Glymour and Quine on Theoretical Equivalence*

Thomas William Barrett†
Hans Halvorson‡

Abstract

Glymour (1970, 1977, 1980) and Quine (1975) propose two different formal criteria for theoretical equivalence. In this paper we examine the relationships between these criteria.

1 Introduction

Philosophers of science have long been concerned with the conditions under which theories might be considered equivalent.[1] One way that this issue has been approached is by proposing different formal criteria for theoretical equivalence. In this paper we will discuss two such criteria. The first was proposed by Glymour (1970, 1977, 1980) and the second was proposed by Quine (1975).

We begin by showing that Quine's criterion is unsatisfactory. It considers some theories to be equivalent that one has good reason to consider *in*equivalent. But Quine's criterion can be amended in such a way that it no longer makes these undesirable verdicts. Indeed, we will isolate a precise sense in which Glymour's criterion is such an amendment.

2 Preliminaries

The criteria for theoretical equivalence that we will discuss are most naturally understood in the framework of first-order logic. We therefore begin by presenting some preliminaries about this framework.[2]

A **signature** Σ is a set of predicate symbols, function symbols, and constant symbols. The Σ-terms, Σ-formulas, and Σ-sentences are recursively defined in the standard way. A **Σ-structure** A is a nonempty set in which the symbols of Σ have been interpreted. One recursively defines when a sequence of elements $a_1, \ldots, a_n \in A$ **satisfy** a Σ-formula $\phi(x_1, \ldots, x_n)$ in a Σ-structure A, written

*Thanks to JB Manchak, Jim Weatherall, Jeff Barrett, Albert Visser for comments and discussion.

†thomaswb@princeton.edu

‡hhalvors@princeton.edu

[1] For example, North (2009), Halvorson (2011), Swanson and Halvorson (2012), Curiel (2014), and Barrett (2014) discuss whether Hamiltonian and Lagrangian mechanics are theoretically equivalent. Glymour (1977), Weatherall (2014), and Knox (2013) discuss standard Newtonian gravitation and geometrized Newtonian gravitation. And Sklar (1982), Halvorson (2012, 2013), Glymour (2013), van Fraassen (2014), and Coffey (2014) discuss more general issues about theoretical equivalence.

[2] The reader is encouraged to consult Hodges (2008) for details and notation.

$A \vDash \phi[a_1, \ldots, a_n]$. If ϕ is a Σ-sentence, then $A \vDash \phi$ just in case the empty sequence satisfies ϕ in A.

A **Σ-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 $\phi \in T$. We will use the notation $\mathrm{Mod}(T)$ to denote the class of models of a theory T. 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, i.e. if $\mathrm{Mod}(T_1) = \mathrm{Mod}(T_2)$.

One can easily verify that T_1 and T_2 are logically equivalent if and only if $\{\phi : T_1 \vDash \phi\} = \{\psi : T_2 \vDash \psi\}$. Note that logical equivalence can only apply to theories T_1 and T_2 that are formulated in the same signature.

3 Glymour and Quine

Although we have the notion of logical equivalence, one might want other criteria for theoretical equivalence. Logical equivalence is too strict to capture the sense in which some theories are equivalent. Indeed, as remarked above, theories can be logically equivalent only if they are formulated in the same signature. And there are many theories in different signatures that are nonetheless equivalent in some sense. For example, group theory can be formulated in different signatures.

Example 1. Let $\Sigma_1 = \{\cdot, e\}$ be a signature where $\cdot$ is a binary function symbol and e is a constant symbol. The theory of groups$_1$ is the following Σ_1-theory:

$$\big\{\forall x \forall y \forall z\big((x \cdot y) \cdot z = x \cdot (y \cdot z)\big), \forall x(x \cdot e = x \wedge e \cdot x = x),\\ \forall x \exists y(x \cdot y = e \wedge y \cdot x = e)\big\}$$

Group theory can also be formulated in the signature $\Sigma_2 = \{\cdot, -1\}$, where $\cdot$ is again a binary function symbol and -1 is a unary function symbol. The theory of groups$_2$ is the following Σ_2-theory:

$$\big\{\forall x \forall y \forall z\big((x \cdot y) \cdot z = x \cdot (y \cdot z)\big),\\ \exists x \forall y\big(y \cdot x = y \wedge x \cdot y = y \wedge y \cdot y^{-1} = x \wedge y^{-1} \cdot y = x\big)\big\}$$

One can easily see that these two theories are not logically equivalent. A model M of the theory of groups$_1$ is a set with a binary function $\cdot^M$ and a distinguished element e^M. A model N of the theory of groups$_2$ is a set with a binary function $\cdot^N$ and a unary function -1^N. These are not the same, so the theories do not have the same class of models. ⌟

If one wants to capture the sense in which these two formulations of group theory are equivalent, then one needs a more general criterion for theoretical equivalence than logical equivalence. Glymour and Quine's proposals are two candidates for such a criterion.

3.1 Glymour's criterion

Glymour (1970, 1977, 1980) proposed definitional equivalence as a criterion for theoretical equivalence.[3] 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 vocabulary that T_2 uses, and in a compatible way, T_2 can define all of the vocabulary that T_1 uses. In order to state this criterion carefully we need to do some work.

