Realism and Empirical Equivalence

Eric Johannesson

Received: 17 October 2018 / Accepted: 1 August 2019 / Published online: 27 August 2019 © The Author(s) 2019

Abstract

The main purpose of this paper is to investigate various notions of empirical equivalence in relation to the two main arguments for realism in the philosophy of science, namely the no-miracles argument and the indispensability argument. According to realism, one should believe in the existence of the theoretical entities (such as numbers and electrons) postulated by empirically adequate theories. According to the no-miracles argument, one should do so because truth is the best explanation of empirical adequacy. According to the indispensability argument, one should do so because the theoretical terms employed in the formulation of an empirically adequate theory are practically indispensable for the formulation of its empirical content. The no-miracles argument might be refuted if one can establish an underdetermination thesis to the effect that every theory has empirically equivalent rivals that are incompatible with each other. Insofar as truth cannot explain the empirical adequacy of two incompatible theories, there is an obvious conflict between the underdetermination thesis and the no-miracles argument. I show that, under certain assumptions, some but not all notions of empirical equivalence support the underdetermination thesis. The indispensability argument might be refuted if one can establish a dispensability thesis to the effect that, for any theory with a practical formulation (e.g. for any axiomatizable theory), there is an empirically equivalent theory with a practical formulation in purely empirical terms. I show that (using axiomatizability as the measure of practicality) some but not all notions of empirical equivalence support this thesis.

Keywords Realism · Underdetermination · Empirical equivalence · Indispensability · Theoretical terms · Quine

Eric Johannesson
johannesson.eric@gmail.com

1 Department of Philosophy, Stockholm University, SE-106 91 Stockholm, Sweden
1 Introduction

Scientific theories help predict the observable behavior of ordinary things such as planets, bridges, and light-bulbs with great accuracy and precision. They typically do so by postulating non-observable things such as electrons, energy, or dark matter. Roughly, if a theory successfully predicts the observable behavior of ordinary things, we say that it is \textit{empirically adequate}. But does empirical adequacy provide reason for thinking that the theory is true and that all non-observable things it postulates exist?

Realists in the philosophy of science argue that it does. According to one of the most influential arguments for realism, \textit{the no-miracles argument},\textsuperscript{1} the best explanation for why a theory is empirically adequate is that it is true. According to a second influential argument, \textit{the indispensability argument},\textsuperscript{2} the theoretical terms employed in the formulation of an empirically adequate theory are practically indispensable for the formulation of its empirical content.

Two theories that make the same empirical predictions are said to be \textit{empirically equivalent}. At least some versions of the no-miracles argument can be refuted if one can establish an \textit{underdetermination thesis} to the effect that

1. For any theory \( T \), there are theories \( T_1 \) and \( T_2 \) empirically equivalent to \( T \) such that \( T_1 \) and \( T_2 \) are incompatible.

Insofar as truth cannot explain the empirical adequacy of two incompatible theories, there is an obvious conflict between this thesis and the no-miracles argument. The status of the thesis remains a matter of dispute, however.\textsuperscript{3} Under certain assumptions, presented in Section 2, it has been shown that the classical notion of empirical equivalence doesn’t support it, but rather the opposite: underdetermination becomes impossible [6, p. 373]. For an exposition, see Section 3. In Section 4, we show that a weaker notion of empirical equivalence still does. A third notion of empirical equivalence is considered in Section 5, but we have not been able to show whether it supports the thesis or not.

The indispensability argument can be refuted if one can establish a \textit{dispensability thesis} to the effect that

2. For any theory with a practical formulation (e.g. for any axiomatizable theory), there’s an empirically equivalent theory with a practical formulation in purely empirical terms.

In Section 6, we show that (using axiomatizability as the measure of practicality), some but not all notions of empirical equivalence support this thesis.

\textsuperscript{1}The no-miracles argument was first formulated by Putnam [19, p. 73]. More refined versions of the argument have been suggested by Boyd [3], Psillos [18], Psillos [24], and Dawid and Hartmann [5].

\textsuperscript{2}The indispensability argument may be traced back to Quine [20], and is more explicitly formulated by Putnam [19, p. 347].

\textsuperscript{3}Arguments for the underdetermination thesis have been offered by Van Fraassen [26], Earman [7], Kukla [12] and Severo [22]. Arguments against it have been offered by Laudan and Leplin [14] and Laudan and Leplin [27].
2 Empirical and Non-Empirical Vocabulary

Following Carnap [4], I shall assume that the distinction between empirical and non-empirical consequences of a theory can be drawn by dividing the vocabulary in which the theory is formulated into an empirical and non-empirical part. Canonically, ‘dog’, ‘table’ and ‘bigger than’ are examples of empirical terms, whereas ‘number’, ‘electron’ and ‘oppositely charged’ are examples of non-empirical (i.e. theoretical) terms. On the classical notion of empirical equivalence, two theories are empirically equivalent just in case they entail the same empirical sentences, where an empirical sentence is one containing only logical and empirical terms (e.g. ‘All dogs bark’). As we shall see, other notions of empirical equivalence are possible. But we shall assume throughout that the vocabulary $L_T$ of any theory $T$ can be split into two parts, an empirical part $E_T$ and a theoretical part $L_T - E_T$. For present purposes, I shall take this distinction as basic, and not attempt to explicate or define it. By theory, I shall mean a set of sentences, not necessarily deductively closed. Unless otherwise stated, all theories are assumed to be first-order.

Admittedly, the distinction between empirical and non-empirical consequences is controversial, as is the distinction between what is directly observable and what is merely deducible from such observations together with some background theory. If one equates empirical consequences with what is directly observable, and requires direct observations to be infallible, it follows that the empirical consequences of a theory basically only pertain to the distribution of patches of color in ones visual field. On these assumptions, there is no epistemological difference between tables and electrons (they are both theoretical entities), and no distinction between empirical and non-empirical consequences that seems relevant to the debate about realism and anti-realism in the philosophy of science.

However, one need not require direct observations to be infallible. We may also grant that there is no metaphysical distinction to be made between tables and electrons, or even numbers. If they exist, they all exist in the same way, so to speak. But realism understood in the above manner is a thesis concerning human epistemology. For such a thesis, the distinction between what human beings considered as measuring instruments can and cannot (fallibly) detect will be relevant. As Ladyman [13, p. 190] puts it:

... realists do not see why our physical constitution, as a contingent feature of our evolution, has any philosophical significance whatsoever. One response to this is simply to restate the opposite intuition: what else but our (biologically determined) observational capacities would one consider relevant to our epistemology?

Some realists do not want to recognize the distinction between empirical and non-empirical consequences in the first place. They seem to think that its mere recognition yields an easy win for anti-realism. As we shall see, that is not true. But there is an easy argument for the underdetermination thesis lurking in the background that might explain this sentiment. The argument, however, is deeply flawed. It runs as follows: take any theory $T$ (not equivalent to its empirical consequences, presumably), and let $T^*$ consist of all the empirical consequences of $T$ plus the assertion that $T$ is
false. Then, it is claimed, $T^*$ will be empirically equivalent to $T$ and yet jointly inconsistent with it. Hence, no matter what one takes the empirical consequences of a theory to be (as long as they don’t encompass all the consequences of the theory), the underdetermination thesis is true. The argument is usually dismissed by realists because they do not think $T^*$ is a genuine alternative to $T$. But the real problem with the argument is that there is no guarantee that $T^*$, constructed in this way, is empirically equivalent to $T$. To see why, let $T_E$ be a set of empirical consequences such that $T_E \not\vdash \epsilon$ and $T_E \not\vdash \epsilon \rightarrow \theta$ for some empirical sentence $\epsilon$, e.g. ‘Pigs can fly’, and some theoretical sentence $\theta$, e.g. ‘Numbers exist’. Consider the theory $T = T_E \cup \{\epsilon \rightarrow \theta\}$. Since $T_E$ has a model where $\epsilon$ is false and therefore $\epsilon \rightarrow \theta$ is true, it follows that $T \not\vdash \epsilon$. Now, let $T^*$ consist of the empirical consequences of $T$ plus the assertion that $T$ is false. Since $T^* \cup T_E \cup \{\epsilon \rightarrow \theta\}$ does not have a model, and $T^* \cup T_E$ is equivalent to $T^*$, it follows that $T^* \vdash \neg(\epsilon \rightarrow \theta)$. Hence, $T^* \vdash \epsilon$, which means that $T^*$ is not empirically equivalent to $T$.

