Synthese (2017) 194:3563–3582
DOI 10.1007/s11229-016-1107-z

Quine's conjecture on many-sorted logic

Thomas William Barrett[1] · **Hans Halvorson**[1]

Received: 6 September 2015 / Accepted: 25 April 2016 / Published online: 7 May 2016

Abstract Quine often argued for a simple, untyped system of logic rather than the typed systems that were championed by Russell and Carnap, among others. He claimed that nothing important would be lost by eliminating sorts, and the result would be additional simplicity and elegance. In support of this claim, Quine conjectured that every many-sorted theory is equivalent to a single-sorted theory. We make this conjecture precise, and prove that it is true, at least according to one reasonable notion of theoretical equivalence. Our clarification of Quine's conjecture, however, exposes the shortcomings of his argument against many-sorted logic.

Keywords Theoretical equivalence · Definitional equivalence · Morita equivalence · Quine · Model theory · Many-sorted logic

Introduction

During the last century, an idea about the relationship between many-sorted logic and single-sorted logic became a piece of folklore. Many logicians endorsed the idea that many-sorted and single-sorted logic are mere "notational variants" of one another, and that the two frameworks have precisely the same expressive power. One particularly natural way to put this idea is as follows:

> Every many-sorted theory is equivalent to a single-sorted theory.

✉ Thomas William Barrett
thomaswb@princeton.edu

[1] Princeton University, Princeton, USA

Quine was one of the most famous proponents of this claim.[1] For this reason, we call the claim *Quine's conjecture*.

In this paper, we aim to capture the sense in which Quine's conjecture is true. Although the basic idea behind the conjecture is clear, before proving it one needs to make precise what it might mean for two many-sorted theories to be "equivalent." We consider two precise versions of Quine's conjecture. We show that the first version of the conjecture is false, but we prove the second.

Preliminaries

We begin with some preliminaries about theories in many-sorted logic.[2] A **signature** Σ is a set of sort symbols, predicate symbols, function symbols, and constant symbols. Σ must contain at least one sort symbol, and the predicate, function, and constant symbols in Σ are assigned **arities** constructed from the sorts in Σ. The only difference from the syntax of single-sorted logic is that the quantifiers $\forall_\sigma$ and $\exists_\sigma$ that appear in Σ-formulas must be indexed by sorts $\sigma \in \Sigma$.

Given a signature Σ, a **Σ-theory** T is a set of Σ-sentences. The sentences $\phi \in T$ are called the **axioms** of T. If the signature Σ has only one sort symbol, then a Σ-theory T is called a **single-sorted theory**, while if Σ has more than one sort symbol, then T is called a **many-sorted theory**. A Σ-structure M is a **model** of a Σ-theory T if $M \vDash \phi$ for all $\phi \in T$, where $\vDash$ is the standard notion of logical consequence. A theory T **entails** a sentence ϕ, written $T \vDash \phi$, if $M \vDash \phi$ for every model M of T.

We begin with the following preliminary criterion for theoretical equivalence.

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

One can verify that theories T_1 and T_2 are logically equivalent if and only if they entail precisely the same sentences. It is therefore easy to see that logical equivalence is too strict to capture any sense in which Quine's conjecture is true. Theories can only be logically equivalent if they are formulated in the same signature, so no many-sorted theory is logically equivalent to a single-sorted theory.

Logical equivalence is a strict criterion for theoretical equivalence, so logicians and philosophers of science have proposed other more general criteria.[3] We will begin

[1] See in particular Quine (1951, pp. 69–71), Quine (1960, pp. 209–210), and Quine (1963, pp. 267–268). He explains the conjecture as follows: "We can always reduce multiple sorts of variables to one sort if we adopt appropriate predicates. Wherever we might have used a special sort of variable we may use instead a general variable and *restrict* it to the appropriate predicate" (Quine 1963, p. 268). Quine (1969, p. 92) expresses the same idea when he remarks that "notations with one style of variables and notations with many are intertranslatable." Quine (1937, 1938, 1956) provides support for the conjecture by describing a method of "translating" between many-sorted and single-sorted logic and applying it to NBG set theory and Russell's theory of types. Of course, the fact that sorts can be eliminated (or, better, unified) was also discussed in the pure logic literature. For example, see Schmidt (1951).

[2] We here take for granted some basic definitions. The reader is encouraged to consult Hodges (2008) and Barrett and Halvorson (2015b) for details and notation.

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

by considering a criterion called *definitional equivalence* that was introduced into philosophy of science by Glymour (1971, 1977, 1980). In order to describe this criterion, we need to do some work.

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** Σ 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. Similarly, 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

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

where $\phi(x_1, \ldots, x_n, y)$ and $\psi(x)$ are both Σ-formulas. Note that in all of these cases it must be that the sorts $\sigma_1, \ldots, \sigma_n, \sigma \in \Sigma$.

Although they are Σ^+-sentences, (1) and (2) have consequences in the signature Σ. In particular, (1) and (2) imply the following sentences, respectively:[4]

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

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

We now have the resources necessary to describe the concept of a definitional extension. Let $\Sigma \subset \Sigma^+$ be signatures and T a Σ-theory. A **definitional extension** of T to Σ^+ is a Σ^+-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 $T \vDash \alpha_s$.

Footnote 3 continued
Glymour (1977), North (2009), Swanson and Halvorson (2012), Curiel (2014), Knox (2011), Knox (2014), Barrett (2015), Weatherall (2015a, b, c), and Rosenstock et al. (2015) for discussion of whether or not particular physical theories should be considered theoretically equivalent. Finally, see de Bouvére (1965), Kanger (1968), Pinter (1978), Pelletier and Urquhart (2003), Andréka et al. (2005), Friedman and Visser (2014), and Barrett and Halvorson (2015a, b) for some results that have been proven about varieties of theoretical equivalence.

[4] We will use the notation $\exists_{\sigma=n} x \, \phi(x)$ throughout to abbreviate the sentence "there exist exactly n things of sort σ that are ϕ."

One can think of a definitional extension of a theory as "saying no more" than the original theory (Barrett and Halvorson 2015b). It simply allows one to add "abbreviations" of old formulas to the theory. With this thought in mind, we can describe 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." The intuition behind definitional equivalence is simple: T_1 and T_2 are definitionally equivalent if T_1 can define all of the vocabulary that T_2 uses, and in a compatible way, T_2 can define all of the vocabulary that T_1 uses. Definitional equivalence captures a sense in which two theories are "intertranslatable" (Barrett and Halvorson 2015a).

One can easily verify that definitional equivalence is a strictly weaker criterion than logical equivalence. Unlike logical equivalence, theories in different signatures can be definitionally equivalent. But definitional equivalence is still incapable of capturing any sense in which Quine's conjecture is true. As we have described it, a definitional extension does not allow one to define new sorts. If T_1 and T_2 are definitionally equivalent, therefore, they must be formulated in signatures with the same sort symbols. So no many-sorted theory is definitionally equivalent to a single-sorted theory.