We first need to formalize the notion of a definition. Let $\Sigma \subset \Sigma^+$ be signatures and let $p \in \Sigma^+ - \Sigma$ be an n-ary predicate symbol. An **explicit definition of p in terms of $\mathbf{\Sigma}$** is a Σ^+-sentence of the form

$$\forall x_1 \dots \forall x_n \big(p(x_1, \dots, x_n) \leftrightarrow \phi(x_1, \dots, x_n)\big)$$

where $\phi(x_1, \dots, x_n)$ is a Σ-formula. Similarly, an explicit definition of an n-ary function symbol $f \in \Sigma^+ - \Sigma$ is a Σ^+-sentence of the form

$$\forall x_1 \dots \forall x_n \forall y \big(f(x_1, \dots, x_n) = y \leftrightarrow \phi(x_1, \dots, x_n, y)\big) \tag{1}$$

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

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

where $\phi(x_1, \dots, x_n, y)$ and $\psi(x)$ are both Σ-formulas.

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

$$\forall x_1 \dots \forall x_n \exists_{=1} y \phi(x_1, \dots, x_n, y)$$
$$\exists_{=1} x \psi(x)$$

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 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 $T \vDash \alpha_s$.

We now have the machinery necessary to state Glymour's criterion.

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 is a definitional extension T_1^+ of T_1 to the signature $\Sigma_1 \cup \Sigma_2$ and a definitional extension T_2^+ of T_2 to the signature $\Sigma_1 \cup \Sigma_2$ such that T_1^+ and T_2^+ are logically equivalent.

One often says that T_1 and T_2 are definitionally equivalent if they have a "common definitional extension."

Definitional equivalence captures a sense in which theories formulated in different signatures might nonetheless be theoretically equivalent. For example, although they are not logically equivalent, the theory of groups$_1$ and the theory of groups$_2$ are definitionally equivalent.

[3] Logicians were familiar with definitional equivalence before the 1970s, but Glymour was the first to introduce the notion into philosophy of science.

Example 2. Recall the two formulations of group theory from Example 1 and consider the following two $\Sigma_1 \cup \Sigma_2$-sentences.

$$\delta_{-1} := \forall x \forall y \big(x^{-1} = y \leftrightarrow (x \cdot y = e \wedge y \cdot x = e)\big)$$
$$\delta_e := \forall x \big(x = e \leftrightarrow \forall z (z \cdot x = z \wedge x \cdot z = z)\big)$$

The theory of groups$_1$ defines the unary function symbol -1 with the sentence δ_{-1} and the theory of groups$_2$ defines the constant symbol e with the sentence δ_e. One can verify that the theory of groups$_1$ satisfies the admissibility condition for δ_{-1} and that the theory of groups$_2$ satisfies the admissibility condition for δ_e. The theory of groups$_1 \cup \{\delta_{-1}\}$ and the theory of groups$_2 \cup \{\delta_e\}$ are logically equivalent. This implies that these two formulations of group theory are definitionally equivalent. ⌟

Definitional equivalence is well-known among logicians, and many results about it have been proven.[4] Here we state one particular fact that will be useful in what follows. Let $\Sigma \subset \Sigma^+$ be signatures. A Σ^+-theory T^+ is a **conservative extension** of a Σ-theory T if for every Σ-sentence ϕ, $T^+ \vDash \phi$ if and only if $T \vDash \phi$.

Proposition 1. *If T^+ is a definitional extension of T, then T^+ is a conservative extension of T (Hodges, 2008, 66).*

3.2 Quine's criterion

The criterion that Quine (1975) proposes is of a different flavor than Glymour's criterion. Quine suggests that two theories should be considered theoretically equivalent if there is a "suitable translation" between the theories. In order to state Quine's criterion carefully we again need to do some work.

We begin by introducing the idea of a reconstrual between signatures Σ_1 and Σ_2. A **reconstrual** F of Σ_1 into Σ_2 is a map from elements of the signature Σ_1 to Σ_2-formulas that satisfies the following three conditions.

- For every n-ary predicate symbol $p \in \Sigma_1$, $Fp(x_1, \ldots, x_n)$ is a Σ_2-formula with n free variables.
- For every n-ary function symbol $f \in \Sigma_1$, $Ff(x_1, \ldots, x_n, y)$ is a Σ_2-formula with $n + 1$ free variables.
- For every constant symbol $c \in \Sigma_1$, $Fc(y)$ is a Σ_2-formula with one free variable.

One can think of the Σ_2-formula $Fp(x_1, \ldots, x_n)$ as a "translation" of the Σ_1-formula $p(x_1, \ldots, x_n)$ into the signature Σ_2. Similarly, $Ff(x_1, \ldots, x_n, y)$ and $Fc(y)$ can be thought of as "translations" of the Σ_1-formulas $f(x_1, \ldots, x_n) = y$ and $c = y$, respectively. We will use the notation $F : \Sigma_1 \to \Sigma_2$ to denote a reconstrual F of Σ_1 into Σ_2.