A more promising algorithm for generating empirically equivalent rivals is suggested by Quine [21, p. 319]:

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 example suggest that joint inconsistency is not sufficient for incompatibility. But in order for the underdetermination thesis to have some bearing on the no-miracles argument, we shall assume it to be necessary. In other words, we shall assume that

\begin{equation}
(3) \quad \text{Incompatible theories are jointly inconsistent.}
\end{equation}

In order to avoid the kind of triviality raised by Quine’s example, there are several options. One option, advocated by Quine himself, is to assume that

\begin{equation}
(4) \quad \text{Consistent incompatible theories are (in some sense) not inter-translatable.}
\end{equation}

The sense Quine had in mind was roughly that two theories $T_1$ and $T_2$ are \emph{inter-translatable} just in case, for all predicates $P_1, \ldots, P_n$ in the vocabulary of $T_1$ there are formulas $\varphi_1, \ldots, \varphi_n$ in the vocabulary of $T_2$ such that replacing all occurrences

\footnote{Cf. Kukla [12, p. 138], who attributes the idea to Van Fraassen (incorrectly, as far as I can tell), and Psillos [18, p. 158].}
of \( P_1, \ldots, P_n \) in \( T_1 \) by \( \varphi_1, \ldots, \varphi_n \) yields a theory equivalent to \( T_2 \), and vice versa. But other notions of inter-translatability are possible.\(^5\)

Another option is to assume that inter-translatability preserves incompatibility:

(5) If \( T_1 \) and \( T_2 \) are incompatible, and \( T_2 \) and \( T'_2 \) are inter-translatable, then \( T_1 \) and \( T'_2 \) are incompatible as well.

This principle, which I take to be plausible enough to merit attention, allows for certain simplifications. Clearly, (3) and (5) already entail (4). Moreover, on any reasonable notion of empirical equivalence and inter-translatability,\(^6\) they also entail that

(6) If \( T_1 \) and \( T_2 \) are incompatible, then there is a theory \( T'_2 \) empirically equivalent to \( T_2 \) such that \( T_1 \) and \( T'_2 \) are incompatible and do not share any theoretical vocabulary.

That means we can assume, without loss of generality, that incompatible theories do not share any theoretical vocabulary. This assumption can be motivated by the idea, most famously advocated by Feyerabend [9], that theoretical terms have no meaning outside of the theoretical context in which they occur. Indeed, that this is part of what makes them theoretical rather than empirical in the first place. In this paper, I shall therefore assume (5). The underdetermination thesis under consideration will in effect be the following:

(7) For any theory \( T \), there are theories \( T_1 \) and \( T_2 \) empirically equivalent to \( T \) such that \( T_1 \) and \( T_2 \) are jointly inconsistent and do not share any theoretical vocabulary.

3 Strong Empirical Equivalence

Arguably, the classical notion of empirical equivalence corresponds to what we shall call strong empirical equivalence. To define it, we first introduce the following notion:

**Definition 3.1** (E-fragment) The \( E \)-fragment of a theory \( T \) is the set of theorems of \( T \) that are \( E_T \)-formulas.

Empirical equivalence is then defined as follows:

**Definition 3.2** (Strong empirical equivalence) Two theories are strongly empirically equivalent just in case they have the same \( E \)-fragments.

\(^{5}\)Cf. Barrett and Halvorson [1].

\(^{6}\)All it takes is that the result of replacing every theoretical predicate \( P \) in a theory \( T \) with a new theoretical predicate \( P^* \) yields a theory \( T^* \) such that \( T \) and \( T^* \) are both empirically equivalent and inter-translatable. As far as empirical equivalence goes, all notions thereof considered in this paper satisfy this constraint.
Granted that the only vocabulary rival theories have in common is empirical, it can be shown that strongly empirically equivalent consistent theories are always jointly consistent, using a theorem by Robinson [2, p. 264]:

**Theorem 3.1** (Robinson’s joint consistency theorem) If $T$ is a complete theory in $L_1 \cap L_2$, and $T_1$ and $T_2$ are satisfiable extensions of $T$ in $L_1$ and $L_2$ respectively, then $T_1 \cup T_2$ is satisfiable.

**Corollary 3.1** If $T_1$ and $T_2$ are strongly empirically equivalent consistent theories sharing no theoretical vocabulary, then $T_1 \cup T_2$ is consistent.

**Proof** Assume, towards contradiction, that $T_1$ and $T_2$ are (i) consistent theories that are (ii) equivalent over their common vocabulary $L$, and that (iii) $T_1 \cup T_2$ is inconsistent. Since $T_1$ is consistent, it has a complete and consistent extension $T_1^*$. The $L$-sentences of $T_1^*$ forms a complete and consistent $L$-theory $T^*$ such that $T^* \subseteq T_1^*$.

Assume, towards contradiction, that $T_2^* = T_2 \cup T^*$ is consistent. By Robinson’s joint consistency theorem, it follows that $T_2^* \cup T_1^*$ is consistent. That means $T_1 \cup T_2$ is consistent, contradicting (iii). Hence, $T_2^* \cup T^*$ is inconsistent. That means there is an $L$-sentence $\varphi$ such that $T^* \vdash \varphi$ and $T_2 \vdash \neg \varphi$. By (ii), it follows that $T_1 \vdash \neg \varphi$. Hence, $T^* \vdash \neg \varphi$, contradicting the fact that $T^*$ is consistent.

Hence, on the strong notion of empirical equivalence, the underdetermination thesis is false. Demopoulos [6, p. 373] essentially makes the same observation using Craig’s Interpolation Lemma. He observes that if one takes the content of a theory to be given by its Ramsey sentence, then one can assume without loss of generality that competing theories do not share any theoretical vocabulary, and hence that strongly empirically equivalent consistent theories are jointly consistent. Similarly, English [8, p. 460] observed that if the Ramsey sentences of two theories are jointly inconsistent, then they cannot be empirically equivalent.

Lastly, we show that

**Corollary 3.2** If $T_1$ and $T_2$ are strongly empirically equivalent consistent theories sharing no theoretical vocabulary, then they are strongly empirically equivalent to $T_1 \cup T_2$.

**Proof** Assume that $T_1$ and $T_2$ are strongly empirically equivalent consistent theories. Suppose there is a sentence $\varphi$ of their common empirical vocabulary such that $T_1 \not\vdash \varphi$ (and hence $T_2 \not\vdash \varphi$). That means $T_1 \cup \{\neg \varphi\}$ and $T_2 \cup \{\neg \varphi\}$ are consistent. Moreover, they are strongly empirically equivalent. For assume, towards contradiction, that there is sentence $\psi$ of their common empirical vocabulary such that $T_1 \cup \{\neg \varphi\} \vdash \psi$ but $T_2 \cup \{\neg \varphi\} \not\vdash \psi$ (or vice versa). That would mean $T_1 \vdash \neg \varphi \rightarrow \psi$ but $T_2 \not\vdash \neg \varphi \rightarrow \psi$ (or vice versa), contradicting our assumption. By Corollary 3.1, it now follows that $T_1 \cup T_2 \cup \{\neg \varphi\}$ is consistent. Hence, $T_1 \cup T_2 \not\vdash \varphi$. □
Together, these two corollaries suggest that, if two consistent theories are strongly empirically equivalent and we believe one of them, we might just as well believe both.

4 Weak Empirical Equivalence

The classical notion of empirical equivalence has been challenged, most famously by Van Fraassen [26, pp. 54-55]:

The empirical import of a theory cannot be isolated in this syntactical fashion, by drawing a distinction among theorems in terms of vocabulary. If that could be done, $T/E$ [the $E$-fragment of $T$, in my terminology] would say exactly what $T$ says about what is observable and what it is like, and nothing more. But any unobservable entity will differ from the observable ones in the way it systematically lacks observable characteristics. As long as we do not abjure negation, therefore, we shall be able to state in the observational vocabulary (however conceived) that there are unobservable entities, and, to some extent, what they are like. The quantum theory, Copenhagen version, implies that there are things which sometimes have a position in space, and sometimes have not. This consequence I have just stated without using a single theoretical term. Newton’s theory implies that there is something (to wit, Absolute Space) which neither has a position nor occupies a volume. Such consequences are by no stretch of the imagination about what there is in the observable world, nor about what any observable thing is like.