Fortunately, there are criteria for theoretical equivalence that are more general than definitional equivalence. The one that will be of particular interest to us is called *Morita equivalence*.[5] Morita equivalence is a natural generalization of definitional equivalence. Indeed, it is essentially the same as definitional equivalence, but it allows one to define new sort symbols in addition to new predicate, function, and constant symbols.

The first version of Quine's conjecture that we will consider is the following.

Quine's conjecture 1

Every theory is Morita equivalent to a single-sorted theory.

In order to understand Quine's conjecture 1, we need to carefully describe Morita equivalence. We begin by discussing how to define new sort symbols. Let $\Sigma \subset \Sigma^+$ be signatures and consider a sort symbol $\sigma \in \Sigma^+ - \Sigma$. One can define the sort σ as a

[5] Quine (1975) proposed his own criteria for equivalence of theories. This criterion suffers from some serious shortcomings, however, so we will not discuss it here (Barrett and Halvorson 2015a). Definitional and Morita equivalence each capture a sense in which two theories are intertranslatable (Barrett and Halvorson 2015a, b). Theories that are equivalent according to these criteria, therefore, can be "mutually interpreted" into one another. One has good reason to prefer these stricter notions of equivalence over mutually interpretability, however, because the latter does not preserve various important theoretical properties. See Visser (2015) for examples of some of these properties, Barrett and Halvorson (2015b) for an introduction to Morita equivalence, and Andréka et al. (2008) and Mere and Veloso (1992) for discussion of closely related ideas.

product sort, a coproduct sort, a subsort, or a quotient sort. In each case one defines σ using old sorts from Σ and new function symbols from $\Sigma^+ - \Sigma$. These new function symbols specify how the new sort σ is related to the old sorts in Σ. We describe in detail these four ways to define new sorts.

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. An explicit 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

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

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. In this case, one 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. An explicit 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 (not necessarily distinct) sort symbols σ_1 and σ_2 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 the subsort σ. An explicit definition of the symbols σ and i as a **subsort** in terms of Σ is a Σ^+-sentence of the form

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

where $\phi(x)$ is a Σ-formula. One should think of σ as "the things of sort σ_1 that are ϕ." The sentence (3) entails the following Σ-sentence:

$$\exists_{\sigma_1} x\, \phi(x)$$

As above, 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 $\sigma_1 \to \sigma$ with $\sigma_1 \in \Sigma$. An explicit 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 $\phi(x_1, x_2)$," and the function symbol ϵ is the "canonical projection" that maps an element to its equivalence class. And indeed, one can verify that the sentence (4) implies that $\phi(x_1, x_2)$ is an equivalence relation. In particular, (4) 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 described 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 three 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 explicit 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 $T \vDash \alpha_s$.

A Morita extension of a theory again "says no more" than the original theory (Barrett and Halvorson 2015b). Like a definitional extension, it can be thought of simply as a way to add in "abbreviations" of old statements into the theory T. A Morita extension, however, allows one to abbreviate old statements using the apparatus of sorts, in addition to the apparatus of predicates, functions, and constant symbols. Our definition of Morita equivalence is perfectly 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 $T_1^1, \ldots, T_1^n$ and $T_2^1, \ldots, T_2^m$ 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 $\Sigma_1 \cup \Sigma_2 \subset \Sigma$.

The intuition behind Morita equivalence is the same as that behind definitional equivalence: T_1 and T_2 are Morita equivalent if they each can, in compatible ways, define all of the vocabulary that the other uses.

One small clarification should be made about this definition. In order to show that two theories are Morita equivalent, one is allowed to take a finite number n of Morita extensions of T_1 and a finite number m of Morita extensions of T_2. One might wonder whether any generality is lost if we were to instead define Morita equivalence as follows:

> ($\star$) T_1 and T_2 are Morita equivalent if there are logically equivalent theories T_1^+ and T_2^+ that are Morita extensions of T_1 and T_2, respectively.

The following example shows that the proposal ($\star$) is significantly less general than the standard definition of Morita equivalence provided above.

Example Let σ_1 and σ_2 be sort symbols, with c a constant symbol of sort σ_1, d a constant symbol of sort σ_2, and i a function symbol of arity $\sigma_2 \to \sigma_1$. We define the signatures $\Sigma_1 = \{\sigma_1, c\}$, $\Sigma_2 = \{\sigma_1, c, \sigma_2, i\}$, and $\Sigma_3 = \{\sigma_1, c, \sigma_2, i, d\}$ and consider the following three theories:

$$
\begin{aligned}
T_1 &= \emptyset \\
T_2 &= \left\{\forall_{\sigma_1} x\big(x = c \leftrightarrow \exists_{\sigma_2} y(i(y) = x)\big) \wedge \forall_{\sigma_2} y_1 \forall_{\sigma_2} y_2\big(i(y_1) = i(y_2) \to y_1 = y_2\big)\right\} \\
T_3 &= T_2 \cup \left\{\forall_{\sigma_2} y\big(y = d \leftrightarrow i(y) = c\big)\right\}
\end{aligned}
$$

It is easy to see that the Σ_2-theory T_2 is a Morita extension of the Σ_1-theory T_1, and that the Σ_3-theory T_3 is a Morita extension of T_2. The theory T_2 defines the sort σ_2 as a subsort containing just the "thing that is c," and T_3 then defines the constant d in such a way that it applies to this new thing. So naturally, one would like to consider T_1 and T_3 equivalent. After all, T_3 is obtained from T_1 simply by defining some new vocabulary.

Although T_1 and T_3 are Morita equivalent according to the standard definition, they are not equivalent according to the proposal ($\star$). The axiom of T_3 that defines d in terms of Σ_2 is *not* a definition of d in terms of Σ_1. And indeed, there simply is no definition of the constant symbol d in terms of Σ_1. Such a definition would have to be a sentence of the form $\forall_{\sigma_2} y\,(y = d \leftrightarrow \psi(y))$, where ψ is a Σ_1-formula. This expression, however, is not well-formed. The constant d and variable y are of sort σ_2, which is not in Σ_1, so ψ cannot be a Σ_1-formula. This means that there is no Morita extension of T_1 that defines the constant symbol d; it takes two Morita extensions of T_1 to define d.

The proposal ($\star$) would therefore not be a satisfactory definition of Morita equivalence, so we use the standard definition for the remainder of this paper.

One can easily verify that Morita equivalence is a strictly weaker criterion than definitional equivalence. If two theories are definitionally equivalent, then they are Morita equivalent. But in general the converse does not hold. As the above example shows, theories can be Morita equivalent even if they are formulated in signatures with different sort symbols.

We now consider Quine's conjecture 1 and show that it is false. The following theorem provides an example of a theory that is not Morita equivalent to any single-sorted theory.

Theorem 1 *Let $\Sigma_1 = \{\sigma_1, \sigma_2, \ldots\}$ be a signature containing a countable infinity of sort symbols. The Σ_1-theory $T_1 = \emptyset$ is not Morita equivalent to any single-sorted theory.*

The idea behind Theorem 1 is simple. One can think of the theory T_1 as saying the following: "Every element is either of kind_1 or of kind_2 or of kind_3 or…, no element is of more than one kind, and there is at least one element of every kind." A single-sorted theory in first-order logic simply does not have the expressive power to say this. (In particular, it cannot express the first conjunct.) The theory T_1 should therefore not be Morita equivalent to any single-sorted theory. Before proving this we need a simple lemma and some notation.