Before stating Quine's criterion we need to take a moment to discuss reconstruals. The important fact about a reconstrual $F : \Sigma_1 \to \Sigma_2$ is that it naturally induces a map from Σ_1-formulas to Σ_2 formulas.

[4]For example, see (Hodges, 2008, 58–62), de Bouvére (1965), Kanger (1968), Pelletier and Urquhart (2003), Andréka et al. (2005), Friedman and Visser (2014) for some results.

In order to describe this map we first need to describe what F does to Σ_1-terms. F extends to a map from Σ_1-terms to Σ_2-formulas. Let $t(x_1, \ldots, x_n)$ be a Σ_1-term. We define the Σ_2-formula $Ft(x_1, \ldots, x_n, y)$ recursively as follows.

- If t is the variable x_i then $Ft(x_i, y)$ is the Σ_2-formula $x_i = y$.
- If t is the constant symbol $c \in \Sigma_1$ then $Ft(y)$ is the Σ_2-formula $Fc(y)$.
- Suppose that t is the term $f(t_1(x_1, \ldots, x_n), \ldots, t_k(x_1, \ldots, x_n))$ and that the Σ_2-formula $Ft_i(x_1, \ldots, x_n, y)$ has been defined for each $i = 1, \ldots, k$. Then $Ft(x_1, \ldots x_n, y)$ is the Σ_2-formula

  $$\exists z_1 \ldots \exists z_k \big(Ft_1(x_1, \ldots, x_n, z_1) \wedge \ldots \wedge Ft_k(x_1, \ldots, x_n, z_k) \wedge Ff(z_1, \ldots, z_k, y)\big)$$

Using this induced map from Σ_1-terms to Σ_2-formulas, we can describe how F maps Σ_1-formulas to Σ_2-formulas.[5] Let $\phi(x_1, \ldots, x_n)$ be a Σ_1-formula. We define the Σ_2-formula $F\phi(x_1, \ldots, x_n)$ recursively as follows.

- If $\phi(x_1, \ldots, x_n)$ is the Σ_1-atom $s(x_1, \ldots, x_n) = t(x_1, \ldots, x_n)$, with s and t Σ_1 terms, then $F\phi(x_1, \ldots, x_n)$ is the Σ_2-formula

  $$\exists z \big(Ft(x_1, \ldots, x_n, z) \wedge Fs(x_1, \ldots, x_n, z)\big)$$

- If $\phi(x_1, \ldots, x_n)$ is the Σ_1-atom $p(t_1(x_1, \ldots, x_n), \ldots, t_k(x_1, \ldots, x_n))$, with $p \in \Sigma_1$ a k-ary predicate symbol, then $F\phi(x_1, \ldots, x_n)$ is the Σ_2-formula

  $$\exists z_1 \ldots \exists z_k \big(Ft_1(x_1, \ldots, x_n, z_1) \wedge \ldots \wedge Ft_k(x_1, \ldots, x_n, z_k) \wedge Fp(z_1, \ldots, z_k)\big)$$

- Lastly, suppose that $F\phi(x_1, \ldots, x_n)$ and $F\psi(x_1, \ldots, x_n)$ have already been defined for Σ_1-formulas $\phi(x_1, \ldots, x_n)$ and $\psi(x_1, \ldots, x_n)$. Then we define $F\neg\phi$ to be $\neg F\phi$, $F(\phi \wedge \psi)$ to be $F(\phi) \wedge F(\psi)$, $F(\phi \to \psi)$ to be $F(\phi) \to F(\psi)$, $F(\phi \vee \psi)$ to be $F(\phi) \vee F(\psi)$, $F(\phi \leftrightarrow \psi)$ to be $F(\phi) \leftrightarrow F(\psi)$, $F(\forall x \phi)$ to be $\forall x F(\phi)$, and $F(\exists x \phi)$ to be $\exists x F(\phi)$.

In this way a reconstrual $F : \Sigma_1 \to \Sigma_2$ gives rise to a map between Σ_1-formulas and Σ_2-formulas. This map allows one to "translate" Σ_1-theories into Σ_2-theories. If T_1 is a Σ_1 theory and $F : \Sigma_1 \to \Sigma_2$ a reconstrual, then the Σ_2-theory $F(T_1)$ is defined by

$$F(T_1) = \{F(\phi) : \phi \in T_1\}$$

We now have the machinery necessary to state Quine's criterion for theoretical equivalence.[6]

[5] We are abusing notation by calling all of these maps F. In what follows, the important map is the map F from Σ_1-formulas to Σ_2-formulas. Whenever we refer to a reconstrual $F : \Sigma_1 \to \Sigma_2$ from now on, we will be referring to this induced map from Σ_1-formulas to Σ_2-formulas.

[6] Quine explains his proposal as follows: "By a *reconstrual* of the predicates of our language, accordingly, let me mean any mapping of our lexicon of predicates into our open sentences (n-place predicates to n-variable sentences). [...] I propose to individuate theories thus: two formulations express the same theory if they are empirically equivalent and there is a reconstrual of predicates that transforms the one theory into a logical equivalent of the other" (Quine, 1975, 320).