However, as argued by Turney [25], this objection does not pertain to every syntactic notion of empirical equivalence. Assuming, with Van Fraassen, that the notion of observability can be expressed in the empirical (observational) vocabulary of each theory, a syntactic notion of empirical equivalence can be defined that is immune to this objection. Essentially following Turney [25], such a notion will now be defined, called weak empirical equivalence. This notion will only apply to theories containing a monadic empirical predicate $O$, the intended interpretation of which is observable. This is no limitation in principle, however. In his quote, Van Fraassen assumes that, for any theory $T$ of the relevant kind, there is an $E_T$-formula $\varphi(x)$ such that the sentence $\exists x \neg \varphi(x)$ states that there are unobservable entities. Now, suppose we have two theories of the relevant kind that are, in some reasonable sense, empirically equivalent. At the very least, they would have to contain the same empirical vocabulary. Arguably, that means the empirical expressions in each theory are intended to apply to the same things. So if an empirical formula $\varphi(x)$ in one theory is intended to apply to all and only observable things, the same formula in the other theory is intended to do the same. That means we can introduce the predicate $O$ in both theories, and regard it as an abbreviation of said formula. Alternatively, we may extend the empirical vocabulary of both theories with the predicate $O$, and extend the old theories with the defining axiom $\forall x (Ox \leftrightarrow \varphi(x))$. The result will be a conservative extension with respect to the old vocabulary, where the intended interpretation of $O$ is observable. The important thing is that, either way, every formula containing
only empirical expressions and the predicate \( O \) may be regarded as an empirical formula.\(^7\) If two theories contain this predicate, then (in sense to be made precise shortly) they may entail the same things with respect to observable objects without being strongly empirically equivalent. As we shall see, one theory may entail that all objects are observable, while the other theory entails that some objects are not observable (Theorem 4.1). Intuitively, however, the theories seem to be empirically equivalent.

In order to ensure that \( O \) can be interpreted as \textit{observable}, we shall require that the theories in question are \textit{\( O \)-theories}, in the following sense:

\textbf{Definition 4.1} (\( O \)-theory) \( T \) is an \textit{\( O \)-theory} just in case (i) \( O \in E_T \), (ii) \( T \vdash \exists x O x \) and (iii) for any constant \( c \in E_T \) and predicate \( P \in E_T \) (other than identity), \( T \vdash O c \) and \( T \vdash \forall x_1...x_n (P x_1...x_n \rightarrow O x_1 \land ... \land O x_n) \).

In order to give a precise syntactic definition of when two \( O \)-theories say the same thing about observable objects, we introduce the following notion:

\textbf{Definition 4.2} (\( O \)-qualification) Let \( O \) be a monadic predicate. For any formula \( \varphi \), its \textit{\( O \)-qualification} \( Q[\varphi] \) is defined recursively:

1. \( Q[Pt_1...t_n] = Pt_1...t_n \).
2. \( Q[\neg \varphi] = \neg Q[\varphi] \).
3. \( Q[\varphi \rightarrow \psi] = Q[\varphi] \rightarrow Q[\psi] \).
4. \( Q[\forall x \varphi] = \forall x (O x \rightarrow Q[\varphi]) \).

If \( \psi = Q[\varphi] \) for some formula \( \varphi \), we say that \( \psi \) is an \textit{\( O \)-qualified formula}.

For convenience, we also introduce the notion of the \textit{\( O \)-fragment} of an \( \textit{\( O \)-theory} \), which is a subset of its \( E \)-fragment:

\textbf{Definition 4.3} (\( O \)-fragment) The \textit{\( O \)-fragment} of an \textit{\( O \)-theory} \( T \) is the set of theorems of \( T \) that are \( O \)-qualified \( E_T \)-formulas, plus the sentence \( \forall x_1...x_n (P x_1...x_n \rightarrow O x_1 \land ... \land O x_n) \) for each \( E_T \)-predicate \( P \) other than identity.

Now we define a weaker notion of empirical equivalence for \( \textit{\( O \)-theories} \):

\textbf{Definition 4.4} (Weak empirical equivalence) Two \textit{\( O \)-theories} are \textit{weakly empirically equivalent} just in case they have the same \( O \)-fragments.

Although it is perhaps intuitively clear already, in order to verify that \( O \)-qualified formulas only talk about observable objects, we introduce the following notion:

\(^7\)I would like to thank an anonymous referee for pressing me on this point.
**Definition 4.5** (O-variant) Let $M$ be a model of an $O$-theory $T$. An $L$-model $M'$ with domain $D'$ is an $O$-variant of $M$ just in case (i) $E_T \subseteq L$, (ii) $O^M \subseteq D'$, and (iii) for any $e \in E_T$, $e^{M'} = e^M$.

The following result, which we shall also make heavy use of later on, verifies our claim:

**Lemma 4.1** Let $M$ be a model of an $O$-theory $T$, and let $M'$ be an $O$-variant of $M$. Then, for any $E_T$-sentence $\varphi$, $M \vDash Q[\varphi]$ iff $M' \vDash Q[\varphi]$.

**Proof** Let $M$ be a model of an $O$-theory $T$, and let $M'$ be an $O$-variant of $M$. Let $X$ be the set of variables. Since $T \vdash Oc$ for any constant $c \in E_T$, it follows that, for any $E_T$-term $t$,

(8) For any assignment $g : X \to O^M$, $t^{M,g} = t^{M',g}$.

Moreover, since $T \vdash \forall x_1...x_n(Px_1...x_n \to Ox_1 \land ... \land Ox_n)$, it follows that, for any $E_T$-predicate $P$ other than identity,

(9) $P^M = P^{M'}$.

Now we show that, for any $E_T$-formula $\varphi$,

(10) For any assignment $g : X \to O^M$, $M, g \vDash Q[\varphi]$ iff $M', g \vDash Q[\varphi]$ by induction on the complexity of $\varphi$:

1. $M, g \vDash Q[s = t]$ iff $M, g \vDash s = t$ iff $s^{M,g} = t^{M,g}$ iff, by (8), $s^{M',g} = t^{M',g}$ iff $M', g \vDash Q[s = t]$.
2. $M, g \vDash Q[P_{t_1...t_n}]$ iff $M, g \vDash P_{t_1...t_n}$ iff $\langle t_1^{M,g}, ..., t_n^{M,g} \rangle \in P^M$ iff, by (8) and (9), $\langle t_1^{M',g}, ..., t_n^{M',g} \rangle \in P^{M'}$ iff $M', g \vDash P_{t_1...t_n}$ iff $M', g \vDash Q[P_{t_1...t_n}]$.
3. $M, g \vDash Q[\neg \varphi]$ iff $M, g \vDash \neg Q[\varphi]$ iff $M, g \nmid Q[\varphi]$ iff (with (10) as induction hypothesis) $M', g \nmid Q[\varphi]$ iff $M', g \vDash Q[\neg \varphi]$.
4. $M, g \vDash Q[\varphi \rightarrow \psi]$ iff $M, g \vDash Q[\varphi] \rightarrow Q[\psi]$ iff $M, g \nmid Q[\varphi]$ or $M, g \vDash Q[\psi]$ iff (with (10) as induction hypothesis) $M', g \nmid Q[\varphi]$ or $M', g \vDash Q[\psi]$ iff $M', g \vDash Q[\varphi] \rightarrow Q[\psi]$ iff $M', g \vDash Q[\varphi] \rightarrow Q[\psi]$.
5. $M, g \vDash Q[\forall x \varphi]$ iff $M, g \vDash \forall x(Ox \rightarrow Q[\varphi])$ iff, for any $d \in O^M$, $M, g[d/x] \vDash Q[\varphi]$ iff (with (10) as induction hypothesis) for any $d \in O^M$, $M', g[d/x] \vDash Q[\varphi]$ iff $M', g \vDash \forall x(Ox \rightarrow Q[\varphi])$ iff $M', g \vDash Q[\forall x \varphi]$.