Lemma *Let $\Sigma \subset \Sigma^+$ be signatures and T a Σ-theory. Suppose that A is a model of T with A_σ a finite set for every sort $\sigma \in \Sigma$. Let T^+ be a Morita extension of T to a signature Σ^+ and A^+ a model of T^+ such that $A^+|_\Sigma = A$.[6] Then A^+_σ is a finite set for every $\sigma \in \Sigma^+$.*

Proof Let $\sigma \in \Sigma^+ - \Sigma$ be a sort symbol. We show that A^+_σ is a finite set in the cases where σ is defined as a product sort or a subsort. If T^+ defines σ as a product sort of σ_1 and σ_2, then A^+_σ has exactly as many elements as $A_{\sigma_1} \times A_{\sigma_2}$, which is finite by assumption. If T^+ defines σ as a subsort of $\sigma_1 \in \Sigma$, then the cardinality of A^+_σ is less than or equal to the cardinality of A_{σ_1}, which is also finite. The coproduct and quotient cases follow analogously. □

Our proof of Theorem 1 will expedited by using the following simple category-theoretic machinery.[7] A first-order theory T has a category 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 $1_c : c \to c$. 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, and an arrow $f : M \to N$ between objects in $\text{Mod}(T)$ is an elementary embedding $f : M \to N$ between the models M and N. One can easily verify that $\text{Mod}(T)$ is a category.

A **functor** $F : C \to D$ between categories C and 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

[6] The model $A^+|_\Sigma$ is the Σ-structure obtained from the Σ^+-structure A^+ by "forgetting" the interpretations of symbols in $\Sigma^+ - \Sigma$. One can show that the model A^+ exists and is unique up to isomorphism (Barrett and Halvorson 2015b, Theorem 4.1).

[7] The reader is encouraged to consult Mac Lane (1971) or Borceux (1994) for further details.

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**.

We now turn to the proof of Theorem 1.

Proof of Theorem 1 Suppose for contradiction that there is a single-sorted theory T_2 that is Morita equivalent to T_1. This means that T_2 is a Σ_2-theory with $\sigma \in \Sigma_2$ the unique sort symbol. Let T be the "common Morita extension" of T_1 and T_2 to a signature $\Sigma \supset \Sigma_1 \cup \Sigma_2$. We consider the model A of T_1 defined by $A_{\sigma_i} = \{i, i'\}$ for each $i \in \mathbb{N}$. For every $i \in \mathbb{N}$ there is an isomorphism $f_i : A \to A$ that is the identity on A_{σ_j} for $j \neq i$, but on A_{σ_i} maps $f_i : i \mapsto i'$ and $f_i : i' \mapsto i$. The $f_i : A \to A$ are isomorphisms, and so are elementary embeddings. This implies that there are infinitely many arrows $f : A \to A$ in the category $\mathrm{Mod}(T_1)$.

Since T_1 and T_2 are Morita equivalent, there is an equivalence of categories $F : \mathrm{Mod}(T_1) \to \mathrm{Mod}(T_2)$ such that for every model M of T_1

$$F(M) = M^+|_{\Sigma_2}$$

for some model M^+ of T that is isomorphic to an expansion of M (Barrett and Halvorson 2015b, Theorem 5.1). We consider the model A^+ of T. The Lemma implies that A^+_σ is a finite set. This implies that $F(A)_\sigma$ is a finite set. Since Σ_2 contains only the sort σ and $F(A)_\sigma$ is finite, there can be at most finitely many arrows $g : F(A) \to F(A)$ in the category $\mathrm{Mod}(T_2)$. But since F is an equivalence and therefore full and faithful, this cannot be the case. □

Theorem 1 immediately implies that Quine's conjecture 1 is false. It is not the case that every many-sorted theory is Morita equivalent to a single-sorted theory. This disproof of Quine's conjecture 1, however, suggests the following slight modification of the conjecture.

Quine's conjecture 2

If Σ is a signature with finitely many sorts, then every Σ-theory is Morita equivalent to a single-sorted theory.

This second version of Quine's conjecture is true. One proves Quine's conjecture 2 by explicitly constructing a "corresponding" single-sorted theory $\widehat{T}$ for every many-sorted theory T. The basic idea behind the construction is intuitive. The theory $\widehat{T}$ simply replaces the sort symbols that the theory T uses with predicate symbols.[8] It takes some work, however, to make this idea precise.

[8] This construction recalls the proof that every theory is definitionally equivalent to a theory that uses only predicate symbols (Barrett and Halvorson 2015a, Prop. 2). Quine (1937, 1938, 1956, 1963) suggests the basic idea behind our proof, as do Burgess (2005, p. 12) and Manzano (1996, pp. 221–222). The theorem that we prove here is much more general than Quine's results because we make no assumption about what the theory T is, whereas Quine only considered Russell's theory of types and NBG set theory.

Let Σ be a signature with finitely many sort symbols $\sigma_1, \ldots, \sigma_n$. We begin by constructing a corresponding signature $\widehat{\Sigma}$ that contains one sort symbol σ. The symbols in $\widehat{\Sigma}$ are defined as follows. For every sort symbol $\sigma_j \in \Sigma$ we let q_{σ_j} be a predicate symbol of sort σ. For every predicate symbol $p \in \Sigma$ of arity $\sigma_{j_1} \times \cdots \times \sigma_{j_m}$ we let q_p be a predicate symbol of arity σ^m (the m-fold product of σ). Likewise, for every function symbol $f \in \Sigma$ of arity $\sigma_{j_1} \times \cdots \times \sigma_{j_m} \to \sigma_j$ we let q_f be a predicate symbol of arity σ^{m+1}. And lastly, for every constant symbol $c \in \Sigma$ we let d_c be a constant symbol of sort σ. The single-sorted signature $\widehat{\Sigma}$ corresponding to Σ is then defined to be

$$\widehat{\Sigma} = \{\sigma\} \cup \{q_{\sigma_1}, \ldots, q_{\sigma_n}\} \cup \{q_p : p \in \Sigma\} \cup \{q_f : f \in \Sigma\} \cup \{d_c : c \in \Sigma\}$$

We can now describe a method of "translating" Σ-theories into $\widehat{\Sigma}$-theories. Let T be an arbitrary Σ-theory. We define a corresponding $\widehat{\Sigma}$-theory $\widehat{T}$, and then show that $\widehat{T}$ is Morita equivalent to T.

We begin by translating the axioms of T into the signature $\widehat{\Sigma}$. This will take two steps. First, we describe a way to translate the Σ-terms into $\widehat{\Sigma}$-formulas. Given a Σ-term $t(x_1, \ldots, x_n)$, we define the $\widehat{\Sigma}$-formula $\widehat{\psi}_t(y_1, \ldots, y_n, y)$ recursively as follows.