Definition. Let T_1 be a Σ_1-theory and T_2 a Σ_2-theory. T_1 is **Quine equivalent** to T_2 if there is a reconstrual $F : \Sigma_1 \to \Sigma_2$ such that the theories $F(T_1)$ and T_2 are logically equivalent.

At first glance, Quine equivalence seems to be a promising way to understand theoretical equivalence. And indeed, one can use it to capture a sense in which the theory of groups$_1$ is equivalent to the theory of groups$_2$.

Example 3. Recall again the theory of groups$_1$ and the theory of groups$_2$. We define a reconstrual $F : \Sigma_1 \to \Sigma_2$ by letting $F\cdot(x_1, x_2, y)$ be the Σ_2-formula $x_1 \cdot x_2 = y$ and $Fe(y)$ be the Σ_2-formula $\forall z(y \cdot z = z \wedge z \cdot y = z \wedge z \cdot z^{-1} = y \wedge z^{-1} \cdot z = y)$. One can then verify that the Σ_2-theory F(theory of groups$_1$) is logically equivalent to the theory of groups$_2$. The theory of groups$_1$ and the theory of groups$_2$ are therefore Quine equivalent. ⌟

3.3 Problems with Quine equivalence

Unlike definitional equivalence, Quine equivalence has not yet been investigated by logicians.[7] And upon investigation one finds that Quine equivalence is unsatisfactory. The following two examples illustrate some of its shortcomings.

These examples show that Quine equivalence is too liberal a criterion for theoretical equivalence. It considers some theories to be equivalent that one has good reason to consider inequivalent.

Example 4. Let $\Sigma = \{p, q\}$ be a signature with p and q both unary predicate symbols. Consider the following two Σ-theories.

$$T_1 = \{\exists_{=1}x(x = x), \forall x(p(x) \wedge q(x))\}$$
$$T_2 = \{\exists_{=1}x(x = x), \forall x p(x)\}$$

The theory T_2 is Quine equivalent to the theory T_1. We define the reconstrual $F : \Sigma \to \Sigma$ by letting $Fp(x)$ be $p(x) \wedge q(x)$ and letting $Fq(x)$ be $q(x)$. One can then easily verify that the following hold.

$$F\big(\exists_{=1}x(x = x)\big) \text{ is equivalent to } \exists_{=1}x(x = x)$$
$$F\big(\forall x(p(x)\big) \text{ is equivalent to } \forall x(p(x) \wedge q(x))$$

This implies that $F(T_2)$ is logically equivalent to T_1. ⌟

T_2 is Quine equivalent to T_1, and that is a strange verdict. In fact, one can make precise a sense in which T_1 and T_2 are inequivalent theories. We call a Σ-theory T **complete** if either $T \vDash \phi$ or $T \vDash \neg\phi$ for every Σ-sentence ϕ. Consider the theory T_1 from the above example. Every model M of T_1 has a domain with one element, and this one element is in both the sets p^M and q^M. So up to isomorphism, T_1 has a unique model. Since this model M satisfies either $M \vDash \phi$ or $M \vDash \neg\phi$ for every Σ-sentence ϕ, it must also be that $T \vDash \phi$ or $T \vDash \neg\phi$ for every Σ-sentence ϕ. The theory T_1 is therefore complete. But the theory T_2 is *not* complete. Consider the Σ-sentence $\forall x q(x)$. One can easily see that $T_2 \nvDash \forall x q(x)$ and $T_2 \nvDash \neg\forall x q(x)$. T_1 is Quine equivalent to T_2, but these two theories are inequivalent in a precise sense: T_1 is complete and T_2 is not.

[7] As of February 24, 2015, according to scholar.google.com, there have been no technical investigations of Quine's proposal.

Example 4 shows that completeness is not an "invariant" under Quine equivalence, and this might strike one as a shortcoming Quine equivalence. The following example provides another case where Quine equivalence makes a puzzling verdict.

Example 5. Let $\Sigma_1 = \emptyset$ and $\Sigma_2 = \{c, d\}$ be signatures with c and d constant symbols. Consider the Σ_1-theory $T_1 = \emptyset$ and the Σ_2-theory $T_2 = \{c = d\}$. The theory T_2 is Quine equivalent to the theory T_1. We define the reconstrual $G : \Sigma_2 \to \Sigma_1$ by letting both $Gc(y)$ and $Gd(y)$ be the Σ_1-formula $y = y$. One can then easily see that

$$G(c = d) \text{ is equivalent to } \exists z(z = z)$$

It follows that $G(T_2)$ is logically equivalent to T_1, so T_2 is Quine equivalent to T_1. ⌟

There are two things to notice about Example 5. First, note that Quine equivalence is again making a strange verdict. There is a sense in which T_1 and T_2 are inequivalent theories. The theory T_2 uses the constant symbols c and d to single out a preferred point in every model. The theory T_1 does not single out a preferred point in any model. For this reason one might not want to consider these two theories equivalent.