It follows that, for any $E_T$-sentence $\varphi$, $M \vDash Q[\varphi]$ iff $M' \vDash Q[\varphi]$. \hfill \square

Now we are in a position to prove the main claim of this section:

**Theorem 4.1** For any $O$-theory $T$ with $O$-fragment $T'$, $T' \cup \{\forall x Ox\}$ and $T' \cup \{\neg \forall x Ox\}$ are $O$-theories that are weakly empirically equivalent to $T$. 

\[\square \text{ Springer} \]
Proof Let $T$ be an $O$-theory, and let $T'$ be its $O$-fragment. Let $M$ be a model of $T'$, and let $M'$ and $M''$ be $O$-variants of $M$ with domains $D' = O^M$ and $D'' \neq O^M$ respectively. By Lemma 4.1, it follows that $M'$ and $M''$ are models of $T'$. Hence, $T' \not\vdash \forall x O x$ and $T' \not\vdash \neg \forall x O x$. Define $O$-theories $T_1 = T' \cup \{ \forall x O x \}$ and $T_2 = T' \cup \{ \neg \forall x O x \}$. It remains to show that $T_1$ and $T_2$ are weakly empirically equivalent to $T'$, and hence to $T$. Assume, towards contradiction, for some closed $O$-qualified formula $\varphi$, that $T_1 \cup \{ \forall x O x \} \vdash \varphi$ but $T_1 \not\vdash \varphi$. Let $M$ be a model of $T'$ such that $M \not\models \varphi$, and let $M'$ be an $O$-variant of $M$ with domain $D = O^M$. Since $\varphi$ is an $O$-qualified formula, it follows by Lemma 4.1 that $M' \not\models \varphi$. Moreover, we have $M' \models \forall x O x$. By assumption, it follows that $M' \models \varphi$, yielding a contradiction. Next assume, towards contradiction, for some closed $O$-qualified formula $\varphi$, that $T_2 \cup \{ \neg \forall x O x \} \vdash \varphi$ but $T_2 \not\vdash \varphi$. Let $M$ be a model of $T'$ such that $M \not\models \varphi$, and let $M'$ be an $O$-variant of $M$ with domain $D' \neq O^M$. By Lemma 4.1, it follows that $M' \not\models \varphi$. Moreover, $M' \models \neg \forall x O x$. By assumption, it follows that $M' \models \varphi$, yielding a contradiction.

Assuming with Van Fraassen that any theory of the relevant kind can be construed as an $O$-theory, the result establishes that weak empirical equivalence supports the underdetermination thesis.

5 Van Fraassen Equivalence

In order to define Van Fraassen’s notion of empirical equivalence, we introduce the following notion:

Definition 5.1 ($O$-part) Let $M$ be a model of an $O$-theory $T$. An $L$-model $M'$ with domain $D'$ is the $O$-part of $M$ just in case (i) $L = E_T$, (ii) $D' = O^M$, and (iii) for any $e \in E_T$, $e^{M'} = e^M$.

Empirical equivalence is then defined as follows:

Definition 5.2 (Van Fraassen equivalence) Two $O$-theories $T$ and $T'$ are Van Fraassen equivalent just in case, for any model $M$ of $T$, there’s a model $M'$ of $T'$ whose $O$-part is isomorphic to the $O$-part of $M$, and vice versa.

Theorem 5.1 For any $O$-theory $T$ with $O$-fragment $T'$, $T' \cup \{ \forall x O x \}$ and $T' \cup \{ \neg \forall x O x \}$ are Van Fraassen equivalent to $T'$.

Proof Let $M$ be a model of $T'$. By Lemma 4.1, $M$ has $O$-variants $M'$ and $M''$ that are models of $T' \cup \{ \forall x O x \}$ and $T' \cup \{ \neg \forall x O x \}$, respectively. Since the $O$-parts of $M$, $M'$ and $M''$ are identical, they are isomorphic. Likewise, for any model of $O$ of $T' \cup \{ \forall x O x \}$ ($T' \cup \{ \neg \forall x O x \}$) (there are $O$-variants satisfying $T'$ and $T' \cup \{ \neg \forall x O x \}$)
Observe that this does not yet establish that the notion of Van Fraassen equivalence supports the underdetermination thesis. We include the result here merely to show that Van Fraassen equivalence does not entail strong empirical equivalence. It does, however, entail weak empirical equivalence. To see that, we introduce to following notion:

Definition 5.3 (Elementary Van Fraassen equivalence) Two $O$-theories $T$ and $T'$ are elementary Van Fraassen equivalent just in case, for any model $\mathcal{M}$ of $T$, there’s a model $\mathcal{M}'$ of $T'$ whose $O$-part is elementary equivalent to the $O$-part of $\mathcal{M}$, and vice versa.

Elementary equivalence (making the same first-order sentences true) is strictly weaker than isomorphism. Clearly, Van Fraassen equivalence entails elementary Van Fraassen equivalence. We go on to establish the identity between elementary Van Fraassen equivalence and weak empirical equivalence:

Lemma 5.1 Let $\mathcal{M}'$ be the $O$-part of a model $\mathcal{M}$ of some $O$-theory $T$. Then, for any $E_T$-sentence $\varphi$, $\mathcal{M}' \models \varphi \iff Q[\varphi]$.

Proof It is straightforward to show, by induction on the complexity of $\varphi$, that for any $E_T$-formula $\varphi$ and assignment $g$, $\mathcal{M}', g \models \varphi$ iff $\mathcal{M}', g \models Q[\varphi]$.

Theorem 5.2 Two $O$-theories are weakly empirically equivalent just in case they are elementary Van Fraassen equivalent.

Proof Assume that $T_1$ and $T_2$ are $O$-theories with the same empirical vocabulary $E_T$ that are not weakly empirically equivalent. Then there is an $E_T$-sentence $\varphi$ such that $T_1 \vdash Q[\varphi]$ and $T_2 \nvdash Q[\varphi]$. Thus, let $\mathcal{M}_2$ be a model of $T_2$ such that $\mathcal{M}_2 \nvdash Q[\varphi]$, and let $\mathcal{M}_1$ be a model of $T_1$. Let $\mathcal{M}_1'$ and $\mathcal{M}_2'$ be the $O$-parts of $\mathcal{M}_1$ and $\mathcal{M}_2$ respectively. Since $\mathcal{M}_2 \nvdash Q[\varphi]$, it follows by Lemma 4.1 that $\mathcal{M}_2' \nvdash Q[\varphi]$. Moreover, since $T_1 \vdash Q[\varphi]$, we have $\mathcal{M}_1 \models Q[\varphi]$. By Lemma 4.1, $\mathcal{M}_1' \models Q[\varphi]$. Hence, for any model $\mathcal{M}_1$ of $T_1$, its $O$-part $\mathcal{M}_1'$ is not elementary equivalent to $\mathcal{M}_2'$, which means that $T_1$ and $T_2$ are not elementary Van Fraassen equivalent.

For the other direction, assume that $T_1$ and $T_2$ are weakly empirically equivalent $O$-theories with common empirical vocabulary $E_T$. Let $\mathcal{M}_1$ be a model of $T_1$ with $O$-part $\mathcal{M}_1'$. Let $T'$ be the set of $E_T$-sentences that are true in $\mathcal{M}_1'$. Assume, towards contradiction, that there is no model $\mathcal{M}_2$ of $T_2$ such that the $O$-part of $\mathcal{M}_2$ is elementary equivalent to $\mathcal{M}_1'$. Let $\mathcal{M}_2$ be a model of $T_2$ with $O$-part $\mathcal{M}_2'$. By assumption, there is $\varphi \in T'$ such that $\mathcal{M}_2' \models \neg \varphi$. By Lemma 5.1, $\mathcal{M}_2' \models Q[\neg \varphi]$, and by Lemma 4.1, $\mathcal{M}_2 \models Q[\neg \varphi]$. Hence, since $\mathcal{M}_2$ is arbitrary, $T_2 \vdash Q[\neg \varphi]$. Since $T_1$ and $T_2$ are weakly empirically equivalent, it follows that $T_1 \vdash Q[\neg \varphi]$. That means $\mathcal{M}_1 \models Q[\neg \varphi]$. By Lemma 4.1, $\mathcal{M}_1 \models Q[\neg \varphi]$. Hence, by Lemma 5.1, $\mathcal{M}_1' \models \neg \varphi$, contradicting the fact that $\varphi \in T'$.
The following result shows that weak empirical equivalence does not entail Van Fraassen equivalence:

**Theorem 5.3** Some $O$-theory is not Van Fraassen equivalent to its $E$-fragment (nor to its $O$-fragment).

**Proof** Let $\langle \mathbb{N}, < \rangle$ be the natural number structure with the binary relation $<$, and let $\langle \emptyset, \langle \rangle \rangle$ be a structure isomorphic to it under bijection $h : \mathbb{N} \to \emptyset$, where $\emptyset \cap \mathbb{N} = \emptyset$. Define a model $\mathcal{M}^*$ with domain $D = \mathbb{N} \cup \emptyset$ interpreting the vocabulary $L^* = \{O, <, O, \preceq\}$ as $O_{\mathcal{M}^*} = \emptyset$, $\preceq_{\mathcal{M}^*} = <$, $\preceq_{\mathcal{M}^*} = <$, $0_{\mathcal{M}^*} = 0$ and, for each $d \in D$,

$$f^*_{\mathcal{M}}(d) = \begin{cases} h(d) & \text{if } d \in \mathbb{N} \\ d & \text{otherwise} \end{cases}$$

$$s^*_{\mathcal{M}}(d) = \begin{cases} d + 1 & \text{if } d \in \mathbb{N} \\ d & \text{otherwise} \end{cases}$$

Let $Th(\mathcal{M}^*)$ be the set of $L^*$-sentences that are true in $\mathcal{M}^*$. Let $c$ be a new constant, let $L = L^* \cup \{c\}$ and $\Gamma = \{0 < c, s(0) < c, s(s(0)) < c, \ldots\}$. Let $T = Th(\mathcal{M}^*) \cup \Gamma$ and $E_T = \{O, \langle\rangle\}$. It follows that

(11) a. $T \vdash \forall x O f(x)$.

b. $T \vdash \forall x \forall y (O x \land O y \land x < y \rightarrow x \neq y)$.

c. For any natural numbers $m < n$, $T \vdash f(m) < f(n) < f(c)$, where $m$ and $n$ are the corresponding numerals.

Clearly, every finite subset of $T$ has a model. By compactness, $T$ has a model. Since $Th(\mathcal{M}^*)$ is complete with respect to $L^*$, and $Th(\mathcal{M}^*) \subseteq T$, that means the $E$-fragments of $T$ is complete and identical to the $E$-fragment of $Th(\mathcal{M}^*)$. Let $\mathcal{M}$ be a model of $T$. By (11), there are infinitely many elements in $O_{\mathcal{M}}$ between $f(0)_{\mathcal{M}}$ and $f(c)_{\mathcal{M}}$ as ordered by $\preceq_{\mathcal{M}}$ (one for each natural number $n > 0$). This is not the case for any pair of elements in $O^*_{\mathcal{M}}$ as ordered by $\preceq^*_\mathcal{M}$. Hence, the $O$-part of $\mathcal{M}$ is not isomorphic to the $O$-part of $\mathcal{M}^*$. Finally, since $\mathcal{M}^*$ is a model also of the $E$-fragment of $Th(\mathcal{M}^*)$, and hence to the $E$- and $O$-fragment of $T$, it follows that $T$ is not Van Fraassen equivalent to its $E$-fragment (nor to its $O$-fragment).

**Corollary 5.1** Some strongly (and hence weakly) empirically equivalent $O$-theories are not Van Fraassen equivalent.

**Proof** Follows from Theorem 5.3 and the fact that every theory is strongly empirically equivalent to its $E$-fragment.

**Remark 5.1** With a little bit of tweaking, we note that the construction used in the proof of Theorem 5.3 shows that the proof of Theorem 5 in Turney [25, p. 437-438] is flawed. In fact, his claim is false. Although Turney assumes that theories
are formulated in a first-order language containing only predicates (each of which has non-zero arity), we can construct a counterexample to his claim satisfying this assumption. Consider the isomorphic structures \( \langle \mathbb{N}, < \rangle \) and \( \langle \mathbb{D}, \prec \rangle \) given in our proof. Define a model \( \mathcal{M}^* \) with domain \( D = \mathbb{N} \cup \mathbb{O} \) interpreting the vocabulary \( L^* = \{O, <, <, Z, S, F\} \) as \( O^{\mathcal{M}^*} = \mathbb{O}, \prec^{\mathcal{M}^*} = <, Z^{\mathcal{M}^*} = \{0\}, S^{\mathcal{M}^*} \) as the successor relation on \( \mathbb{N} \) and \( F^{\mathcal{M}^*} \) as the graph of the bijection \( h : \mathbb{N} \rightarrow \mathbb{O} \) under which the structures are isomorphic. Let \( Th(\mathcal{M}^*) \) be the set of \( L^* \)-sentences that are true in \( \mathcal{M}^* \). Let \( C \) be a new unary predicate, let \( L = L^* \cup \{C\} \) and let

\[
\Gamma = \{\exists x(Cx \land \forall y(Cy \rightarrow y = x)),
\forall x_0 \forall y(Zx_0 \land Cy \rightarrow x_0 < y),
\forall x_0 \forall x_1 \forall y(Zx_0 \land Sx_0x_1 \land Cy \rightarrow x_1 < y),
\forall x_0 \forall x_1 \forall x_2 \forall y(Zx_0 \land Sx_0x_1 \land Sx_1x_2 \land Cy \rightarrow x_2 < y), \ldots\}
\]

Let \( T = Th(\mathcal{M}^*) \cup \Gamma \) and \( E_T = \{O, <\} \). By compactness, \( T \) has a model. Moreover, for each model \( \mathcal{M} \) of \( T \), \( \prec^\mathcal{M} \) contains infinitely many objects between some pair of objects. Let \( L' = \{<\} \), and define the set of \( L' \)-sentences \( T' \) such that, for each \( L' \)-sentence \( \varphi, \varphi \in T' \) iff \( T \vdash Q[\varphi] \). Since \( T \) and \( Th(\mathcal{M}^*) \) have the same \( O \)-fragment (because \( Th(\mathcal{M}^*) \) is complete), and the \( O \)-part of \( \mathcal{M}^* \) is a model of the \( O \)-fragment of \( Th(\mathcal{M}^*) \) according to Lemma 4.1, it follows by Lemma 5.1 that the \( O \)-part of \( \mathcal{M}^* \) is a model of \( T' \). By Definition 14 in Turney \[25, p. 436\], \( T' \) is a theory fragment of \( T \) satisfying the conditions of his claim. But since it is not the case that \( \prec^{\mathcal{M}^*} \) contains infinitely many objects between some pair of objects, the \( O \)-part of \( \mathcal{M}^* \) is not isomorphic to the \( O \)-part of any model of \( T \). A fortiori, the \( O \)-part of \( \mathcal{M}^* \) is not a so called model fragment of any model of \( T \), per Definition 16 in Turney \[25, p. 437\]. In effect, Turney uses his Theorem 5 to show that weak empirical equivalence entails Van Fraassen equivalence. As we have seen, this is not true.

To sum up, weak empirical equivalence does not entail Van Fraassen equivalence. Moreover, both strong empirical equivalence and Van Fraassen equivalence entail weak empirical equivalence, but neither of the two entails the other. The question is whether Van Fraassen equivalence supports the underdetermination thesis, i.e. whether the following conjecture is true:

**Conjecture 5.1** For any \( O \)-theory \( T \), there are \( O \)-theories \( T_1 \) and \( T_2 \) Van Fraassen equivalent to \( T \) such that \( T_1 \) and \( T_2 \) do not share any theoretical vocabulary and \( T_1 \cup T_2 \) is inconsistent.

In view of Theorem 5.3, the answer is not obvious. For some \( O \)-theories, it is certainly possibly to find Van Fraassen equivalent rivals that are jointly inconsistent. Take, for instance, the \( O \)-fragment \( T' \) of an arbitrary \( O \)-theory \( T \). By Theorem 5.1, \( T' \) will have Van Fraassen equivalent rivals \( T' \cup \{\forall x Ox\} \) and \( T' \cup \{\neg \forall x Ox\} \). However, by Theorem 5.3, these need not be Van Fraassen equivalent to \( T \).
6 Indispensability

Quine and Putnam famously argued that we should believe that theoretical terms refer because these terms are indispensable for the practical formulation of empirically adequate theories. In response to the second part of their claim, the following results are sometimes mentioned:

**Theorem 6.1** (Craig) *Every recursively enumerable set of sentences is axiomatizable (i.e. equivalent to some recursive set of sentences).*

**Remark 6.1** Not every recursively enumerable set of sentences is *finitely* axiomatizable, however. To see why, let $T$ be any finitely axiomatizable consistent theory having only infinite models (e.g. Robinson arithmetic), and let $T_\equiv$ be the set of theorems of $T$ containing no non-logical vocabulary. Then $T_\equiv$ is recursively enumerable.

