Erkenntnis
https://doi.org/10.1007/s10670-022-00609-8

ORIGINAL RESEARCH

Quine's Underdetermination Thesis

Eric Johannesson[1]

Received: 10 February 2022 / Accepted: 10 August 2022

Abstract

In *On Empirically Equivalent Systems of the World* from 1975, Quine formulated a thesis of underdetermination roughly to the effect that every scientific theory has an empirically equivalent but logically incompatible rival, one that cannot be discarded merely as a terminological variant of the former. For Quine, the truth of this thesis was an open question. If true, some would argue that it undermines any belief in scientific theories that is based purely on their empirical success. But despite its potential significance, surprisingly little has been done by way of establishing or refuting it. My aim is to establish the thesis for as large a class of theories as possible. Under various precisifications of the concepts involved, I show that it holds for all consistent and recursively axiomatizable theories that postulate infinitely many theoretical entities.

1 Introduction

Quine (1975) formulated a thesis of underdetermination, roughly saying that

(UT) Every scientific theory has an empirically equivalent but logically incompatible rival, one that cannot be discarded merely as a terminological variant of the former.

For Quine, the truth of this thesis was an open question. If true, some would argue that it undermines any belief in scientific theories that is based purely on their empirical success.[1] But despite its potential significance, surprisingly little has been

[1] I am referring to what I take to be a folklore argument for anti-realism in the philosophy of science, namely *The underdetermination argument* (see, for instance, Okasha (2002, pp. 71–76)). The underlying observation is that, since – by definition – empirically equivalent theories always enjoy the same amount of empirical success, deciding between them purely on that basis is impossible. Although, in practice, theory choice may be influenced by other theoretical virtues (such as simplicity), it can be argued that these virtues do not carry any epistemic or justificatory weight.

Eric Johannesson
johannesson.eric@gmail.com

[1] Department of Philosophy, Stockholm University, Stockholm, Sweden

Published online: 26 August 2022

done by way of establishing or refuting it. My aim is to establish the thesis for as large a class of theories as possible.

I will make the idealizing assumption that a theory is a set of sentences of a first-order single-sorted language without function symbols whose predicates can be partitioned into an empirical and a theoretical part.[2] Relative to such a partition, various notions of *empirical equivalence* can be defined, corresponding to proposals in the literature. In Sect. 2, these notions are presented and ordered by logical strength. In Sect. 3, I briefly investigate some necessary and sufficient conditions for finding empirically equivalent but logically incompatible rivals to a given theory, and then introduce the problem of saying when such a rival is to be regarded as a terminological variant of the former, also known as the problem of *theoretical equivalence*. In Sect. 4, I present various solutions to this problem from the literature and order them by logical strength. In Sect. 5, I show that, under any combination of said notions of empirical and theoretical equivalence, Quine's thesis applies to all consistent and recursively axiomatizable theories that postulate infinitely many theoretical entities.

2 Empirical Equivalence

According to what Laudan and Leplin (1991) and Worrall (2011) refer to as the *traditional* notion of empirical equivalence, two theories are empirically equivalent just in case they entail the same empirical sentences, where an empirical sentence is a sentence containing only empirical predicates. Although I have never been able to find any explicit endorsement of this notion in the literature, I believe it is implicit in the writings of Carnap and Quine. This syntactic notion of empirical equivalence was criticized by Van Fraassen (1980), on the grounds that it does not apply to theories that, intuitively speaking, make the same claims about the observable part of the world. Indeed, granted that 'observable' may be considered an empirical predicate, the traditional notion classifies 'some things are not observable' as an empirical claim which, intuitively, it is not. Instead, Van Fraassen suggested a semantic notion on which two theories are empirically equivalent just in case, for every model of one of them, there is a model of the other whose observable substructure is isomorphic to the observable substructure of the former. In and of itself, however, Van Fraassen's objection does not target syntactic notions of empirical equivalence as such. In response, Turney (1990) proposed another syntactic notion of empirical equivalence, immune to said objection, on which two theories are empirically equivalent just in case they entail the same *qualified* empirical sentences, where a qualified empirical sentence is an empirical sentence whose quantifiers are restricted by an observability predicate. This is also how empirical sentences are defined in Schurz (2013, pp. 107–108).

With L being the set of empirical predicates, and $\delta(x)$ being an observability predicate, these three notions correspond to what I will call *syntactic L-equivalence*,

[2] For an excellent defense of the distinction between empirical and theoretical concepts, see Schurz (2013, pp. 63–75).

semantic L-equivalence over δ, and *syntactic L-equivalence over* δ, respectively. A fourth notion, *semantic L-equivalence*, will also be defined, although it does not correspond to anything suggested in the literature (as far as I know).

Two theories are said to be *syntactically* equivalent just in case they prove the same sentences, and *semantically* equivalent just in case they are satisfied by the same models. By soundness and completeness, these notions coincide in the case of classical first-order logic. In either of those senses, non-equivalent theories may still be equivalent with respect to a limited range of objects and their properties. Such theories may agree about the distribution of *certain* properties and relations among objects, or about the distribution of properties and relations among *certain* objects, or about the distribution of *certain* properties and relations among *certain* objects. As we shall see, when equivalence is limited to certain objects and properties, the syntactic and semantic notions come apart.

In order to make this claim precise, we need to introduce some standard definitions, namely those of *relativization*, *reduct*, and *part*:

Definition 2.1 (Relativization) Let $\delta(x)$ be a formula.[3] For any formula φ, its δ-relativization (written $[\varphi]_\delta$) is defined recursively:

(i) $[P\bar{x}]_\delta = P\bar{x}$.
(ii) $[\neg\varphi]_\delta = \neg[\varphi]_\delta$.
(iii) $[\varphi \to \psi]_\delta = [\varphi]_\delta \to [\psi]_\delta$.
(iv) $[\forall x\varphi]_\delta = \forall x(\delta(x) \to [\varphi]_\delta)$.

Remark 2.1 We may regard the other standard operators as defined in terms of $\neg$, $\to$, and $\forall$. In particular, with $\exists x\varphi$ defined as $\neg\forall x\neg\varphi$, it follows that $[\exists x\varphi]_\delta = [\neg\forall x\neg\varphi]_\delta = \neg[\forall x\neg\varphi]_\delta = \neg\forall x(\delta(x) \to [\neg\varphi]_\delta) = \neg\forall x(\delta(x) \to \neg[\varphi]_\delta)$, which is equivalent to $\exists x(\delta(x) \wedge [\varphi]_\delta)$.