Second, note that although T_2 is Quine equivalent to T_1, T_1 is not Quine equivalent to T_2. There is a unique reconstrual $F : \Sigma_1 \to \Sigma_2$, the "empty reconstrual." The theory $F(T_1)$ is the empty theory in the signature Σ_2. And so $F(T_1)$ is not logically equivalent to T_2. This implies that Quine equivalence is not a symmetric relation on theories.[8]

Neither of these examples pose a problem for definitional equivalence. The two theories from Example 4 are not definitionally equivalent, and neither are the two theories from Example 5. Quine equivalence, however, makes puzzling verdicts in both of these cases.

4 Intertranslatability

One might wonder whether there is a way to amend Quine's original proposal so that it no longer makes these undesirable verdicts. The basic idea behind Quine equivalence is that two theories should count as theoretically equivalent if there exists a "suitable translation" between them. Definitional equivalence is often thought of as imposing a similar requirement. For example, Glymour (1970) remarks that definitional equivalence captures the idea that two theories are "intertranslatable."[9] In this final section we make this remark precise, and in doing so, we show that definitional equivalence can itself be understood as an amendment to Quine's original proposal.

The way that Quine explicates the notion of a "suitable translation" between theories makes Quine equivalence too liberal a criterion for theoretical equivalence. But one can be more restrictive about what counts as a "suitable

[8] Coffey (2014) argues that symmetry is not a good feature for a proposed criterion for theoretical equivalence to have. But both Coffey and Quine suggest that Quine equivalence is an equivalence relation. We have shown here that this is not the case.

[9] Knox (2013) and Coffey (2014) make this same remark.

translation" than Quine is. We call a reconstrual $F : \Sigma_1 \to \Sigma_2$ a **translation** of a Σ_1-theory T_1 into a Σ_2-theory T_2 if $T_1 \vDash \phi$ implies that $T_2 \vDash F\phi$ for all Σ_1-sentences ϕ. We will use the notation $F : T_1 \to T_2$ to denote a translation of T_1 into T_2.

We can then consider an amendment to Quine's criterion, which we call "intertranslatability."

Definition. Let T_1 be a Σ_1-theory and T_2 a Σ_2-theory. T_1 and T_2 are **intertranslatable** if there are translations $F : T_1 \to T_2$ and $G : T_2 \to T_1$ such that

$$T_1 \vDash \forall x_1 \dots \forall x_n \big(\phi(x_1, \dots, x_n) \leftrightarrow GF\phi(x_1, \dots, x_n)\big) \tag{3}$$

$$T_2 \vDash \forall x_1 \dots \forall x_n \big(\psi(x_1, \dots, x_n) \leftrightarrow FG\psi(x_1, \dots, x_n)\big) \tag{4}$$

for every Σ_1-formula $\phi(x_1, \dots, x_n)$ and every Σ_2-formula $\psi(x_1, \dots, x_n)$.

The conditions (3) and (4) can be thought of as requiring the translations $F : T_1 \to T_2$ and $G : T_2 \to T_1$ to be "almost inverse" to one another. Note, however, that F and G need not be literal inverses. The Σ_1-formula $GF\phi(x_1, \dots, x_n)$ is not required to be *equal* to the Σ_1-formula $\phi(x_1, \dots, x_n)$. Rather, these two formulas are merely required to be *equivalent* according to the theory T_1.

Like Quine's original proposal, intertranslatability judges theories to be equivalent when "suitable translations" exist between them. It is just more restrictive about what should count as a "suitable translation." Our goal in the remainder of this section is to show how intertranslatability relates to definitional equivalence. The following theorem provides a partial answer to this question.[10]

Theorem 1. *If T_1 and T_2 are definitionally equivalent, then they are intertranslatable.*

Proof. Suppose that T is a common definitional extension of a Σ_1-theory T_1 and a Σ_2-theory T_2. We define a reconstrual $F : \Sigma_1 \to \Sigma_2$ as follows. Let $p \in \Sigma_1$ be an n-ary predicate symbol. Then since T is a definitional extension of T_2,

$$T \vDash \forall x_1 \dots \forall x_n \big(p(x_1, \dots, x_n) \leftrightarrow \delta_p(x_1, \dots, x_n)\big)$$

for some Σ_2-formula $\delta_p(x_1, \dots, x_n)$. We define $Fp(x_1, \dots, x_n)$ to be the Σ_2-formula $\delta_p(x_1, \dots, x_n)$. Let $f \in \Sigma_1$ be an n-ary function symbol. We define $Ff(x_1, \dots, x_n, y)$ to be $\delta_f(x_1, \dots, x_n, y)$, where δ_f is the formula that T uses to define the function symbol f. Lastly, let $c \in \Sigma_1$ be a constant symbol. We again define $Fc(y)$ to be $\delta_c(y)$. A reconstrual $G : \Sigma_2 \to \Sigma_1$ is defined in the analogous way. One can then verify by induction on complexity that

$$T \vDash \forall x_1 \dots \forall x_n (\phi(x_1, \dots, x_n) \leftrightarrow F\phi(x_1, \dots, x_n)) \tag{5}$$

$$T \vDash \forall x_1 \dots \forall x_n (\psi(x_1, \dots, x_n) \leftrightarrow G\psi(x_1, \dots, x_n)) \tag{6}$$