Let $\Gamma_1$ be the set of sentences saying, for each natural number $n$, that there are at least $n$ objects. We know that $\Gamma_1$ is complete with respect to the logical vocabulary, but not finitely axiomatizable. Since $T_\equiv$ is consistent and $\Gamma_1 \subseteq T_\equiv$, it follows that $T_\equiv$ is equivalent to $\Gamma_1$, and hence not finitely axiomatizable.

**Lemma 6.1** *The $E$-fragment of an axiomatizable theory is recursively enumerable.*

**Corollary 6.1** *For any axiomatizable theory $T$, there is a strongly empirically equivalent axiomatizable theory $T'$ in the language of $ET$.*

The relevance of this result was noted, for instance, by Hempel [10]. It shows that, insofar as axiomatizability is the relevant notion of practicality, the strong notion of empirical equivalence supports the dispensability thesis. On the other hand, if *finite* axiomatizability is what matters for practicality, it follows from Remark 6.1 that strong empirical equivalence does not support the thesis (as witnessed by $T_\equiv$ considered as the $E$-fragment of $T$). For various other reasons, Hempel resists the conclusion that theoretical terms are dispensable. List [15] also resists the dispensability thesis in spite of Corollary 6.1, but rather because he takes the strong notion of empirical equivalence to be irrelevant (citing Van Fraassen). More specifically, he argues that the relevant empirical consequences of an axiomatizable theory may not be recursively enumerable. In response to List, we make the following observation:

**Lemma 6.2** *The $O$-fragment of an axiomatizable theory is recursively enumerable.*

Using Theorem 6.1 and Lemma 6.2, together with Theorem 4.1, it follows that

**Corollary 6.2** *For any axiomatizable $O$-theory $T$ with $O$-fragment $T'$, $T' \cup \{\forall x O x\}$ and $T' \cup \{\neg\forall x O x\}$ are axiomatizable $O$-theories in the language of $ET$ that are weakly empirically equivalent to $T$.*

The result shows that, assuming weak empirical equivalence to be the relevant notion of empirical equivalence, and axiomatizability to be the relevant notion of
practicality, the dispensability thesis is true. Again, in view of Remark 6.1, assuming finite axiomatizability to be the relevant notion of practicality instead, it is not hard to construct a counterexample to dispensability also on the weak notion of empirical equivalence.

Lastly, we show that Van Fraassen equivalence together with axiomatizability does not support indispensability. To do that, we shall use the following result, as reported by Keisler [11, p. 177, Corollary 10.8]:

**Theorem 6.2** (Keisler-Shelah isomorphism theorem) Two models are elementary equivalent just in case, for some ultrafilter, the corresponding ultrapowers of the two models are isomorphic.

Using that result, we first prove the following lemma:

**Lemma 6.3** Let $L$ be a first-order vocabulary containing only predicates, and let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two elementary equivalent $L$-models with the same domain $D$. Let $\mathcal{M}'_1$ and $\mathcal{M}'_2$ be two $L$-models with the same domain $D'$ such that $D \subseteq D'$. Assume that, for each predicate $P \in L$, $P^{\mathcal{M}_1'} = P^{\mathcal{M}_1}$ and $P^{\mathcal{M}_2'} = P^{\mathcal{M}_2}$. Then $\mathcal{M}_1'$ and $\mathcal{M}_2'$ are elementary equivalent.

**Proof** Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two elementary equivalent $L$-models with the same domain $D$. By Theorem 6.2, that means there is an ultrafilter $U$ over a non-empty set $I$ such that the corresponding ultrapowers of $\mathcal{M}_1$ and $\mathcal{M}_2$ are isomorphic, i.e. $\Pi_U \mathcal{M}_1 \cong \Pi_U \mathcal{M}_2$. By definition of ultrapowers, the domain $D_U$ of both $\Pi_U \mathcal{M}_1$ and $\Pi_U \mathcal{M}_2$ is given by

$$D_U = \{[f]_U : f \in D^I\}$$

where, for each $f \in D^I$,

$$[f]_U = \{g \in D^I : \{i \in I : f(i) = g(i)\} \in U\}$$

Moreover, $\Pi_U \mathcal{M}_1$ and $\Pi_U \mathcal{M}_2$ are defined so that, for each $n$-place predicate $P$ and $f_1, \ldots, f_n \in D^I$,

$$P^{\Pi_U \mathcal{M}_1}([f_1]_U, \ldots, [f_n]_U) \text{ iff } \{i \in I : P^{\mathcal{M}_1}(f_1(i), \ldots, f_n(i))\} \in U$$

and

$$P^{\Pi_U \mathcal{M}_2}([f_1]_U, \ldots, [f_n]_U) \text{ iff } \{i \in I : P^{\mathcal{M}_2}(f_1(i), \ldots, f_n(i))\} \in U.$$  

It can be shown that $D_U$ is a partitioning of $D^I$ into equivalence classes $[f]_U$ for each $f \in D^I$. The models are well-defined because it can be shown that, for any $f_1, \ldots, f_n \in D^I$ and $g_1 \in [f_1]_U, \ldots, g_n \in [f_n]_U$, $\{i \in I : P^{\mathcal{M}}(f_1(i), \ldots, f_n(i))\} \in U$ iff $\{i \in I : P^{\mathcal{M}}(g_1(i), \ldots, g_n(i))\} \in U$.

Let $\mathcal{M}_1'$ and $\mathcal{M}_2'$ be two $L$-models with the same domain $D'$ such that $D \subseteq D'$. Assume that, for each predicate $P \in L$, $P^{\mathcal{M}_1'} = P^{\mathcal{M}_1}$ and $P^{\mathcal{M}_2'} = P^{\mathcal{M}_2}$. The domain $D'_U$ of both $\Pi_U \mathcal{M}_1'$ and $\Pi_U \mathcal{M}_2'$ is given by

$$D'_U = \{[f]_U : f \in D'^I\}$$
where, for each \( f \in D^I \), the equivalence class \([f]_U\) is given by

\[
[f]_U = \{ g \in D^I : \{ i \in I : f(i) = g(i) \} \in U \}
\]

(16) \( \Pi U M'_1 \) and \( \Pi U M'_2 \) are defined so that, for each \( n \)-place predicate \( P \) and \( f_1, \ldots, f_n \in D^I \),

\[
P_{\Pi U M'_1}([f_1]_U, \ldots, [f_n]_U) \text{ iff } \{ i \in I : P_{M'_1}(f_1(i), \ldots, f_n(i)) \} \in U.
\]

(17) a. \( P_{\Pi U M'_1}([f_1]_U, \ldots, [f_n]_U) \text{ iff } \{ i \in I : P_{M'_1}(f_1(i), \ldots, f_n(i)) \} \in U. \)

b. \( P_{\Pi U M'_2}([f_1]_U, \ldots, [f_n]_U) \text{ iff } \{ i \in I : P_{M'_2}(f_1(i), \ldots, f_n(i)) \} \in U. \)

Now, since \( \Pi U M'_1 \cong \Pi U M'_2 \), there is a bijection \( \alpha : D_U \rightarrow D_U \) such that, for each \( n \)-place predicate \( P \) and \( [f_1]_U, \ldots, [f_n]_U \in D_U \),

\[
P_{\Pi U M'_1}([f_1]_U, \ldots, [f_n]_U) \text{ iff } P_{\Pi U M'_2}([\alpha([f_1]_U), \ldots, \alpha([f_n]_U)]). \]

Assuming the Axiom of Choice,\(^8\) let \( r : D_U \rightarrow D^I \) be a choice function picking out a representative of each equivalence class, and define \( a : D^I \rightarrow D^I \) so that it maps every element \( f \in D^I \) to the representative of \( \alpha([f]_U) \), i.e. \( a(f) = r(\alpha([f]_U)) \). It follows by (18) that, for each \( n \)-place predicate \( P \) and \( f_1, \ldots, f_n \in D^I \),

\[
P_{\Pi U M'_1}([f_1]_U, \ldots, [f_n]_U) \text{ iff } P_{\Pi U M'_2}([a(f_1)]_U, \ldots, [a(f_n)]_U). \]