Remark 2.2 When there is no risk of ambiguity, we shall write φ_δ instead of $[\varphi]_\delta$.

Intuitively speaking, δ-relativized sentences only talk about objects satisfying δ. This claim is made precise by Lemma 2.2 below.

Definition 2.2 (Reducts and parts) Let $L \subseteq L'$ be vocabularies, let $\delta(x)$ be an L'-formula, and let $\mathcal{M}$ be an L'-model.

(i) The *L-reduct of* $\mathcal{M}$ (written $\mathcal{M}|L$) is the L-model with the same domain as $\mathcal{M}$ such that, for any predicate $P \in L$, $P^{\mathcal{M}|L} = P^{\mathcal{M}}$.
(ii) Provided that $\mathcal{M} \vDash \exists x\delta$, the *$\delta$-part of* $\mathcal{M}$ (written $\mathcal{M}_\delta$) is the L'-model whose domain D consist of all objects satisfying δ in $\mathcal{M}$ and such that, for any n-place predicate $P \in L'$, $P^{\mathcal{M}_\delta} = P^{\mathcal{M}} \cap D^n$.

[3] As is customary, when we say 'let $\delta(x)$ be a formula', we mean 'let δ be a formula with at most one free variable x'. To avoid cluttering, we will thenceforth usually refer to the formula simply as 'δ'.

It is easy to establish that the L-reduct of a model satisfies an L-sentence just in case the model satisfies it, and that the δ-part of a model satisfies a sentence just in case the model satisfies its δ-relativization:

Lemma 2.1 *Let $L \subseteq L'$ be vocabularies, let $\delta(x)$ be an L'-formula, let $\mathcal{M}$ be an L'-model such that $\mathcal{M} \vDash \exists x\delta$, and let φ be an L-sentence. Then we have $\mathcal{M}_\delta | L \vDash \varphi$ iff $\mathcal{M} \vDash \varphi_\delta$.*

Proof Let X be the set of variables, let D be the set of objects satisfying δ in $\mathcal{M}$, and let φ be an L-formula. It is straightforward to show, by induction on the complexity of φ, that for any variable assignment $v : X \to D$, we have $\mathcal{M}_\delta | L, v \vDash \varphi$ iff $\mathcal{M}, v \vDash \varphi_\delta$. The atomic cases are obvious, since relativization do not affect them, and the cases of boolean operators are straightforward. In the case of universal quantification, we have $\mathcal{M}_\delta | L, v \vDash \forall x\varphi$ iff, for all $a \in D$, $\mathcal{M}_\delta | L, v(a/x) \vDash \varphi$ iff (by induction hypothesis), for all $a \in D$, $\mathcal{M}, v(a/x) \vDash \varphi_\delta$ iff, for all $a \in |\mathcal{M}|$, $\mathcal{M}, v(a/x) \vDash \delta \to \varphi_\delta$ iff $\mathcal{M}, v \vDash \forall x(\delta \to \varphi_\delta)$ iff $\mathcal{M}, v \vDash [\forall x\varphi]_\delta$. It now follows that, for any L-sentence φ, we have $\mathcal{M}_\delta | L \vDash \varphi$ iff $\mathcal{M} \vDash \varphi_\delta$. □

It follows that models with identical L-reducts satisfy the same L-sentences, and that models with identical δ-parts satisfy the same δ-relativized sentences:

Lemma 2.2 *Let $L \subseteq L'$ be vocabularies, let $\delta(x)$ be an L'-formula, and let $\mathcal{M}$ and $\mathcal{M}'$ be L'-models such that $\mathcal{M}, \mathcal{M}' \vDash \exists x\delta$ and $\mathcal{M}_\delta | L = \mathcal{M}'_\delta | L$. Then, for any L-sentence φ, $\mathcal{M} \vDash \varphi_\delta$ iff $\mathcal{M}' \vDash \varphi_\delta$.*

Proof We have $\mathcal{M} \vDash \varphi_\delta$ iff (by Lemma 2.1) $\mathcal{M}_\delta | L \vDash \varphi$ iff (by assumption) $\mathcal{M}'_\delta | L \vDash \varphi$ iff (by Lemma 2.1) $\mathcal{M}' \vDash \varphi_\delta$. □

We shall say that two theories are *syntactically L-equivalent (over δ)* just in case they entail the same (δ-relativized) L-sentences, and *semantically L-equivalent (over δ)* just in case the models satisfying them have the same (δ-restricted) L-reducts. More precisely:

Definition 2.3 (Syntactic and semantic equivalence) Let T_1 and T_2 be theories in L_1 and L_2, respectively, and let $L \subseteq L_1 \cap L_2$. We say that

1. T_1 and T_2 are *syntactically L-equivalent* just in case, for any L-sentence φ, $T_1 \vdash \varphi$ iff $T_2 \vdash \varphi$.
2. T_1 and T_2 are *semantically L-equivalent* just in case $\{\mathcal{M}|L : \mathcal{M} \vDash T_1\} = \{\mathcal{M}|L : \mathcal{M} \vDash T_2\}$.

Moreover, provided that $\delta(x)$ is an L-formula such that $T_1 \vdash \exists x\delta$ and $T_2 \vdash \exists x\delta$, we say that

$$\begin{tikzcd}
& \text{Syntactic } L\text{-equivalence over } \delta & \\
\text{Syntactic } L\text{-equivalence} \arrow[ur] & & \text{Semantic } L\text{-equivalence over } \delta \arrow[ul] \\
& \text{Semantic } L\text{-equivalence} \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

Fig. 1 The relation of entailment between the four notions of empirical equivalence, where L is the set of empirical predicates, and $\delta(x)$ is a formula applying to all and only empirical entities

1. T_1 and T_2 are *syntactically L-equivalent over* δ just in case, for any L-sentence φ, $T_1 \vdash \varphi_\delta$ iff $T_2 \vdash \varphi_\delta$.
2. T_1 and T_2 are *semantically L-equivalent over* δ just in case $\{\mathcal{M}_\delta|L : \mathcal{M} \vDash T_1\} = \{\mathcal{M}_\delta|L : \mathcal{M} \vDash T_2\}$.

The relation of entailment between these four notions is given by the following facts, summarized in Fig. 1:

Fact 2.1 *Syntactic/semantic L-equivalence entails syntactic/semantic L-equivalence over* δ.