- If $t(x_1, \ldots, x_n)$ is the variable x_i, then $\widehat{\psi}_t$ is the $\widehat{\Sigma}$-formula $y_i = y$.
- If $t(x_1, \ldots, x_n)$ is the constant c, then $\widehat{\psi}_t$ is the $\widehat{\Sigma}$-formula $d_c = y$.
- Suppose that $t(x_1, \ldots, x_n)$ is the term $f(t_1(x_1, \ldots, x_n), \ldots, t_k(x_1, \ldots, x_n))$ and that each of the $\widehat{\Sigma}$-formulas $\widehat{\psi}_{t_i}(y_1, \ldots, y_n, y)$ have been defined. Then $\widehat{\psi}_t(y_1, \ldots, y_n, y)$ is the $\widehat{\Sigma}$-formula

$$\exists_\sigma z_1 \ldots \exists_\sigma z_k \big(\widehat{\psi}_{t_1}(y_1, \ldots, y_n, z_1) \wedge \ldots \wedge \widehat{\psi}_{t_k}(y_1, \ldots, y_n, z_k) \wedge q_f(z_1, \ldots, z_k, y)\big)$$

One can think of the formula $\psi_t(y_1, \ldots, y_n, y)$ as the translation of the expression "$t(x_1, \ldots, x_n) = x$" into the signature $\widehat{\Sigma}$.

Second, we use this map from Σ-terms to $\widehat{\Sigma}$-formulas to describe a map from Σ-formulas to $\widehat{\Sigma}$-formulas. Given a Σ-formula $\psi(x_1, \ldots, x_n)$, we define the $\widehat{\Sigma}$-formula $\widehat{\psi}(y_1, \ldots, y_n)$ recursively as follows.

- If $\psi(x_1, \ldots, x_n)$ is $t(x_1, \ldots, x_n) = s(x_1, \ldots, x_n)$, where s and t are Σ-terms of sort σ_i, then $\widehat{\psi}(y_1, \ldots, y_n)$ is the $\widehat{\Sigma}$-formula

$$\exists_\sigma z \big(\widehat{\psi}_t(y_1, \ldots, y_n, z) \wedge \widehat{\psi}_s(y_1, \ldots, y_n, z) \wedge q_{\sigma_i}(z)\big)$$

- If $\psi(x_1, \ldots, x_n)$ is $p(t_1(x_1, \ldots, x_n), \ldots, t_k(x_1, \ldots, x_n))$, where $p \in \Sigma$ is a predicate symbol, then $\widehat{\psi}(y_1, \ldots, y_n)$ is the $\widehat{\Sigma}$-formula

$$\exists_\sigma z_1 \ldots \exists_\sigma z_k \big(\widehat{\psi}_{t_1}(y_1, \ldots, y_n, z_1) \wedge \ldots \wedge \widehat{\psi}_{t_k}(y_1, \ldots, y_n, z_k) \wedge q_p(z_1, \ldots, z_k)\big)$$

- This definition extends to all Σ-formulas in the standard way. We define the $\widehat{\Sigma}$-formulas $\widehat{\neg\psi} := \neg\widehat{\psi}$, $\widehat{\psi_1 \wedge \psi_2} := \widehat{\psi_1} \wedge \widehat{\psi_2}$, $\widehat{\psi_1 \vee \psi_2} := \widehat{\psi_1} \vee \widehat{\psi_2}$, and $\widehat{\psi_1 \to \psi_2} :=$

$\widehat{\psi_1} \to \widehat{\psi_2}$. Furthermore, if $\psi(x_1, \ldots, x_n, x)$ is a Σ-formula, then we define both of the following:

$$\widehat{\forall_{\sigma_i} x \psi} := \forall_\sigma y(q_{\sigma_i}(y) \to \widehat{\psi}(y_1, \ldots, y_n, y))$$
$$\widehat{\exists_{\sigma_i} x \psi} := \exists_\sigma y(q_{\sigma_i}(y) \wedge \widehat{\psi}(y_1, \ldots, y_n, y))$$

One should think of the formula $\widehat{\psi}$ as the translation of the Σ-formula ψ into the signature $\widehat{\Sigma}$.

This allows us to consider the translations $\widehat{\alpha}$ of the axioms $\alpha \in T$. The single-sorted theory $\widehat{T}$ will have the $\widehat{\Sigma}$-sentences $\widehat{\alpha}$ as some of its axioms. But $\widehat{T}$ will have more axioms than just the sentences $\widehat{\alpha}$. It will also have some **auxiliary axioms**. These auxiliary axioms will guarantee that the symbols in $\widehat{\Sigma}$ "behave like" their counterparts in Σ. We define auxiliary axioms for the predicate symbols $q_{\sigma_1}, \ldots, q_{\sigma_n} \in \widehat{\Sigma}$, $q_p \in \widehat{\Sigma}$, and $q_f \in \widehat{\Sigma}$, and for the constant symbols $d_c \in \widehat{\Sigma}$. We discuss each of these four cases in detail.

We first define auxiliary axioms to guarantee that the symbols $q_{\sigma_1}, \ldots, q_{\sigma_n}$ behave like sort symbols. The $\widehat{\Sigma}$-sentence ϕ is defined to be $\forall_\sigma y(q_{\sigma_1}(y) \vee \ldots \vee q_{\sigma_n}(y))$.[9] Furthermore, for each sort symbol $\sigma_j \in \Sigma$ we define the $\widehat{\Sigma}$-sentence ϕ_{σ_j} to be

$$\exists_\sigma y(q_{\sigma_j}(y)) \wedge \forall_\sigma y\big(q_{\sigma_j}(y) \to (\neg q_{\sigma_1}(y) \wedge \ldots \wedge \neg q_{\sigma_{j-1}}(y) \wedge \neg q_{\sigma_{j+1}}(y) \wedge \ldots \wedge \neg q_{\sigma_n}(y))\big)$$

One can think of the sentences $\phi_{\sigma_1}, \ldots, \phi_{\sigma_n}$, and ϕ as saying that "everything is of some sort, nothing is of more than one sort, and every sort is nonempty."

Next we define auxiliary axioms to guarantee that the symbols q_p, q_f, and d_c behave like their counterparts p, f, and c in Σ. For each predicate symbol $p \in \Sigma$ of arity $\sigma_{j_1} \times \cdots \times \sigma_{j_m}$, we define the $\widehat{\Sigma}$-sentence ϕ_p to be

$$\forall_\sigma y_1 \ldots \forall_\sigma y_m \Big(q_p(y_1, \ldots, y_m) \to \Big(q_{\sigma_{j_1}}(y_1) \wedge \ldots \wedge q_{\sigma_{j_m}}(y_m)\Big)\Big)$$

This sentence restricts the extension of q_p to the subdomain of n-tuples satisfying $q_{\sigma_{j_1}}, \ldots, q_{\sigma_{j_m}}$, guaranteeing that the predicate q_p has "the appropriate arity." Consider, for example, the case of a unary predicate p of sort σ_i. In that case, ϕ_p says that

$$\forall_\sigma y(q_p(y) \to q_{\sigma_i}(y)),$$

which means that nothing outside the subdomain q_{σ_i} satisfies q_p. Note, however, that here we have made a conventional choice. We could just as well have stipulated that q_p applies to *everything* outside of the subdomain q_{σ_i}. All that matters here is that q_p is trivial (either trivially true, or trivially false) except on the subdomain of objects satisfying q_{σ_i}.