Hence, by (14), for any \( f_1, \ldots, f_n \in D^I \),

\[
\{ i \in I : P_{M'_1}(f_1(i), \ldots, f_n(i)) \} \in U \text{ iff } \{ i \in I : P_{M'_2}(a(f_1)(i), \ldots, a(f_n)(i)) \} \in U.
\]

Moreover, since \( P_{M'_1} = P_{M'_1} \) and \( P_{M'_2} = P_{M'_2} \), we have for each \( f_1, \ldots, f_n \in D^I \) and \( i \in I \),

\[
a_{M'_1}(f_1(i), \ldots, f_n(i)) \text{ iff } P_{M'_1}(f_1(i), \ldots, f_n(i)).
\]

b. \( P_{M'_2}(f_1(i), \ldots, f_n(i)) \text{ iff } P_{M'_2}(f_1(i), \ldots, f_n(i)). \)

Observe that, by (13) and (16), for any \( f, g \in D^I \),

\[
a' : D^I \rightarrow D^I \text{ such that, for each } f \in D^I,
\]

\[
a'(f) = \begin{cases} a(g) & \text{if } g \in [f]_U \cap D^I \\ f & \text{if } [f]_U \cap D^I = \emptyset \end{cases}
\]

and define \( \alpha' : D'_U \rightarrow D'_U \) such that, for each \( f \in D'^I \),

\[
\alpha'([f]_U) = [a'(f)]_U.
\]

We show that \( \alpha' \) is a bijection. To show that \( \alpha' \) is surjective, assume that \( [f]_U \in D'_U \). We get two cases:

\(^8\)The Axiom of Choice is assumed anyway in the proof of Theorem 6.2 by Shelah [23].
1. There is \( g \in D^1 \) such that \( g \in [f]_U \). Then \([f]_U = [g]_U\). By (22-b), \([g]_U \subseteq [f]_U\). Moreover, since \( \alpha \) is a bijection, there is \( h \in D^1 \) such that \( \alpha'(h)_U = [a(h)]_U \). By (22-b), \([g]_U \subseteq \alpha'(h)_U = [a(h)]_U\). Hence, \( \alpha'(h)_U \cap [f]_U \neq \emptyset \), in which case \( \alpha'(h)_U = [f]_U\).

2. There is no \( g \in D^1 \) such that \( g \in [f]_U \). Then \( \alpha'(f)_U = [a'(f)]_U = [f]'_U\).

To show that \( \alpha' \) is injective, assume that \( \alpha'([f]'_U) = \alpha'([g]'_U) \). Clearly, \([f]'_U \cap D^1 = \emptyset \) iff \([g]'_U \cap D^1 = \emptyset \). Otherwise \( \alpha'([f]'_U) = \alpha'([g]'_U) \), since only one of them will contain a member of \( D^1 \). Hence, we get two cases:

1. There is \( f' \in [f]'_U \cap D^1 \) and \( g' \in [g]'_U \cap D^1 \). Hence, \([f]'_U = [f]'_U \) and \([g]'_U = [g]'_U \). Then \( \alpha'([f]'_U) = [a(f')]_U \) and \( \alpha'([g]'_U) = [a(g')]_U \). Hence, \([a(f')]_U = [a(g')]_U \). By (22-c), \([a(f')]_U = [a(g')]_U \). Hence, \( \alpha([f]'_U) = \alpha([g]'_U) \). Since \( \alpha \) is a bijection, \([f]'_U = [g]'_U \). By (22-c), \([f]'_U = [g]'_U\).

2. \([f]'_U \cap D^1 = [g]'_U \cap D^1 = \emptyset \). Then \( \alpha'([f]'_U) = [f]'_U \) and \( \alpha'([g]'_U) = [g]'_U \). Hence, \([f]'_U = [g]'_U\).

Finally we show that, for any \( f_1, \ldots, f_n \in D^1 \), \( P\Pi_U M'_1([f_1]'_U, \ldots, [f_n]'_U) \) iff \( P\Pi_U M'_2(\alpha'([f_1]'_U), \ldots, \alpha'([f_n]'_U)) \). We get two cases:

1. There are \( g_1, \ldots, g_n \in D^1 \) such that \( g_1 \in [f_1]'_U \), \( \ldots \), \( g_n \in [f_n]'_U \). That means \([f_1]'_U = [g_1]'_U \), \( \ldots \), \([f_n]'_U = [g_n]'_U \). Hence, \( P\Pi_U M'_1([f_1]'_U, \ldots, [f_n]'_U) \) iff \( P\Pi_U M'_2(\alpha'([f_1]'_U), \ldots, \alpha'([f_n]'_U)) \).

2. For some \( 1 \leq k \leq n \), there is no \( g \in D^1 \) such that \( g \in [f_k]'_U \), i.e. no \( g \in D^1 \) such that \( [i \in I : f_k(i) = g(i)] \subseteq D \). It follows that \( [i \in I : f_k(i) \in D] \neq U \) and, since \( \alpha'(f_k(i)) = f_k \), \( [i \in I : \alpha'(f_k(i)) \in D] \neq U \). Since \( P\Pi_U M'_1(f_1(i), \ldots, f_n(i)) \) entails \( f_k(i) \in D \), and \( P\Pi_U M'_2(\alpha'(f_1(i)), \ldots, \alpha'(f_n(i))) \) entails \( \alpha'(f_k(i)) \in D \), we have \( [i \in I : f_1(i), \ldots, f_n(i)] \neq U \) and \([i \in I : f_k(i), \ldots, f_n(i)] \neq U \) by upwards closure of \( U \). Hence, by (17) and (24), \( P\Pi_U M'_1([f_1]'_U, \ldots, [f_n]'_U) \) iff \( P\Pi_U M'_2(\alpha'([f_1]'_U), \ldots, \alpha'([f_n]'_U)). \)

Now we have shown that \( \Pi_U M'_1 \cong \Pi_U M'_2 \). By Theorem 6.2 it follows that \( M'_1 \) and \( M'_2 \) are elementarily equivalent.

This lemma has the following important corollary:
Corollary 6.3  If an $O$-theory $T$ is not Van Fraassen equivalent to its $O$-fragment, and $E_T$ only contains predicates, then $T$ is not Van Fraassen equivalent to any $O$-theory in the language of $E_T$.

Proof  Let $T$ be an $O$-theory such that $E_T$ only contains predicates. Assume that $T$ is not Van Fraassen equivalent to its $O$-fragment $T_O$. That means $T_O$ has a model $M$ with $O$-part $M_O$ no $O$-variant of which is a model of $T$. Assume, towards contradiction, that $T$ is Van Fraassen equivalent to some $O$-theory $T'$ in the language of $E_T$. It follows that no $O$-variant of $M_O$ is a model of $T'$. Since Van Fraassen equivalence entails weak empirical equivalence, it follows by Theorem 5.2 that $T'$ is elementary Van Fraassen equivalent to $T_O$. Hence, $T'$ has a model $M'$ whose $O$-part $M'_O$ is elementary equivalent to $M_O$. By Löwenheim-Skolem, we can assume without loss of generality that $M'_O$ has the same domain as $M_O$. By Lemma 6.3, that means $M'$ is elementary equivalent to some $O$-variant of $M_O$. Hence, some $O$-variant of $M_O$ is a model of $T$, contradicting what we have already established.

Now we can prove that Van Fraassen’s notion of empirical equivalence, together with axiomatizability as the criterion for practicality, does not support the dispensability thesis:

Theorem 6.3  For some axiomatizable $O$-theory $T$, there is no Van Fraassen equivalent $O$-theory in the language of $E_T$.