Proof In the syntactic case, the result follows immediately by $\delta(x)$ being an L-formula. In the semantic case, assume that semantic L-equivalence obtains between T_1 and T_2. Let $\mathcal{M}$ be an L-model, and suppose that there is an L_1-model $\mathcal{M}'$ of T_1 with $\mathcal{M}'_\delta|L = \mathcal{M}$. Since $\mathcal{M}'|L$ is an L-model, and $\mathcal{M}'$ is a model of T_1, it follows by assumption that there is an L_2-model $\mathcal{M}^*$ of T_2 with $\mathcal{M}^*|L = \mathcal{M}'|L$. Furthermore, $\mathcal{M}^*_\delta|L = \mathcal{M}'_\delta|L = \mathcal{M}$. The other direction is symmetrical. □

Fact 2.2 *Semantic L-equivalence* (*over* δ) *entails syntactic L-equivalence* (*over* δ).

Proof Assume semantic L-equivalence over δ between T_1 and T_2. Let φ be an L-sentence such that $T_1 \vdash \varphi_\delta$. Let $\mathcal{M}$ be a model of T_2. Since $\mathcal{M}_\delta|L$ is an L-model, it follows by assumption that there is a model $\mathcal{M}'$ of T_1 with $\mathcal{M}'_\delta|L = \mathcal{M}_\delta|L$. Since $\mathcal{M}' \vDash \varphi_\delta$, it follows by Lemma 2.2 that $\mathcal{M} \vDash \varphi_\delta$. Hence, $T_2 \vdash \varphi_\delta$. The other direction is symmetrical. □

Fact 2.3 *Semantic L-equivalence over* δ *does not entail syntactic L-equivalence*, *and vice versa*.

Proof For left to right, let $T_1 = \{\forall x Px\}$ and $T_2 = \{\exists x Px\}$. Clearly, T_1 and T_2 are semantically L-equivalent over Px, but not syntactically L-equivalent. For the other direction, let P and Q be unary predicates and R binary, let $L = L_1 = \{P, Q\}$ and $L_2 = \{P, Q, R\}$, and let $\delta(x)$ be $x = x$. Let T_2 be a theory saying that R is a bijective relation between the P:s and the Q:s, namely

$$T_2 = \{\forall x(Px \to \exists! y(Qy \wedge Rxy)), \forall y(Qy \to \exists! x(Px \wedge Rxy))\}$$

and let T_1 be the set of all L-consequences of T_2. Clearly, T_1 and T_2 are syntactically L-equivalent. Let $\mathcal{M}$ be an L-model where $P^{\mathcal{M}}$ is countably infinite and $Q^{\mathcal{M}}$ is uncountable, and let T be the set of L-sentences true in $\mathcal{M}$. By Löwenheim-Skolem, T has a countable model $\mathcal{M}'$, one in which both $P^{\mathcal{M}'}$ and $Q^{\mathcal{M}'}$ are countably infinite. Since $\mathcal{M}'$ can be expanded to a model of T_2, $\mathcal{M}'$ is a model of T_1. And since $T_1 \subseteq T$, so is $\mathcal{M}$. But $\mathcal{M}$ cannot be expanded to a model of T_2. Hence, T_1 and T_2 are not semantically L-equivalent over δ. □

Other examples highlighting the distinction between syntactic and semantic equivalence have been offered by van Benthem (1978, p. 324), Melia (2000, pp. 459–461), Ketland (2004, p. 297), and Johannesson (2020, pp. 492–493).

3 Constructing Empirically Equivalent Rivals

Without committing to any particular notion of empirical equivalence, let us stipulate that a theory T has the **underdetermination property** just in case there is a theory T' such that (i) T and T' are empirically equivalent, and (ii) T and T' are jointly inconsistent. In that case, we shall say that T and T' are *empirically equivalent rivals*.

By Craig's interpolation theorem, in order for T and T' to be jointly inconsistent, there has to be a sentence φ in their common vocabulary such that $T \vdash \varphi$ and $T' \vdash \neg\varphi$. If, moreover, T and T' are syntactically L-equivalent (over δ) and consistent on their own, it follows that φ cannot be derived from the set of (δ-relativized) L-sentence that are theorems of T. In this sense, we may say that φ is a *proper* non-empirical consequence of T.[4] Thus, for a consistent theory to have the underdetermination property, it has to have some proper non-empirical consequences. Having such consequences, however, is not sufficient for having the underdetermination property. To construct a counterexample, let Γ be a consistent but incomplete set of (δ-relativized) L-sentences, with $\Gamma \nvdash \sigma$ and $\Gamma \nvdash \neg\sigma$ for some (δ-relativized) L-sentence σ. Let $P \notin L$ be a theoretical predicate, and consider the theory $T = \Gamma \cup \{\sigma \to \exists\bar{x}P\bar{x}\}$. It follows that $\sigma \to \exists\bar{x}P\bar{x}$ is a proper non-empirical consequence of T.[5] Since, by assumption, Γ has a model where σ is false and therefore $\sigma \to \exists\bar{x}P\bar{x}$ is true, it also

[4] Since, for any theoretical predicate P, we have $\vdash \exists\bar{x}P\bar{x} \vee \neg\exists\bar{x}P\bar{x}$, every theory has non-empirical consequences that are not proper.

[5] To see why, let Δ be the set of all empirical consequences of T. Since Γ has a model where both σ and $\exists\bar{x}P\bar{x}$ are true, it follows that $T \nvdash \neg\sigma$, and thus $\Delta \nvdash \neg\sigma$. Since P does not occur in $\Delta \cup \{\sigma\}$, there is a model of it where $\exists\bar{x}P\bar{x}$ is false. Hence, $\Delta \nvdash \sigma \to \exists\bar{x}P\bar{x}$.

follows that $T \nvdash \sigma$. Assume, towards contradiction, that T has an empirically equivalent rival T'. Since, by assumption, $T' \cup \Gamma \cup \{\sigma \rightarrow \exists\bar{x}P\bar{x}\}$ is inconsistent, and T' is equivalent to $T' \cup \Gamma$, it follows that $T' \vdash \neg(\sigma \rightarrow \exists\bar{x}P\bar{x})$ and thus $T' \vdash \sigma$, contrary to our assumption. Hence, T does not have the underdetermination property.[6] It is still possible, of course, to find theories T_1 and T_2 that are both are empirically equivalent to T but jointly inconsistent with each other. To achieve this, just introduce a new theoretical predicate P and define $T_1 = T \cup \{\exists x Px\}$ and $T_2 = T \cup \{\neg\exists x Px\}$. Quine (1975, p. 323) regards this "gratuitous branching of theories" as insignificant, since the new theories are compatible with the old one. Although I am not quite convinced by this line of argument, I will go along with it.