[9] Note that if there were infinitely many sort symbols in Σ, then we could not define the $\widehat{\Sigma}$-sentence ϕ in this way.

For each function symbol $f \in \Sigma$ of arity $\sigma_{j_1} \times \cdots \times \sigma_{j_m} \to \sigma_j$ we define the $\widehat{\Sigma}$-sentence ϕ_f to be the conjunction

$$\forall_\sigma y_1 \ldots \forall_\sigma y_m \forall_\sigma y \big(q_f(y_1, \ldots, y_m, y) \to (q_{\sigma_{j_1}}(y_1) \wedge \ldots \wedge q_{\sigma_{j_m}}(y_m) \wedge q_{\sigma_j}(y))\big)$$
$$\wedge \forall_\sigma y_1 \ldots \forall_\sigma y_m \big((q_{\sigma_{j_1}}(y_1) \wedge \ldots \wedge q_{\sigma_{j_m}}(y_m)) \to \exists_{\sigma=1} y(q_f(y_1, \ldots, y_m, y))\big)$$

The first conjunct guarantees that the symbol q_f has "the appropriate arity," and the second conjunct guarantees that q_f behaves like a function. Lastly, if $c \in \Sigma$ is a constant symbol of arity σ_j, then we define the $\widehat{\Sigma}$-sentence ϕ_c to be $q_{\sigma_j}(d_c)$. This sentence guarantees that the constant symbol d_c also has "the appropriate arity."

We now have the resources to define a $\widehat{\Sigma}$-theory $\widehat{T}$ that is Morita equivalent to T.

$$\begin{aligned}\widehat{T} = \{\widehat{\alpha} : \alpha \in T\} &\cup \{\phi, \phi_{\sigma_1}, \ldots, \phi_{\sigma_n}\} \\ &\cup \{\phi_p : p \in \Sigma\} \\ &\cup \{\phi_f : f \in \Sigma\} \\ &\cup \{\phi_c : c \in \Sigma\}\end{aligned}$$

The theory $\widehat{T}$ has two kinds of axioms, the translated axioms of T and the auxiliary axioms. These axioms allow $\widehat{T}$ to imitate the theory T in the signature $\widehat{\Sigma}$. Indeed, one can prove the following result.

Theorem 2 *The theories T and $\widehat{T}$ are Morita equivalent.*

The proof of Theorem 2 requires some work, and has therefore been placed in an appendix. But the idea behind the proof is simple. The theory T needs to define symbols in $\widehat{\Sigma}$. It defines the sort symbol σ as a "universal sort," by taking the coproduct of the sorts $\sigma_1, \ldots, \sigma_n \in \Sigma$. The theory T then defines the symbols q_p, q_f, and d_c in $\widehat{\Sigma}$ simply by using the corresponding symbols p, f, and c in Σ. Likewise, the theory $\widehat{T}$ needs to define the symbols in Σ. It defines the sort symbol σ_j as the subsort of "things that are q_{σ_j}" for each $j = 1, \ldots, n$. And $\widehat{T}$ defines the symbols p, f, and c again by using the corresponding symbols q_p, q_f, and d_c.

Since the theory T was arbitrary, Theorem 2 immediately implies that Quine's conjecture 2 is true.

Philosophical implications

Our proof of Quine's conjecture 2 substantiates Quine's original thought about the relationship between many-sorted logic and single-sorted logic. It captures a precise sense in which single-sorted logic has exactly the same expressive power as (finitely) many-sorted logic.

In capturing this relationship, our results provide a reason to prefer Morita equivalence over definitional equivalence. The original motivation for taking definitional equivalence as the standard for theoretical equivalence is intuitive: Definitional equivalence captures a sense in which two theories are intertranslatable. If T_1 and T_2 are

definitionally equivalent, then there are two precise ways to translate between the theories.

- The first way of translating between T_1 and T_2 is purely syntactic. Formulas of T_1 can be translated into the signature of T_2, and vice versa, and these translations have some nice features. In particular, theorems are translated into theorems, and the two translations are "almost inverse" to one another (Barrett and Halvorson 2015a, Theorem 1).[10]
- The second way of translating is semantic. Models of T_1 can be translated into models of T_2, and vice versa, and these translations also have some nice features. In particular, the two translations induce a categorical equivalence between the categories of models of the two theories (Barrett and Halvorson 2015b, Theorems 3.1 and 5.1).

Morita equivalence also captures a sense in which two theories are intertranslatable. Indeed, if two theories are Morita equivalent, then one can again translate formulas and convert models between them (Barrett and Halvorson 2015b, Theorems 4.3 and 5.1).

The motivation for taking definitional equivalence as the standard for theoretical equivalence therefore carries over to Morita equivalence. But our results show that Morita equivalence has a virtue that definitional equivalence lacks. In particular, if one follows Glymour (1971, 1977, 1980) in taking definitional equivalence as the standard for theoretical equivalence, then one cannot recover any sense in which Quine's conjecture is true. If one wants to substantiate Quine's conjecture and capture the relationship between the many-sorted and single-sorted frameworks, one needs to move to Morita equivalence.

Our discussion also yields a cautionary remark concerning a particular argument of Quine's. One might be tempted to conclude from Theorem 2 that many-sorted logic is dispensable. In fact, this is precisely the conclusion that Quine himself drew. Quine says that explicitly naming sorts and specifying the sorts of the vocabulary in the signature of a theory is an "artificial device" (Quine and Carnap 1990, p. 409), and that we are licensed to ignore it. He sums up his negative stance on many-sorted logic as follows:

> All in all, I find an overwhelming case for a single unpartitioned universe of values of bound variables, and a simple grammar of predication which admits general terms all on an equal footing. Subsidiary distinctions can still be drawn as one pleases, both on methodological considerations and on considerations of natural kind; but we may think of them as distinctions special to the sciences and unreflected in the structure of our notation (Quine 1960, p. 229).[11]

[10] Intuitively, two translations F and G between theories are "almost inverse" if both $F \circ G$ and $G \circ F$ map every formula to a provably equivalent (but not necessarily equal) formula. The reader is invited to consult Barrett and Halvorson (2015a) for a precise definition. See Friedman and Visser (2014) for alternative characterizations of definitional equivalence and intertranslatability.

[11] For other expressions of this same attitude, see Quine (1951, pp. 69–71), Quine (1963, pp. 267–268), and Quine and Carnap (1990, p. 409).

Quine's reasoning here may be seductive, but it is nonetheless misleading. We have shown that for each theory T in a signature Σ with finitely many sorts, there is a Morita equivalent single-sorted theory $\widehat{T}$. This result does *not* imply that we can ignore sorts altogether. There are a number of reasons for this.