[10] Glymour (1970) remarks that definitional equivalence "guarantees that all and only theorems of $[T_1]$ are translated as theorems of $[T_2]$, and conversely" (Glymour, 1970, 279). Here we provide a strengthening of Glymour's remark. Theorems 1 and 2 make precise a sense in which this requirement is no stronger and no weaker than definitional equivalence. Friedman and Visser (2014) and Pinter (1978) state these two results, but do not provide proofs. Ingredients for proofs using tools of category theory are contained in Visser (2006). Further ingredients are contained in Hoehnke (1966). Pelletier and Urquhart (2003) provide proofs for the special case of propositional logic. Here we extend the results to full first-order logic using only elementary methods.

for every Σ_1-formula $\phi(x_1, \dots, x_n)$ and every Σ_2-formula $\psi(x_1, \dots, x_n)$.

We need to show that $F : \Sigma_1 \to \Sigma_2$ and $G : \Sigma_2 \to \Sigma_1$ are translations. Without loss of generality we show that F is. Suppose that $T_1 \vDash \phi$ for some Σ_1-sentence ϕ. Then by equation (5) above, $T \vDash \phi \leftrightarrow F\phi$. Proposition 1 guarantees that $T \vDash \phi$, so it must be that $T \vDash F\phi$. Since $F\phi$ is a Σ_2-sentence, Proposition 1 then implies that $T_2 \vDash F\phi$. So $F : T_1 \to T_2$ is a translation.

Now let $\phi(x_1, \dots, x_n)$ be any Σ_1-formula. Equations (5) and (6) together imply that $T \vDash \forall x_1 \dots \forall x_n(\phi(x_1, \dots, x_n) \leftrightarrow GF\phi(x_1, \dots, x_n))$. Since T is a conservative extension of T_1 by Proposition 1, it must be that

$$T_1 \vDash \forall x_1 \dots \forall x_n \big(\phi(x_1, \dots, x_n) \leftrightarrow GF\phi(x_1, \dots, x_n)\big)$$

A similar argument yields $T_2 \vDash \forall x_1 \dots \forall x_n(\psi(x_1, \dots, x_n) \leftrightarrow FG\psi(x_1, \dots, x_n))$ for every Σ_2-formula $\psi(x_1, \dots, x_n)$. □

This result immediately provides us with many examples of theories that are intertranslatable. For example, we have already seen that the theory of groups$_1$ and the theory of groups$_2$ are definitionally equivalent. Theorem 1 implies that they are intertranslatable too.

The converse of Theorem 1, however, does not hold. The following example provides a case of theories that are intertranslatable but not definitionally equivalent.

Example 6. Let $\Sigma = \{p\}$ be the signature containing a unary predicate symbol p. Consider the following two Σ-theories.

$$T_1 = \{\exists_{=1} x(x = x), \forall x p(x)\}$$
$$T_2 = \{\exists_{=1} x(x = x), \neg\forall x p(x)\}$$

T_1 and T_2 are not definitionally equivalent since they do not have a common conservative extension. But T_1 and T_2 are intertranslatable. Consider the reconstrual $F : \Sigma \to \Sigma$ defined by letting $Fp(x)$ be $\neg p(x)$. One can verify that $F : T_1 \to T_2$ is a translation and that both

$$T_1 \vDash \forall x_1 \dots \forall x_n \big(\phi(x_1, \dots, x_n) \leftrightarrow FF\phi(x_1, \dots, x_n)\big)$$
$$T_2 \vDash \forall x_1 \dots \forall x_n \big(\phi(x_1, \dots, x_n) \leftrightarrow FF\phi(x_1, \dots, x_n)\big)$$

hold for every Σ-formula $\phi(x_1, \dots, x_n)$. This implies that T_1 and T_2 are intertranslatable. ⌟

Although the converse of Theorem 1 is not true in general, the following theorem establishes that it *is* true when one only considers theories in disjoint signatures.

Theorem 2. *Let Σ_1 and Σ_2 be disjoint signatures with T_1 a Σ_1-theory and T_2 a Σ_2-theory. If T_1 and T_2 are intertranslatable, then they are definitionally equivalent.*

Before proving Theorem 2 we need to do some work. Consider a translation $F : T_1 \to T_2$ between a Σ_1-theory T_1 and a Σ_2-theory T_2. The translation F gives rise to a map $F^* : \mathrm{Mod}(T_2) \to \mathrm{Mod}(T_1)$, which takes models of the theory T_2 to models of the theory T_1.[11] For every model A of T_2 we first define a Σ_1-structure $F^*(A)$ as follows.

[11] One can compare this with Gajda et al. (1987).

- $\mathrm{dom}(F^*(A)) = \mathrm{dom}(A)$.
- $(a_1, \ldots, a_n) \in p^{F^*(A)}$ if and only if $A \vDash Fp[a_1, \ldots, a_n]$.
- $f^{F^*(A)}(a_1, \ldots, a_n) = b$ if and only if $A \vDash Ff[a_1, \ldots, a_n, b]$.
- $c^{F^*(A)} = a$ if and only if $A \vDash Fc[a]$.