For a theory to have the underdetermination property, it is sufficient that it entails $\exists\bar{x}P\bar{x}$ for some theoretical predicate P.[7] We can then construct a rival by replacing every occurrence of P in the original theory with a new predicate P^* of the same arity, and adding the sentence $\neg\exists\bar{x}P\bar{x}$. The two theories will be jointly consistent but semantically equivalent with respect to their empirical vocabulary. *A fortiori*, the same holds for any weaker notion of empirical equivalence.

Quine (1975, p. 319) has a similar construction:

> Take some theory formulation and select two of its terms, say 'electron' and 'molecule'. I am supposing that these do not figure essentially in any observation sentences; they are purely theoretical. Now let us transform our theory formulation merely by switching these two terms throughout. The new theory formulation will be logically incompatible with the old: it will affirm things about so-called electrons that the other denies.

Quine contends, however, that the example is not a genuine case of underdetermination. The quote continues:

> Yet their only difference, the man in the street would say, is terminological; the one theory formulation uses the technical terms 'molecule' and 'electron' to name what the other formulation calls 'electron' and 'molecule'. The two formulations express, he would say, the same theory.

The intuitive notion of *expressing the same theory* can be explicated in different ways. In the philosophical literature, it is called *theoretical equivalence*. We shall consider various proposals for making this notion precise in the next section. According to Quine's own proposal, two formulations may be taken to express the same theory if they are "reconcilable by reconstrual of predicates", meaning roughly that there is a way of substituting the predicates of one by formulas of the other to

[6] Incidentally, this also works as a counterexample to a claim made by Kukla (1996, p. 138) and Psillos (1999, p. 158) to the effect that, for any theory T (having at least some proper non-empirical consequences, presumably), one can construct an empirically equivalent rival consisting of the empirical consequences of T plus the assertion that T is false. Clearly, this construction does not work in the case at hand.

[7] Here, $\bar{x}$ is a sequence of variables whose length is the same as the arity of P.

obtain a formulation logically equivalent to the latter. Ultimately, Quine (1975, p. 327) settles for the following version of the underdetermination thesis:

> The thesis of under-determination, even in my latest tempered version, asserts that our system of the world is bound to have empirically equivalent alternatives that are not reconcilable by reconstrual of predicates however devious. This, for me, is an open question.

Again without committing to any particular explication of the notions involved, we shall say that a theory T has the **genuine underdetermination property** just in case there is a theory T' such that (i) T and T' are empirically equivalent, (ii) T and T' are jointly inconsistent, and (iii) T and T' are not theoretically equivalent. Observe that, when it comes to establishing this property for a given theory, it is sufficient to do so under the strongest notion of empirical equivalence in combination with the weakest notion of theoretical equivalence.

4 Theoretical Equivalence

In many cases, it is reasonable to regard logically non-equivalent theories as essentially one and the same. For instance, Peano arithmetic formulated in a relational vocabulary (with $(n+1)$-place predicates instead of the standard n-place function symbols) is still essentially Peano arithmetic. In the terminology of Tarski et al. (1953), the two theories are *mutually interpretable*: by adding a set of definitions to one of them, one can prove every theorem of the other, and vice versa. As it stands, this definition is only intended to apply to theories in disjoint vocabularies. In the general case, Tarski et. al. define two theories as mutually interpretable just in case the original definition applies to any of their *copies*, where a *copy* is the result of replacing each predicate in one with a new predicate not occurring in the other. This is equivalent to what we will call *mutual translatability*[8], which is defined in terms of the following notions:

Definition 4.1 (Translation) A *translation* is a function τ from L_1-formulas to L_2-formulas such that $\tau(x=y)$ is $x=y$ and, for any n-place L_1-predicate P, there is an L_2-formula $\varphi(x_1,...,x_n)$ such that $\tau(P\bar{x})=\varphi(\bar{x})$, and for any L_1-formulas φ and ψ,

(i) $\tau(\neg\varphi)=\neg\tau(\varphi)$
(ii) $\tau(\varphi\rightarrow\psi)=\tau(\varphi)\rightarrow\tau(\psi)$
(iii) $\tau(\forall x\varphi)=\forall x\tau(\varphi)$

Definition 4.2 (Translatability) An L_1-theory T_1 is *translatable* into an L_2-theory T_2 just in case there is a translation τ from L_1-formulas to L_2-formulas such that, for any L_1-sentence φ, if $T_1 \vdash \varphi$ then $T_2 \vdash \tau(\varphi)$.

[8] Following modern usage, the term *interpretation* will instead be reserved for what Tarski et al. (1953) call *relative interpretation*, and is defined below.

Two theories are said to be *mutually translatable* just in case each is translatable into the other.

A stronger notion of theoretical equivalence for theories in disjoint vocabularies was suggested by Glymour (1970), called *definitional equivalence*[9]:

Definition 4.3 (Definitional equivalence) Let T_1 and T_2 be theories in disjoint vocabularies L_1 and L_2, respectively. We say that T_1 and T_2 are *definitionally equivalent* just in case there is a set D_{21} of definitions of L_2-predicates in terms of L_1-formulas and a set D_{12} of definitions of L_1-predicates in terms of L_2-formulas such that $T_1 \cup D_{21} \equiv T_2 \cup D_{12}$.[10]

Obviously, this definition can also be generalized to theories in overlapping vocabularies by considering their copies, thus yielding a proper equivalence relation. The generalized notion is shown by Barrett and Halvorson (2016a) to be equivalent to what they call *intertranslatability*:

Definition 4.4 (Intertranslatability) Two theories T_1 and T_2 in L_1 and L_2 respectively are *intertranslatable* just in case there are translations τ_1 and τ_2 such that,

(i) For any L_1-formula φ, if $T_1 \vdash \varphi$ then $T_2 \vdash \tau_1(\varphi)$.
(ii) For any L_2-formula φ, if $T_2 \vdash \varphi$ then $T_1 \vdash \tau_2(\varphi)$.
(iii) For any L_1-formula φ, $T_1 \vdash \forall \bar{x}(\varphi \leftrightarrow \tau_2(\tau_1(\varphi)))$.
(iv) For any L_2-formula φ, $T_2 \vdash \forall \bar{x}(\varphi \leftrightarrow \tau_1(\tau_2(\varphi)))$.