First, Quine's argument against many-sorted logic is circular. It relies on a theorem that can only be stated in the framework of many-sorted logic. In capturing the sense in which Quine's conjecture is true, one employs Morita equivalence, a standard of equivalence that is itself inherently many-sorted. If one does away with the many-sorted framework altogether, one can no longer capture this equivalence; the single-sorted framework simply does not provide one with a criterion for equivalence that is capable of substantiating Quine's conjecture.

The many-sorted framework simply allows us to better recognize the variety of ways in which theories might be equivalent. For example, one might want to consider Euclidean geometry formulated using lines to be equivalent to Euclidean geometry formulated using points. Although both of these theories are single-sorted, the single-sorted framework does not allow one to consider them equivalent. With Morita equivalence in hand, however, the many-sorted framework does allow one capture this equivalence (Barrett and Halvorson 2016). The issue here is that two single-sorted theories might have different sorts, in which case there is a non-trivial question about whether these sorts can be defined from each other. One theory's single sort might, for example, be definable as a product of another theory's single sort. In particular, this means that two single-sorted theories can be Morita equivalent even if they are not definitionally equivalent. The following simple example serves to illustrate this fact.

Example Consider the signatures $\Sigma_1 = \{\sigma_1\}$ and $\Sigma_2 = \{\sigma_2, p\}$, where σ_1 and σ_2 are sort symbols, and p is a unary predicate of sort σ_2. We define the following two theories.

$$T_1 = \{\exists_{\sigma_1=1} x\ (x = x)\}$$
$$T_2 = \{\exists_{\sigma_2=2} y\ (y = y), \exists_{\sigma_2=1} y\ p(y)\}$$

The Σ_1-theory T_1 says "there is only one thing", while the Σ_2-theory T_2 says "there are exactly two things, the first is p, and the second is $\neg p$." These two theories are Morita equivalent. In order to prove this, one shows that T_1 can define all of the symbols that T_2 uses, and vice versa. The theory T_1 defines the sort symbol σ_2 as the coproduct of the sort σ_1 with itself, and then defines the predicate p to hold of the element in the image of the canonical injection ρ_1 that is associated with the new coproduct sort σ_2. And on the other hand, the theory T_2 defines the sort symbol σ_1 as the subsort of "things that are p." T_1 and T_2 are therefore Morita equivalent.

Suppose, however, that we were to follow Quine and ignore the apparatus of sorts when considering these theories T_1 and T_2. That is, suppose that we were to treat T_1 and T_2 as if they were both formulated using the same single sort symbol. The single-sorted framework then gives us no way of recognizing any sense in which these theories are equivalent. The theory T_1 implies that "there is one thing", while the theory T_2 implies that "there are two things." These implications contradict one another, so T_1 and T_2 do not even have a common conservative extension, let alone a common

definitional extension.[12] They are therefore not definitionally equivalent. If we ignore the apparatus of sorts when considering the theories T_1 and T_2, then we are also forced to forget a natural sense in which they are equivalent theories.[13]

When Quine argues that nothing of philosophical significance turns on the use of many-sorted as opposed to single-sorted logic, therefore, he is mistaken. Indeed, recall that Quine used the fact that many-sorted logic could be reduced to single-sorted logic as a tool in his debates with Carnap. While Carnap finds many-sorted logic helpful for capturing the phenomenon of category mistakes, Quine claims that it is more convenient to say that sentences committing category mistakes are meaningful, but trivially false:

> […] the many-sorted is translatable into one-sorted. Generally such translation has the side effect of admitting as meaningful some erstwhile meaningless predications. E.g., if the predicate "divisible by 3" is henceforth to be trained on general variables instead of number variables, we must make sense of calling things other than numbers divisible by 3. But this is easy; we may count such attributions as false instead of meaningless (Quine 1969, p. 96).

Quine has again missed a beat here. Recall that in eliminating sort distinctions, we had to make a conventional choice.[14] For a predicate that originally was restricted to things of a particular sort, what should we say about its application to things of another sort? Should we say that it applies to all such things, or to no such things? There is no default answer here because we have to consider the question both for a predicate and for its negation (in which case our answers will be opposite to each other).

Consider a particular example. The predicates "prime" and "composite" are defined on the domain of natural numbers. Now suppose that we unify the domain of natural numbers with the domain of 20th century philosophers. Following Quine's suggestion, we should count the attribution of "prime" to philosophers as false, e.g. "Carnap is prime," is considered false. By the same reasoning, we also count the attribution of "composite" (i.e. "not prime") to philosophers as false, e.g. "Carnap is composite," is considered false. But then Carnap is neither prime nor not-prime, which cannot be tolerated within the framework of classical logic.

Quine is therefore wrong when he claims that "we may count such attributions as false." If we count *all* such attributions as false, then we land ourselves in a contradiction. We have to choose which attributions are true, and which are false. The many-sorted framework allows us to refrain from making such conventional choices, which could easily be mistaken for assertions. And this is an advantage of the frame-

[12] Note that if we remember that T_1 and T_2 use different sort symbols (as we did in the example above), then we can recognize that they have a common conservative extension. In this case, T_1 implies that "there is one thing of sort σ_1", while T_2 implies that "there are two things of sort σ_2", and these implications no longer contradict one another.

[13] Manzano (1996, pp. 221–222) puts our attitude nicely: "It is well-known that many-sorted logic reduces to [single-sorted] logic, and this approach is the one commonly used in textbooks. […] What is not usually said in textbooks is that the reduction has a price."

[14] See our definition of the sentence ϕ_p and the surrounding discussion in the construction of the theory $\hat{T}$.

work. It does not force us to apply predicates in cases where we have no good reason to say that they do (or do not) hold of the items in question.

Recent philosophical literature has devoted significant attention to logic and quantifiers, but relatively little to the device of sorts.[15] Some of us felt justified to ignore sorts because we believed Quine when he claimed that every theory can be replaced by a single-sorted variant. But when we look at the actual mechanics of how this replacement works, it becomes clear that Quine was mistaken about the philosophical upshot of the logical facts. Many-sorted theories are eliminable only to the extent that Morita equivalence is a reasonable notion of equivalence, and Morita equivalence only makes sense in the framework of many-sorted logic. When we forget the apparatus of sorts, we are also foced forget a particularly natural and useful criterion for theoretical equivalence. And in addition, eliminating sort distinctions forces us to make unnecessary conventional choices about how to extend predicates beyond their original range of application.

Through a bit of logical legerdemain, Quine convinced a generation of philosophers that formal logic can do without the apparatus of sorts. But the formal result — that each theory can be replaced with a single-sorted variant — does not support the philosophical conclusion. If we forget about the apparatus of sorts and replace each theory with a single-sorted variant, then philosophical questions will be begged and an interesting point of view will be lost.