One needs to verify that $f^{F^*(A)}$ and $c^{F^*(A)}$ are well-defined. For example, in the first case one needs to check that $f^{F^*(A)}$ is a function. We trivially know that $T_1 \vDash \forall x_1 \ldots \forall x_n \exists_{=1} y f(x_1, \ldots, x_n) = y$. Since F is an interpretation, this implies that

$$T_2 \vDash F\big(\forall x_1 \ldots \forall x_n \exists_{=1} y f(x_1, \ldots, x_n) = y\big)$$

Unraveling the sentence on the right hand side, we see that this means that for all $a_1, \ldots, a_n \in F^*(A)$ there is a unique $b \in A$ such that $A \vDash Ff[a_1, \ldots, a_n, b]$. So $f^{F^*(A)}$ is indeed a function. One goes through a similar argument to show that the element $c^{F^*(A)}$ is well-defined.

For every model A of T_2 we have defined a Σ_1-structure $F^*(A)$. In order to show that $F^*(A)$ is actually a model of T_1 we need the following lemma.[12]

Lemma. *Let A be a model of T_2 and $\phi(x_1, \ldots, x_n)$ a Σ_1-formula. Then $A \vDash F\phi[a_1, \ldots, a_n]$ if and only if $F^*(A) \vDash \phi[a_1, \ldots, a_n]$.*

Proof. By induction on the complexity of ϕ. □

Using the Lemma we can verify that for every model A of T_2, $F^*(A)$ is a model of T_1. We let $\phi \in T_1$ be an axiom of T_1 and we show that $F^*(A) \vDash \phi$. Since F is a translation, it must be that $T_2 \vDash F\phi$, which means that $A \vDash F\phi$ since A is a model of T_2. Then the Lemma implies that $F^*(A) \vDash \phi$. So indeed, $F^*(A)$ is a model of T_1 and we have successfully defined a map $F^* : \mathrm{Mod}(T_2) \to \mathrm{Mod}(T_1)$ between models of T_2 and models of T_1.

We conclude our discussion of intertranslatability and definitional equivalence with a proof of Theorem 2.

Proof of Theorem 2. Suppose that T_1 and T_2 are intertranslatable, with $F : T_1 \to T_2$ and $G : T_2 \to T_1$ the relevant translations. We begin by defining definitional extensions T_1^+ and T_2^+ of T_1 and T_2 to the signature $\Sigma_1 \cup \Sigma_2$.

We define $T_1^+ = T_1 \cup \{\delta_s : s \in \Sigma_2\}$, where for each symbol $s \in \Sigma_2$ the Σ_2-sentence δ_s is an explicit definition of s. If $q \in \Sigma_2$ is an n-ary predicate symbol then we let the definition δ_q be the sentence

$$\forall x_1 \ldots \forall x_n \big(q(x_1, \ldots, x_n) \leftrightarrow Gq(x_1, \ldots, x_n)\big)$$

If $g \in \Sigma_2$ is an n-ary function symbol then we let the definition δ_g be the sentence $\forall x_1 \ldots \forall x_n \forall y (g(x_1, \ldots, x_n) = y \leftrightarrow Gg(x_1, \ldots, x_n, y))$. And if $d \in \Sigma_2$ is a constant symbol then we let δ_d be the sentence $\forall y(d = y \leftrightarrow Gd(y))$. Using the Lemma one can verify that T_1 satisfies the admissibility conditions for δ_g and δ_d.

We define $T_2^+ = T_2 \cup \{\delta_t : t \in \Sigma_1\}$ in the same manner. If $p \in \Sigma_1$ is an n-ary predicate symbol then we let the definition δ_p be the sentence

$$\forall x_1 \ldots \forall x_n \big(p(x_1, \ldots, x_n) \leftrightarrow Fp(x_1, \ldots, x_n)\big)$$

[12] See Pinter (1978, 2.4) for a similar lemma.

If $f \in \Sigma_1$ is an n-ary function symbol then we let the definition δ_f be the sentence $\forall x_1 \dots \forall x_n \forall y(f(x_1, \dots, x_n) = y \leftrightarrow Ff(x_1, \dots, x_n, y))$. If $c \in \Sigma_1$ is a constant symbol then we let δ_c be the sentence $\forall y(c = y \leftrightarrow Fc(y))$. Using the above Lemma one can verify that T_2 satisfies the admissibility conditions for δ_f and δ_c.

We show that T_1^+ and T_2^+ are logically equivalent. Without loss of generality, we show that every model of T_2^+ is a model of T_1^+. The converse follows via an analogous argument. Let A be a model of T_2^+. We show that A is a model of T_1^+. There are two cases that need checking.

First, we show that $A \vDash \phi$ for every $\phi \in T_1$. We know that $F^*(A)$ is a model of T_1, so $F^*(A) \vDash \phi$. By the Lemma this means that $A \vDash F\phi$. One can verify by induction on the complexity of ψ that for every model A of T_2^+ and every Σ_1-formula $\psi(x_1, \dots, x_n)$,

$$A \vDash \psi[a_1, \dots, a_n] \text{ iff } A \vDash F\psi[a_1, \dots, a_n]. \tag{7}$$

In particular, (7) implies that $A \vDash \phi$.