A third notion of theoretical equivalence is provided by Quine (1975):

Definition 4.5 (Reconcilability) An L_1-theory T_1 is *reconcilable* with an L_2-theory T_2 just in case there is a translation τ form L_1-formulas to L_2-formulas such that $\tau(T_1) \equiv T_2$.[11]

This is what Barrett and Halvorson (2016a) call *Quine equivalence*, which is something of a misnomer, since (as they themselves point out) it is not an equivalence relation:

Fact 4.1 *Reconcilability is not symmetric.*

Proof Let $L_1 = \{P\}$, $L_2 = \emptyset$, $T_1 = \{\forall x Px\}$ and $T_2 = \emptyset$. Then there is a translation τ_1 such that $\tau_1(T_1) \equiv T_2$, namely the one defined by letting $\tau_1(Px)$ be $x = x$, but there is obviously no translation τ_2 such that $\tau_2(T_2) \equiv T_1$. □

[9] Thus defined, this is actually not an equivalence relation, since it is not reflexive. An essentially identical (but reflexive) notion can be found already in De Bouvère (1965), and a slightly different one based on the same idea in Kanger (1968).

[10] $\equiv$ denotes logical equivalence.

[11] Here, and elsewhere, we shall write $\tau(T)$ for $\{\tau(\varphi) : \varphi \in T\}$.

To get a proper notion of theoretical equivalence in terms of reconcilability, the most natural solution is perhaps to define it as *mutual* reconcilability. However, as noted by Barrett and Halvorson (2016a):

Fact 4.2 *Mutual reconcilability does not preserve completeness.*

Proof Let $L = L_1 = L_2 = \{P, Q\}$, $T_1 = \{\forall x \forall y(x = y), \forall x Px\}$ and $T_2 = \{\forall x \forall y(x = y), \forall x(Px \wedge Qx)\}$. Clearly, T_2 but not T_1 is complete with respect to L. But T_1 and T_2 are mutually reconcilable, as witnessed by the translations τ_1 and τ_2 defined by $\tau_1(Px) = Px \wedge Qx$, $\tau_1(Qx) = Qx$, $\tau_2(Px) = Px$, and $\tau_2(Qx) = Px$. □

This is their chief complaint against Quine's proposal (and its modification). By contrast, they show that their own proposal does not suffer from such drawbacks:

Fact 4.3 *Intertranslatability preserves completeness and decidability.*

Proof Assume that T_1 and T_2 are intertranslatable with τ_1 and τ_2, and that T_1 is complete. Let φ be an L_2-sentence. By completeness of T_1, we get two cases:

(i) $T_1 \vdash \tau_2(\varphi)$, in which case $T_2 \vdash \tau_1(\tau_2(\varphi))$ and therefore $T_2 \vdash \varphi$.
(ii) $T_1 \vdash \neg\tau_2(\varphi)$, in which case $T_2 \vdash \tau_1(\tau_2(\neg\varphi))$ and therefore $T_2 \vdash \neg\varphi$.

Hence, T_2 is complete. Next, assume that T_1 is decidable. Let φ be an L_2-sentence. $T_2 \vdash \varphi$ implies $T_1 \vdash \tau_2(\varphi)$, which implies $T_2 \vdash \tau_1(\tau_2(\varphi))$, which implies $T_2 \vdash \varphi$. Hence, for any L_2-sentence φ, $T_2 \vdash \varphi$ iff $T_1 \vdash \tau_2(\varphi)$, which means that T_2 is decidable. □

As a consequence, Barrett and Halvorson (2016a) are able to conclude that the two proposals are logically independent:

Fact 4.4 *Intertranslatability does not entail mutual reconcilability, and vice versa.*

Proof For left to right, let $L_1 = \{P\}$, $L_2 = \emptyset$, $T_1 = \{\forall x Px\}$ and $T_2 = \emptyset$. Let τ_1 be the translation defined by letting $\tau_1(Px)$ be $x = x$, and let τ_2 be the identity function. It is easy to verify that T_1 and T_2 are intertranslatable with τ_1 and τ_2. But there is obviously no translation τ such that $\tau(T_2) \equiv T_1$. The other direction follows from the fact that intertranslatability preserves completeness (Fact 4.3) while mutual reconcilability does not (Fact 4.2). □

Our fourth and final notion of theoretical equivalence is called *mutual interpretability*. An *interpretation* is very much like a translation, except that it is allowed to restrict the quantifiers. In a nutshell, interpretation is translation followed by relativization:

Definition 4.6 (Interpretation) An *interpretation* is a function I from L_1-formulas to L_2-formulas such that $I(x = y)$ is $x = y$ and, for any n-place L_1-predicate P, there is

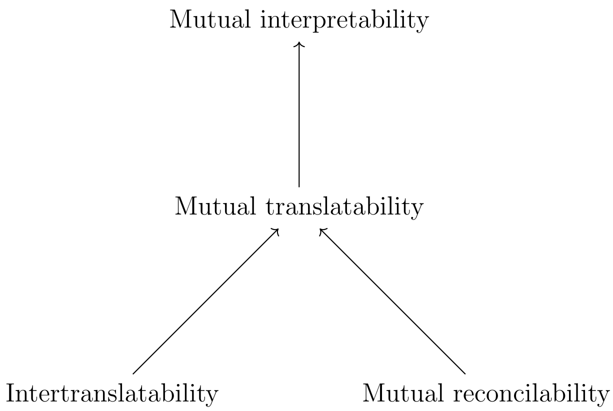

Fig. 2 The relation of entailment between the four notions of theoretical equivalence

an L_2-formula $\varphi(x_1, ..., x_n)$ such that $I(P\bar{x}) = \varphi(\bar{x})$, and there is an L_2-formula $\delta(x)$ (a so-called *domain formula*) such that, for any L_1-formulas φ and ψ,

(i) $I(\neg\varphi) = \neg I(\varphi)$
(ii) $I(\varphi \rightarrow \psi) = I(\varphi) \rightarrow I(\psi)$
(iii) $I(\forall x\varphi) = \forall x(\delta(x) \rightarrow I(\varphi))$

Definition 4.7 (Interpretability) An L_1-theory T_1 is *interpretable* by an L_2-theory T_2 just in case there is an interpretation I from L_1-formulas to L_2-formulas such that, for any L_1-sentence φ, if $T_1 \vdash \varphi$ then $T_2 \vdash I(\varphi)$. Relative to I, we then say that T_2 *interprets* T_1.

As in the case of translatability, *mutual interpretability* is defined as interpretability in both directions. For instance, Zermelo-Fraenkel set theory can interpret Peano arithmetic (e.g by representing the natural numbers as finite von Neumann ordinals), but the latter cannot interpret the former. Hence, these two theories are not mutually interpretable.