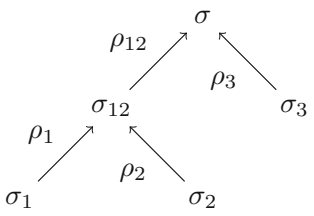
Appendix

The objective of this appendix is to prove Theorem 2. We prove a special case of the result for convenience. We will assume that Σ has only three sort symbols $\sigma_1, \sigma_2, \sigma_3$ and that Σ does not contain function or constant symbols. A perfectly analogous (though more tedious) proof goes through in the general case.

We prove the result by explicitly constructing a "common Morita extension" $T_4 \cong \widehat{T}_4$ of T and $\widehat{T}$ to the following signature.

$$\Sigma^+ = \Sigma \cup \widehat{\Sigma} \cup \{\sigma_{12}\} \cup \{\rho_1, \rho_2, \rho_{12}, \rho_3\} \cup \{i_1, i_2, i_3\}$$

The symbol $\sigma_{12} \in \Sigma^+$ is a sort symbol. The symbols denoted by subscripted ρ are function symbols. Their arities are expressed in the following figure.

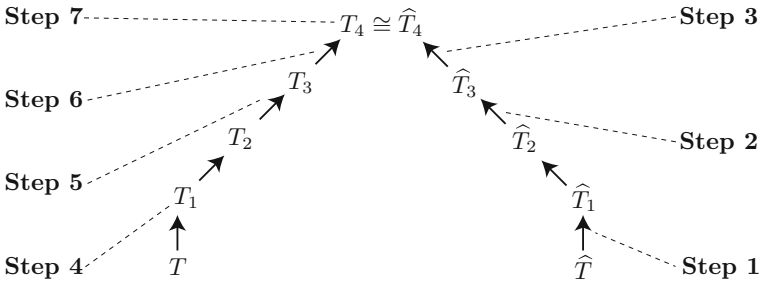


[15] Of course, there are exceptions, such as Turner (2010).

The symbols i_1, i_2, and i_3 are function symbols with arity $\sigma_1 \to \sigma$, $\sigma_2 \to \sigma$, and $\sigma_3 \to \sigma$, respectively.

We now turn to the proof.

Proof of Theorem 2 The following figure illustrates how our proof will be organized.



Steps 1–3 define the theories $\widehat{T}_1, \ldots, \widehat{T}_4$, steps 4–6 define $T_1, \ldots, T_4$, and step 7 shows that T_4 and $\widehat{T}_4$ are logically equivalent.

Step 1 We begin by defining the theory $\widehat{T}_1$. For each sort $\sigma_j \in \Sigma$ we consider the following sentence.

$$\forall_\sigma y \left(q_{\sigma_j}(y) \leftrightarrow \exists_{\sigma_j} x (i_j(x) = y)\right) \qquad (\theta_{\sigma_j})$$
$$\wedge\, \forall_{\sigma_j} x_1 \forall_{\sigma_j} x_2 (i_j(x_1) = i_j(x_2) \to x_1 = x_2)$$

The sentence θ_{σ_j} defines the symbols σ_j and i_j as the subsort of "things that are q_{σ_j}." The auxiliary axioms ϕ_{σ_j} of $\widehat{T}$ guarantee that the admissibility conditions for these definitions are satisfied. The theory $\widehat{T}_1 = \widehat{T} \cup \{\theta_{\sigma_1}, \theta_{\sigma_2}, \theta_{\sigma_3}\}$ is therefore a Morita extension of $\widehat{T}$ to the signature $\widehat{\Sigma} \cup \{\sigma_1, \sigma_2, \sigma_3, i_1, i_2, i_3\}$.

Step 2 We now define the theories $\widehat{T}_2$ and $\widehat{T}_3$. Let $\theta_{\sigma_{12}}$ be a sentence that defines the symbols σ_{12}, ρ_1, ρ_2 as a coproduct sort. The theory $\widehat{T}_2 = \widehat{T}_1 \cup \{\theta_{\sigma_{12}}\}$ is clearly a Morita extension of $\widehat{T}_1$.

We have yet to define the function symbols ρ_{12} and ρ_3. The following two sentences define these symbols.

$$\forall_{\sigma_3} x \forall_\sigma y (\rho_3(x) = y \leftrightarrow i_3(x) = y) \qquad (\theta_{\rho_3})$$
$$\forall_{\sigma_{12}} x \forall_\sigma y (\rho_{12}(x) = y \leftrightarrow \psi(x, y)) \qquad (\theta_{\rho_{12}})$$

The sentence θ_{ρ_3} simply defines ρ_3 to be equal to the function i_3. For the sentence $\theta_{\rho_{12}}$, we define the formula $\psi(x, y)$ to be

$$\exists_{\sigma_1} z_1 \big(\rho_1(z_1) = x \wedge i_1(z_1) = y\big) \vee \exists_{\sigma_2} z_2 \big(\rho_2(z_2) = x \wedge i_2(z_2) = y\big)$$

We should take a moment here to understand the definition $\theta_{\rho_{12}}$. We want to define what the function ρ_{12} does to an element a of sort σ_{12}. Since the sort σ_{12} is the coproduct of the sorts σ_1 and σ_2, the element a must "actually be" of one of the sorts σ_1 or σ_2. (The

disjuncts in the formula $\psi(x, y)$ correspond to these possibilities.) The definition $\theta_{\rho_{12}}$ stipulates that if a is "actually" of sort σ_j, then the value of ρ_{12} at a is the same as the value of i_j at a. One can verify that $\widehat{T}_2$ satisfies the admissibility conditions for θ_{ρ_3} and $\theta_{\rho_{12}}$, so the theory $\widehat{T}_3 = \widehat{T}_2 \cup \{\theta_{\rho_3}, \theta_{\rho_{12}}\}$ is a Morita extension of $\widehat{T}_2$ to the signature

$$\widehat{\Sigma} \cup \{\sigma_1, \sigma_2, \sigma_3, \sigma_{12}, i_1, i_2, i_3, \rho_1, \rho_2, \rho_3, \rho_{12}\}$$

Step 3 We now describe the Σ^+-theory $\widehat{T}_4$. This theory defines the predicates in the signature Σ. Let $p \in \Sigma$ be a predicate symbol of arity $\sigma_{j_1} \times \cdots \times \sigma_{j_m}$. We consider the following sentence.

$$\forall_{\sigma_{j_1}} x_1 \ldots \forall_{\sigma_{j_m}} x_m \left(p(x_1, \ldots, x_m) \leftrightarrow q_p(i_{j_1}(x_1), \ldots, i_{j_m}(x_m))\right) \qquad (\theta_p)$$

The theory $\widehat{T}_4 = \widehat{T}_3 \cup \{\theta_p : p \in \Sigma\}$ is therefore a Morita extension of $\widehat{T}_3$ to the signature Σ^+.

Step 4 We turn to the left-hand side of our organizational figure and define the theories T_1 and T_2. We proceed in an analogous manner to the first part of Step 2. The theory $T_1 = T \cup \{\theta_{\sigma_{12}}\}$ is a Morita extension of T to the signature $\Sigma \cup \{\sigma_{12}, \rho_1, \rho_2\}$. Now let θ_σ be the sentence that defines the symbols σ, ρ_{12}, ρ_3 as a coproduct sort. The theory $T_2 = T_1 \cup \{\theta_\sigma\}$ is a Morita extension of T_1 to the signature $\Sigma \cup \{\sigma_{12}, \sigma, \rho_1, \rho_2, \rho_3, \rho_{12}\}$.