Second, we show that $A \vDash \delta_s$ for every $s \in \Sigma_2$. Let $q \in \Sigma_2$ be an n-ary predicate symbol. We show that $A \vDash \delta_q$. This follows since for all $a_1, \dots, a_n \in A$

$$A \vDash q[a_1, \dots, a_n] \text{ iff } A \vDash FGq[a_1, \dots, a_n] \text{ iff } A \vDash Gq[a_1, \dots, a_n]$$

The first equivalence follows from condition (4) in the definition of intertranslatability and the fact that A is a model of T_2^+. The second equivalence follows from (7). This string of equivalences implies that $A \vDash \delta_q$. In a similar manner one can verify that $A \vDash \delta_g$ for every function symbol $g \in \Sigma_2$ and that $A \vDash \delta_d$ for every constant symbol $d \in \Sigma_2$.

We have therefore shown that A is a model of T_2^+. We conclude that T_1^+ and T_2^+ are logically equivalent, so T_1 and T_2 are definitionally equivalent. □

5 Conclusion

It is important to note that intertranslatability does not suffer from the same shortcomings as Quine's original criterion. Theorem 2 implies that the two theories from Example 5 are not intertranslatable. Example 4 does not pose a problem for intertranslatability either. One can easily verify that if T_1 and T_2 are intertranslatable, then T_1 is complete if and only if T_2 is complete. So the two theories in Example 4 are not intertranslatable.

Quine originally considered theoretical equivalence because of its relationship to underdetermination.[13] Our discussion of Quine equivalence yields a small remark concerning this relationship. A theory T is underdetermined if there is another theory T' such that (i) T and T' are empirically equivalent but (ii) T and T' are theoretically *in*equivalent. Quine proposed Quine equivalence as a way to make condition (ii) precise. We have shown here, however, that it is too liberal a criterion for theoretical equivalence. According to Quine equivalence, there are few pairs of theories that satisfy (ii), and so there are few cases of underdetermination. If one takes Quine equivalence as the standard for theoretical equivalence, one underestimates the threat of underdetermination.

[13] See Quine (1975, 318–322) and Coffey (2014).

Our discussion also allows one to evaluate Quine's criterion against Glymour's. Theorems 1 and 2 demonstrate a natural and precise sense in which definitional equivalence is an amendment to Quine equivalence. Quine's original criterion for theoretical equivalence was unsatisfactory, and when it is fixed it becomes essentially the same as Glymour's criterion.*

References

Andréka, H., Madarász, J. X., and Németi, I. (2005). Mutual definability does not imply definitional equivalence, a simple example. *Mathematical Logic Quarterly.*

Barrett, T. W. (2014). On the structure of classical mechanics. *The British Journal for the Philosophy of Science (forthcoming).*

Coffey, K. (2014). Theoretical equivalence as interpretive equivalence. *The British Journal for the Philosophy of Science.*

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

de Bouvére, K. L. (1965). Synonymous theories. In *Symposium on the Theory of Models.* North-Holland Publishing Company.

Friedman, H. and Visser, A. (2014). When bi-interpretability implies synonymy. *Manuscript.*

Gajda, A., Krynicki, M., and Szczerba, L. (1987). A note on syntactical and semantical functions. *Studia Logica.*

Glymour, C. (1970). Theoretical realism and theoretical equivalence. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association.*

Glymour, C. (1977). The epistemology of geometry. *Nous.*

Glymour, C. (1980). *Theory and Evidence.* Princeton.

Glymour, C. (2013). Theoretical equivalence and the semantic view of theories. *Philosophy of Science.*

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

Halvorson, H. (2012). What scientific theories could not be. *Philosophy of Science.*

Halvorson, H. (2013). The semantic view, if plausible, is syntactic. *Philosophy of Science.*

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

Hoehnke, H. (1966). Zur strukturgleichheit axiomatischer klassen. *Mathematical Logic Quarterly.*

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

Kanger, S. (1968). Equivalent theories. *Theoria*.

Knox, E. (2013). Newtonian spacetime structure in light of the equivalence principle. *The British Journal for the Philosophy of Science*.

North, J. (2009). The 'structure' of physics: A case study. *Journal of Philosophy*.

Pelletier, F. J. and Urquhart, A. (2003). Synonymous logics. *Journal of Philosophical Logic*.

Pinter, C. C. (1978). Properties preserved under definitional equivalence and interpretations. *Mathematical Logic Quarterly*.

Quine, W. V. O. (1975). On empirically equivalent systems of the world. *Erkenntnis*.

Sklar, L. (1982). Saving the noumena. *Philosophical Topics*.

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

van Fraassen, B. C. (2014). One or two gentle remarks about Hans Halvorson's critique of the semantic view. *Manuscript*.

Visser, A. (2006). Categories of theories and interpretations. In *Logic in Tehran. Proceedings of the workshop and conference on Logic, Algebra and Arithmetic, held October 18–22, 2003*. ASL.

Weatherall, J. O. (2014). Are Newtonian gravitation and geometrized Newtonian gravitation theoretically equivalent? *Manuscript*.