Let us compare the four notions in terms of strength (the results are summarized in Fig.2). We have already established the independence between intertranslatability and mutual reconcilbability (Fact 4.4). By definition, intertranslatability entails mutual translatability. By taking $x = x$ as the domain formula, it is obvious that translatability entails interpretability. To see the failure of the converse, it is enough to consider the theories $T_1 = \{\exists!xPx\}$ and $T_2 = \{\exists!xPx, \forall xPx\}$. Since $T_2 \vdash \exists!x(x = x)$ but $T_1 \nvdash \exists!x(x = x)$, it follows that T_2 is not translatable into T_1. But it should be sufficiently obvious that we can interpret T_2 in T_1 with Px as our domain formula. Finally, to see that reconcilability entails translatability, we may use the following lemma:

Lemma 4.1 *Let T be an L-theory, and let I be an interpretation on L-formulas with domain formula $\delta(x)$. Then, for any L-formula φ, if $T \vdash \varphi$ then $I(T) \cup \{\exists x\delta\} \vdash I(\varphi)$.*

Proof An interpretation I with domain formula $\delta(x)$ provides a recipe for transforming any model $\mathcal{M}$ of $I(T) \cup \{\exists x\delta\}$ into a model $\mathcal{M}'$ of T. The domain of $\mathcal{M}'$ is defined as the extension of δ in $\mathcal{M}$, and each predicate P in L is interpreted by $\mathcal{M}'$ as the extension of $I(P\bar{x})$ in $\mathcal{M}$ intersected with the domain of $\mathcal{M}'$. Let X be the set of variables. It is easy to verify, by induction on the complexity of the formula φ, that for any assignment $v : X \to |\mathcal{M}'|$, we have $\mathcal{M}', v \vDash \varphi$ iff $\mathcal{M}, v \vDash I(\varphi)$. The base case is immediate from the construction, and the cases of boolean operators are straightforward. In the case of universal quantification, we get $\mathcal{M}', v \vDash \forall x\varphi$ iff, for all $a \in |\mathcal{M}'|$, $\mathcal{M}', v(a/x) \vDash \varphi$ iff (by induction hypothesis), for all $a \in |\mathcal{M}'|$, $\mathcal{M}, v(a/x) \vDash I(\varphi)$ iff, for all $a \in |\mathcal{M}|$, $\mathcal{M}, v(a/x) \vDash \delta \to I(\varphi)$ iff $\mathcal{M}, v \vDash \forall x(\delta \to I(\varphi))$ iff $\mathcal{M}, v \vDash I(\forall x\varphi)$. Having established as much, it now follows by soundness that $I(T) \cup \{\exists x\delta\} \nvDash I(\varphi)$ implies $T \nvDash \varphi$, yielding the desired result by contraposition. □

Fact 4.5 *Reconcilability* (*witnessed by a certain translation*) *entails translatability* (*witnessed by the same translation*).

Proof Assume that τ is a translation such that $\tau(T_1) \equiv T_2$. Define an interpretation I with domain formula $x = x$ that is identical to τ with respect to all atomic formulas. Clearly, for any formula φ, we have $\tau(\varphi) \equiv I(\varphi)$. Assume that $T_1 \vdash \varphi$. By Lemma 4.1, we have $I(T_1) \vdash I(\varphi)$. By assumption of reconcilability, it follows that $T_2 \vdash I(\varphi)$, and thus $T_2 \vdash \tau(\varphi)$. Hence, τ is a translation of T_1 into T_2. □

Hence, of these four notions, mutual interpretability is the weakest one. It is widely considered too weak to capture the intuitive concept of *expressing the same theory*. But in the context of establishing the genuine underdetermination property, the question is rather whether it is weak enough. In recent years, several other notions of theoretical equivalence have been proposed using concepts from category theory.[12] The weakest of them (as far as I know) is *categorical equivalence*. There is also a stronger notion called *Morita equivalence*. Both of these naturally apply to theories in many-sorted languages. With mutual interpretability suitably generalized to the many-sorted case, McEldowney (2020, p. 16) provides an example of Morita equivalent theories that are not mutually interpretable. Hence, in the many-sorted case, categorical equivalence does not entail mutual interpretability. Whether it does so in the singe-sorted case seems to be an open question.[13] But McEldowney (2020, p. 19, Proposition 5.11) also shows that, for theories implying that there at least two things, Morita equivalence entails something called *bi-interpretability*, which in turn entails mutual interpretability. In addition, Barrett and Halvorson (2016b, p. 575) conjecture that, for theories in finite vocabularies, categorical equivalence entails Morita equivalence. Thus, if their conjecture is correct, categorical equivalence entails mutual interpretability for all single-sorted theories in finite

[12] See, for instance, Barrett and Halvorson (2016b), Weatherall (2016), and Hudetz (2019).

[13] It was posted on MathOverflow some years ago: https://mathoverflow.net/questions/146343/the-interplay-between-certain-aspects-of-interpretability-model-theory-and-cate/152695

vocabularies implying that there are at least two things. For present purposes, that would be quite sufficient. In any case, it seems reasonable to regard mutual interpretability as a necessary condition for theoretical equivalence.

5 Answering Quine's Question

As we saw earlier, given a theory T such that $T \vdash \exists\bar{x}P\bar{x}$ for some theoretical predicate P, constructing an empirically equivalent rival T^* is a trivial matter: just replace P everywhere with a new predicate P^*, and add the sentence $\neg\exists\bar{x}P\bar{x}$. But the two theories are mutually translatable: we can translate every theorem of T to a theorem of T^* by replacing P with P^*, and we can translate every theorem of T^* to a theorem of T by replacing P^* with P and $P\bar{x}$ with $\neg\forall x(x = x)$. It is easy to check that these translations also satisfy the requirements for intertranslatability, making T and T^* theoretically equivalent in a rather strong sense.

If T is consistent, recursively axiomatizable, and does not have any finite models, then T^* will inherit these properties. In that case, as a consequence of Theorem 5.1 below, one can extend T^* with a single sentence (saying, essentially, that T^* is consistent), thereby producing a theory with the same empirical content as T^*, but one that T *cannot* interpret. In particular, partially answering Quine's question, it follows that any consistent and recursively axiomatizable theory that postulates infinitely many theoretical entities (numbers, for instance) is bound to have empirically equivalent alternatives that are not reconcilable by reconstrual of predicates however devious.