Step 5 This step defines the function symbols i_1, i_2, and i_3. We consider the following sentences.

$$\forall_{\sigma_3} x_3 \forall_\sigma y(i_3(x_3) = y \leftrightarrow \rho_3(x_3) = y) \qquad (\theta_{i_3})$$

$$\forall_{\sigma_2} x_2 \forall_\sigma y\big(i_2(x_2) = y \leftrightarrow \exists_{\sigma_{12}} z(\rho_2(x_2) = z \wedge \rho_{12}(z) = y)\big) \qquad (\theta_{i_2})$$

$$\forall_{\sigma_1} x_1 \forall_\sigma y\big(i_1(x_1) = y \leftrightarrow \exists_{\sigma_{12}} z(\rho_1(x_1) = z \wedge \rho_{12}(z) = y)\big) \qquad (\theta_{i_1})$$

The sentence θ_{i_3} defines the function symbol i_3 to be equal to ρ_3. The sentence θ_{i_2} defines the function symbol i_2 to be equal to the composition "$\rho_{12} \circ \rho_2$." Likewise, the sentence θ_{i_1} defines the function symbol i_1 to be "$\rho_{12} \circ \rho_1$." The theory $T_3 = T_2 \cup \{\theta_{i_1}, \theta_{i_2}, \theta_{i_3}\}$ is a Morita extension of T_2 to the signature $\Sigma \cup \{\sigma_{12}, \sigma, \rho_1, \rho_2, \rho_3, \rho_{12}, i_1, i_2, i_3\}$.

Step 6 We still need to define the predicate symbols in $\widehat{\Sigma}$. Let $\sigma_j \in \Sigma$ be a sort symbol and $p \in \Sigma$ a predicate symbol of arity $\sigma_{j_1} \times \cdots \times \sigma_{j_m}$. We consider the following sentences.

$$\forall_\sigma y(q_{\sigma_j}(y) \leftrightarrow \exists_{\sigma_j} x(i_j(x) = y)) \qquad (\theta_{q_{\sigma_j}})$$

$$\forall_\sigma y_1 \ldots \forall_\sigma y_m\big(q_p(y_1, \ldots, y_m) \leftrightarrow \exists_{\sigma_{j_1}} x_1 \ldots \exists_{\sigma_{j_m}} x_m(i_{j_1}(x_1) = y_1 \wedge \ldots \wedge i_{j_m}(x_m) = y_m \wedge p(x_1, \ldots, x_m))\big) \qquad (\theta_{q_p})$$

These sentences define the predicates $q_{\sigma_j} \in \widehat{\Sigma}$ and $q_p \in \widehat{\Sigma}$. One can verify that T_3 satisfies the admissibility conditions for the definitions $\theta_{q_{\sigma_j}}$. And therefore the theory

$T_4 = T_3 \cup \{\theta_{q_{\sigma_1}}, \theta_{q_{\sigma_2}}, \theta_{q_{\sigma_3}}\} \cup \{\theta_{q_p} : p \in \Sigma\}$ is a Morita extension of T_3 to the signature Σ^+.

Step 7 It only remains to show that the Σ^+-theories T_4 and $\widehat{T}_4$ are logically equivalent. One can verify by induction on the complexity of ψ that

$$T_4 \vDash \psi \leftrightarrow \widehat{\psi} \text{ and } \widehat{T}_4 \vDash \psi \leftrightarrow \widehat{\psi}. \tag{5}$$

for every Σ-sentence ψ. One then uses (5) to show that $\mathrm{Mod}(T_4) = \mathrm{Mod}(\widehat{T}_4)$. The argument involves a number of cases, but since each case is straightforward we leave them to the reader to verify. The theories T_4 and $\widehat{T}_4$ are logically equivalent, which implies that T and $\widehat{T}$ are Morita equivalent. □

References

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

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.

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. *Journal of Philosophical Logic*.

Barrett, T. W., & Halvorson, H. (2015b). Morita equivalence. *arXiv preprint*.

Barrett, T. W., & Halvorson, H. (2016). From geometry to conceptual relativity.

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

Burgess, J. P. (2005). *Fixing frege*. Princeton: Princeton 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. In: *Symposium on the Theory of Models*, pp. 402–406. Amsterdam: North-Holland Publishing Company.

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. In *PSA 1970*, pp. 275–288. Berlin: Springer.

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

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

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

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. In *Oxford Handbooks Online*.

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

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

Knox, E. (2011). Newton-Cartan theory and teleparallel gravity: The force of a formulation. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *42*(4), 264–275.

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

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

Manzano, M. (1996). *Extensions of first order logic*. Cambridge: cambridge University Press.

Mere, M. C., & Veloso, P. (1992). On extensions by sorts. *Monografias em Ciências da Computaçao, DI, PUC-Rio*, *38*, 92.
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. (1937). New foundations for mathematical logic. *The American Mathematical Monthly*, *44*(2), 70–80.
Quine, W. V. O. (1938). On the theory of types. *The Journal of Symbolic Logic*, *3*(04), 125–139.
Quine, W. V. O. (1951). On Carnap's views on ontology. *Philosophical Studies*, *2*(5), 65–72.
Quine, W. V. O. (1956). Unification of universes in set theory. *The Journal of Symbolic Logic*, *21*(03), 267–279.
Quine, W. V. O. (1960). *Word and object*. Cambridge: MIT Press.
Quine, W. V. O. (1963). *Set theory and its logic*. Cambridge: Harvard University Press.
Quine, W. V. O. (1969). Existence and quantification. In *Ontological relativity and other essays* (pp. 91–113). New York: Columbia University Press.
Quine, W. V. O. (1975). On empirically equivalent systems of the world. *Erkenntnis*, *9*(3), 313–328.
Quine, W. V. O., & Carnap, R. (1990). *Dear Carnap, Dear Van: The Quine-Carnap correspondence and related work*. Berkeley, CA: University of California Press.
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.
Sklar, L. (1982). Saving the noumena. *Philosophical Topics*, *13*, 89–110.
Swanson, N., & Halvorson, H. (2012). On North's 'The structure of physics'.
Turner, J. (2010). Ontological pluralism. *The Journal of Philosophy*, *107*(1), 5–34.
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.
Visser, A. (2015). On Q. In *Logic Group Preprint Series*, p. 330.
Weatherall, J. O. (2015a). Are Newtonian gravitation and geometrized Newtonian gravitation theoretically equivalent? *Erkenntnis*, 1–19.
Weatherall, J. O. (2015b). Categories and the foundations of classical field theories. In E. Landry (Ed.), *Forthcoming in categories for the working philosopher*. Oxford: Oxford University Press.
Weatherall, J. O. (2015c). Understanding gauge. *Philosophy of Science*.