We proceed as follows. Let PA be Peano arithmetic formulated in vocabulary L_{PA}. Recall that an L_{PA} formula $\varphi(x_1, ..., x_k)$ is said to *represent* a relation $R \subseteq \mathbb{N}^k$ in PA just in case, for any $n_1, ..., n_k \in \mathbb{N}$, we have $PA \vdash \varphi(\underline{n_1}, ..., \underline{n_k})$ if $n_1, ..., n_k \in R$, and $PA \vdash \neg\varphi(\underline{n_1}, ..., \underline{n_k})$ if $n_1, ..., n_k \notin R$, where $\underline{n}$ is the numeral associated with each natural number n. A fundamental result in mathematical logic says that *a relation on the natural numbers is recursive just in case there is a* Σ_1*-formula representing it in* PA, where a Σ_1-formula is an L_{PA}-formula of the form $\exists\bar{x}\varphi$, with φ only containing restricted quantifiers.[14]

In a derivative sense, and relative to a given Gödel-numbering # of finite sequences of symbols of some first-order language, a theory T in this language is said to be recursive (or represented by a formula in PA) just in case $\{\#(\varphi) : \varphi \in T\}$ is.[15] Thus, if T is recursive, there is a Σ_1-formula $\alpha(x)$ representing it in PA. Moreover, we can then construct an L_{PA}-formula $Prf_\alpha(x, y)$ representing the fact that x is the Gödel-number of a proof whose premises belong to T and whose last sentence has

[14] The property of *only containing restricted quantifiers* can be defined inductively as follows. All atomic formulas have the property, and the property is preserved under boolean operators. Furthermore, if φ has the property, and t is a term not containing the variable x, then $\exists x(x < t \wedge \varphi)$ and $\forall x(x < t \rightarrow \varphi)$ also have the property.

[15] A theory is said to be *recursively axiomatizable* just in case it is logically equivalent to a recursive theory.

Gödel-number y. Finally, given an L_{PA}-formula $Neg(x, y)$ representing the fact that x is the Gödel-number of the negation of a formula with Gödel-number y, we can define the L_{PA}-sentence Con_α as

$$\neg\exists y(\exists x Prf_\alpha(x, y) \wedge \exists x \exists z(Prf_\alpha(x, z) \wedge Neg(z, y)))$$

essentially saying that T is consistent. Gödel's second incompleteness theorem then states that, *if T is a consistent recursive extension of PA, then $T \nvdash Con_\alpha$*. To establish the genuine underdetermination property, we use a corollary of Gödel's theorem, due to Feferman (1966, p. 90, Theorem 8.8):

Lemma 5.1 (Feferman) *Let T be a consistent theory, let I be an interpretation such that $I(PA) \subseteq T$, and assume that $\alpha(x)$ is a Σ_1-formula representing T in PA. Then T cannot interpret $T \cup \{I(Con_\alpha)\}$.*

Remark 5.1 Clearly, the requirement that $I(PA) \subseteq T$ can be replaced by the weaker requirement that I is an interpretation of PA in T.

To strengthen our result, we shall also use a well-known result by Craig and Vaught (1958, p. 292, Theorem 2.1):

Lemma 5.2 (Craig and Vaught) *Let T be a recursive theory in a finite vocabulary L, and assume that T does not have any finite models. Then there is a finite theory T' in a vocabulary $L' \supseteq L$ semantically L-equivalent to T.*

Here is our main result:

Theorem 5.1 *Let T be a theory in vocabulary L_T, and let $L \subseteq L_T$. Assume that (i) T is consistent, (ii) T does not have any finite models, and that (iii) there is a recursive theory T^* semantically L-equivalent to T such that T and T^* are jointly inconsistent. If T^* can interpret T, there is a finite extension*[16] *T' of T^* such that (iv) T and T' are semantically L-equivalent, and (v) T cannot interpret T'.*

Proof Let T be a theory in vocabulary L_T, with empirical part $L \subseteq L_T$. Assume that (i) T is consistent, (ii) T does not have any finite models, and that (iii) there is a recursive theory T^* in a vocabulary $L_{T^*} \supseteq L$ semantically L-equivalent to T such that T and T^* are jointly inconsistent. By Lemma 5.2, there is a finite theory PA' in a vocabulary $L'_{PA} \supseteq L_{PA}$ semantically L_{PA}-equivalent to PA. Let L^*_{PA} be a copy of L'_{PA} disjoint from $L_T \cup L_{T^*}$ that also contains a new unary predicate N. For any L'_{PA}-formula φ, let φ^*_N be the L^*_{PA}-formula you get by first replacing every L'_{PA}-symbol in φ with the corresponding L^*_{PA}-symbol, and then relativizing the whole thing to N. Let

[16] To be clear, a set A is a *finite extension* of a set B just in case there is a finite set C such that $A = B \cup C$.

$\mathcal{N}' \vDash PA'$ be an expansion of the standard model of arithmetic, and let $\mathcal{N}^*$ be the corresponding L^*_{PA}-model with $\mathcal{N}^* \vDash \forall x Nx$.

Let $PA^*_N = \{\varphi^*_N : \varphi \in PA'\}$ and $T^+ = T^* \cup PA^*_N \cup \{\exists x Nx\}$. Observe that, since $PA \vdash \varphi$ implies $PA' \vdash \varphi$, which by Lemma 4.1 implies $PA^*_N \cup \{\exists x Nx\} \vdash \varphi^*_N$, it follows that

(1) The map $\varphi \mapsto \varphi^*_N$ is an interpretation of PA in T^+.

We first show that

(2) The L-reduct of every model of T can be expanded to a model of T^+.

Suppose that $\mathcal{M} \vDash T$. By assumption (iii), $\mathcal{M}|L$ can be expanded to a model of T^*. Since, by assumption (ii), $\mathcal{M}|L$ is infinite, we can further expand it by interpreting L^*_{PA} on a countably infinite subset of its domain, yielding a model $\mathcal{M}'$ such that $\mathcal{M}'_N|L^*_{PA} \cong \mathcal{N}^*$. Hence, $\mathcal{M}' \vDash T^+$.

Since T is consistent, it follows that T^+ is too. It follows by assumption (iii) that there is a Σ_1-formula $\alpha(x)$ representing T^+ in PA. Let us write $Con(T^+)$ for the corresponding L_{PA}-sentence Con_α, and let $T' = T^+ \cup \{Con(T^+)^*_N\}$. It now follows by (1) and Lemma 5.1 that

(3) T^+ cannot interpret T'.

Consistency of T^+ also implies that $\mathcal{N}' \vDash Con(T^+)$, and thus $\mathcal{N}^* \vDash Con(T^+)^*_N$. By the same line of reasoning as before, it follows that

(4) The L-reduct of every model of T can be expanded to a model of T'.

Since T' is an extension of T^*, it already follows form assumption (iii) that

(5) The L-reduct of every model of T' can be expanded to a model of T.

Hence, T and T' are semantically L-equivalent. Finally, we show that

(6) If T^* can interpret T, then T cannot interpret T'.

Let I be an interpretation of T in T^*. Assume, towards contradiction, that J is an interpretation of T' in T. Since $T \vdash J(\varphi)$ implies $T^* \vdash I(J(\varphi))$, which implies $T^+ \vdash I(J(\varphi))$, the map $\varphi \mapsto I(J(\varphi))$ is an interpretation of T' in T^+, contrary to (3). □

Remark 5.2 Obviously, the construction can be iterated, generating an infinite sequence of empirically equivalent but theoretically non-equivalent rivals.

Suppose, for instance, that T is a theory postulating infinitely many so-called 'numbers'. First, let T^* be just like T, except that it says about 'numbers*' what T says about 'numbers', and that 'numbers' do not exist. Secondly, extend T^* to a theory T^+ by adding a version of Peano arithmetic formulated in a vocabulary disjoint from that of T and T^*, talking about 'numbers**'. Provided that T (and thereby T^+) is recursive, we can express the claim that no so-called 'number**' codes a proof of a contradiction in T^+. Finally, adding this claim to T^+ yields a theory T' which T cannot interpret (provided that T is consistent).

Thus constructed, there is an obvious sense in which the theory about the so-called 'numbers**' included in T' is superfluous: it can be removed without loss of empirical content. However, it is easy to construct a theory logically equivalent to T' to which this does not apply: just transform all the other claims of T' to claims conditional on the number**-claims. Relative this new theory, there is no obvious sense in which the number**-claims are superfluous, lest *all* non-empirical consequences of the theory are rendered superfluous.[17]

6 Concluding Remarks

We have established Quine's underdetermination thesis for all consistent and recursively axiomatizable theories postulating infinitely many theoretical entities. It is clear that any straightforward first-order axiomatization of, say, *General relativity theory* will satisfy this requirement, namely by postulating infinitely many real numbers.[18] Quine (1975, p. 324) thought that postulating theoretical entities might be necessary for a theory to be finite:

> Here, evidently, is the nature of under-determination. There is some infinite lot of observation conditionals that we want to capture in a finite formulation. Because of the complexity of the assortment, we cannot produce a finite formulation that would be equivalent merely to their infinite conjunction. Any finite formulation that will imply them is going to have to imply also some trumped-up matter, or stuffing, whose only service is to round out the formulation. There is some freedom of choice of stuffing, and such is the under-determination.

It can indeed be shown that, in order for a finite theory to imply certain sets of empirical claims, it has to postulate an infinite number of theoretical entities.[19]

Realists may respond to the challenge of underdetermination in a number of ways. In my view, the most promising response is perhaps to weaken their claim, by insisting only on the *approximate* truth of our best scientific theories. This is what most realists seem to do anyway. Intuitively, two logically incompatible theories may both be

[17] I am not saying that no such sense exists. For instance, Schurz (2009) has a notion of a theoretical expression *yielding* the empirical success of a theory to which it belongs, which may be applicable in this context.

[18] See, for instance, Andréka et al. (2007).

[19] See Johannesson (2022, pp. 22–23).

approximately true. Insofar as the notion of *approximate truth* can be made precise, a corresponding underdetermination thesis may be formulated and evaluated (with 'logically incompatible' replaced by 'not both approximately true').

Funding Open access funding provided by Stockholm University.

References

Andréka, H., Madarász, J. X., & Németi, I. (2007). *Logic of Space-Time and Relativity Theory*. Netherlands: Springer.

Barrett, T. W., & Halvorson, H. (2016). Glymour and Quine on theoretical equivalence. *Journal of Philosophical Logic, 45*(5), 467–483.

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

Craig, W., & Vaught, R. L. (1958). Finite axiomatizability using additional predicates. *The Journal of Symbolic Logic, 23*(3), 289–308.

De Bouvère, K. (1965). Logical synonymity. *Indagationes Mathematicae (Proceedings), 68,* 622–629.

Feferman, S. (1966). Arithmetization of metamathematics in a general setting. *Journal of Symbolic Logic, 31*(2), 269–270.

Glymour, C. (1970). Theoretical realism and theoretical equivalence. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1970:275–288.

Hudetz, L. (2019). Definable categorical equivalence. *Philosophy of Science, 86*(1), 47–75.

Johannesson, E. (2020). Realism and empirical equivalence. *Journal of Philosophical Logic, 49*(3), 475–495.

Johannesson, E. (2022). On the indispensability of theoretical terms and entities. *Synthese, 200*(136), 1–25.

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

Ketland, J. (2004). Empirical adequacy and ramsification. *British Journal for the Philosophy of Science, 55*(2), 287–300.

Kukla, A. (1996). Does every theory have empirically equivalent rivals? *Erkenntnis, 44*(2), 137–166.

Laudan, L., & Leplin, J. (1991). Empirical equivalence and underdetermination. *Journal of Philosophy, 88*(9), 449–472.

McEldowney, P. A. (2020). On morita equivalence and interpretability. *The Review of Symbolic Logic, 13*(2), 388–415.

Melia, J. (2000). Weaseling away the indispensability argument. *Mind, 109*(435), 455–479.

Okasha, S. (2002). *Philosophy of science : a very short introduction*. Oxford: Oxford University Press.

Psillos, S. (1999). *Scientific realism : How science tracks truth*. New York: Routledge.

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

Schurz, G. (2009). When empirical success implies theoretical reference: A structural correspondence theorem. *The British Journal for the Philosophy of Science, 60*(1), 101–133.

Schurz, G. (2013). *Philosophy of Science*. Hoboken: Routledge.

Tarski, A., Mostowski, A., & Robinson, R. M. (1953). *Undecidable theories*. Amsterdam: North-Holland.

Turney, P. (1990). Embeddability, syntax, and semantics in accounts of scientific theories. *Journal of Philosophical Logic, 19*(4), 429–451.

van Benthem, J. F. A. K. (1978). Ramsey eliminability. *Studia Logica, 37*(4), 321–336.

Van Fraassen, B. C. (1980). *The scientific image*. Oxford: Clarendon P.

Weatherall, J. O. (2016). Are newtonian gravitation and geometrized newtonian gravitation theoretically equivalent? *Erkenntnis, 81*(5), 1073–1091.

Worrall, J. (2011). Underdetermination, realism and empirical equivalence. *Synthese, 180*(2), 157–172.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.