Proof  Let $L^* = \{O, \prec, <, 0, s, f\}$, and let $T^*$ consist of the following groups of axioms. The first group says that $\prec$ is a strict total order on observable objects:

(25)  a. $\forall x \forall y (x \prec y \rightarrow O x \land O y)$
     b. $\forall x \forall y (x \prec y \rightarrow x \neq y)$
     c. $\forall x \forall y \forall z (x \prec y \land y \prec z \rightarrow x \prec z)$
     d. $\forall x \forall y (O x \land O y \rightarrow x \prec y \lor x = y \lor y \prec x)$

The second group says that $<$ is a strict total order on unobservable objects:

(26)  a. $\forall x \forall y (x < y \rightarrow \neg O x \land \neg O y)$
     b. $\forall x \forall y (x < y \rightarrow x \neq y)$
     c. $\forall x \forall y \forall z (x < y \land y < z \rightarrow x < z)$
     d. $\forall x \forall y (\neg O x \land \neg O y \rightarrow x < y \lor x = y \lor y < x)$

The third group says that $0$ is the smallest unobservable object:

(27)  a. $\neg O 0$
     b. $\exists x (x < 0)$

The fourth group says that $s$ is the successor function on the unobservable objects:

(28)  a. $\forall x (\neg O x \rightarrow x < s(x))$
     b. $\forall x \neg \exists y (x < y \land y < s(x))$
The fifth group says that $f$ is an order-preserving bijection from unobservable to observable objects:

\[(29)\]
\[
a. \ \forall x O f(x) \\
b. \ \forall x \forall y (\neg O x \land \neg O y \land x \neq y \rightarrow f(x) \neq f(y)) \\
c. \ \forall x (O x \rightarrow \exists y (x = f(y))) \\
d. \ \forall x \forall y (f(x) < f(y) \leftrightarrow x < y)
\]

Let $c$ be a new constant, let $\Gamma_1 = \{0 < c, s(0) < c, s(s(0)) < c, \ldots\}$, let $L = L^* \cup \{c\}$ and $T = T^* \cup \Gamma$. Let $E_T = \{O, <\}$. By $(29-a)$ and $(25-a)$, $T$ is an $O$-theory. By compactness, it has a model. Observe that, for any model $M$ of $T$, $\prec^M$ contains infinitely many objects between some pair of objects, namely $f(0)^M$ and $f(c)^M$.

Next, we show that

\[(30)\]

For any $L^*$-sentence $\varphi$, $T \vdash \varphi$ iff $T^* \vdash \varphi$.

Since $T^* \subseteq T$, we only need to show the direction from left to right. Assume, towards contradiction, that there is an $L^*$-sentence $\varphi$ such that $T \vdash \varphi$ but $T^* \nvdash \varphi$. That means there is a finite subset $\Gamma' \subseteq \Gamma$ such that $T^* \cup \Gamma' \cup \{\neg \varphi\}$ does not have a model but $T^* \cup \{\neg \varphi\}$ has. Let $\mathcal{M}$ be a model of $T^* \cup \{\neg \varphi\}$. Let $t$ be the longest term such that $t < c \in \Gamma'$. Let $\mathcal{M}'$ be an expansion of $\mathcal{M}$ such that $c^\mathcal{M}' = s(t)^\mathcal{M}$. Then $\mathcal{M}'$ is a model of $T^* \cup \Gamma'$. By locality, $\mathcal{M}'$ is also a model of $\neg \varphi$. Hence, $\mathcal{M}'$ is a model of $T^* \cup \Gamma' \cup \{\neg \varphi\}$, contrary to our assumption.

It follows from $(30)$ that the $O$-fragments of $T$ and $T^*$ are identical. But $(\mathbb{N}, <)$ is a model of this fragment, and non-isomorphic to the $O$-part of any model of $T$. Hence, $T$ is not Van Fraassen equivalent to its $O$-fragment. By Corollary 6.3, it follows that $T$ is not Van Fraassen equivalent to any $O$-theory in the language of $E_T$.

**Remark 6.2** In a way, this result strengthens that of Melia [17], who shows (in my terminology) that some axiomatizable theory $T^*$ is not Van Fraassen equivalent to its $E$-fragment $T$. Melia’s aim is essentially to show that the trivial nominalization-strategy suggested by Craig’s theorem does not work if one adopts Van Fraassen’s notion of empirical equivalence. My aim is to show that, for some theories, no nominalization-strategy works in that case. It is not obvious from his result that no other $E_{T^*}$-theory is Van Fraassen equivalent to his theory $T^*$, but it can be shown using Corollary 6.3.

As mentioned in the beginning, all these result pertain to first-order theories only. I suspect that, for any theory $T$, one always can find a Van Fraassen equivalent higher-order theory in the empirical vocabulary of $T$. For instance, if the theory is finite, it is enough to consider its Ramsey sentences. If so, the choice between Van Fraassen equivalence and weak empirical equivalence (i.e. elementary Van Fraassen equivalence) boils down to whether or not one takes the empirical content of a theory to include the higher order empirical consequences of the theory. Lutz [16] has shown that, in the context of higher-order logic, Van Fraassen’s notion of empirical equivalence can be defined syntactically. Thus, if one thinks that the higher order empirical consequences are important, then one might opt for Van Fraassen...
equivalence. If, on the other hand, one thinks that sentences of higher-order logic do not even express empirical propositions (since they quantify over non-observable entities such as relations), one might opt for weak empirical equivalence.

7 Conclusion

We take the underdetermination thesis to say that, for any theory $T$, there are theories $T_1$ and $T_2$ empirically equivalent to $T$ such that $T_1$ and $T_2$ do not share any theoretical vocabulary and are jointly inconsistent. Is it true? On the classical notion of empirical equivalence (Definition 3.2), where two theories are empirically equivalent just in case they entail the same things with respect to their shared empirical vocabulary, the answer is no (Corollary 3.1). However, on a weaker understanding of empirical equivalence (Definition 4.4), where two theories are empirically equivalent just in case they entail the same things about observable objects, the answer is yes (Theorem 4.1). Using this notion of weak empirical equivalence, it can be argued further that theoretical terms are dispensable for science (Corollary 6.2). However, if one adopts Van Fraassen’s slightly stronger notion of empirical equivalence (Definition 5.2), where two theories are empirically equivalent just in case they are satisfied by the same types of observable structures, dispensability cannot be established on the same grounds (Theorem 6.3). The question is whether this notion still supports underdetermination (Conjecture 5.1).

Acknowledgments I want to thank Hans Mathlein, Dag Westerståhl and Peter Pagin for helpful comments and suggestions on earlier versions of this paper.

Funding Information Open access funding provided by Stockholm University.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Barrett, T.W., & Halvorson, H. (2016). Glymour and Quine on theoretical equivalence. Journal of Philosophical Logic, 45(5), 467–483.
2. Boolos, G., Burgess, J.P., Jeffrey, R.C. (2007). Computability and logic, 5th edn. Cambridge: Cambridge University Press.
3. Boyd, R. (1989). What realism implies and what it does not? Dialectica, 43(1/2), 5–29.
4. Carnap, R. (1975). Observation language and theoretical language. In Hintikka, J. (Ed.) Rudolf Carnap, logical empiricist: materials and perspectives (pp. 75–85): Springer.
5. Dawid, R., & Hartmann, S. (2018). The no miracles argument without the base rate fallacy. Synthese, 195(9), 4063–4079.
6. Demopoulos, W. (2008). Some remarks on the bearing of model theory on the theory of theories. Synthese, 164(3), 359–383.
7. Earman, J. (1993). Underdetermination, realism, and reason. Midwest Studies in Philosophy, 18(1), 19–38.
8. English, J. (1973). Underdetermination: Craig and Ramsey. *The Journal of Philosophy, 70*(14), 453–462.

9. Feyerabend, P.K. (1967). Explanation, reduction and empiricism. *Critica, 1*(2), 103–106.

10. Hempel, C.G. (1958). The theoretician’s dilemma: a study in the logic of theory construction. *Minnesota Studies in the Philosophy of Science, 2*, 173–226.

11. Keisler, H.J. (2010). The ultraproduct construction. In Bergelson, V. (Ed.) *Ultrafilters across mathematics: international congress, ultramath 2008, applications of ultrafilters and ultraproducts in mathematics, June 1-7, 2008, Pisa, Italy* (Vol. 530 pp. 163–179): American Mathematical Soc.

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

13. Ladyman, J. (2001). *Understanding philosophy of science*. Routledge.

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

15. List, C. (1999). Craig’s theorem and the empirical underdetermination thesis reassessed. *Disputatio, 7*, 28–39.

16. Lutz, S. (2014). Empirical adequacy in the received view. *Philosophy of Science, 81*(5), 1171–1183.

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

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

19. Putnam, H. (1979). *Philosophical papers. Mathematics, matter and method*, 2nd edn. Vol. 1. Cambridge: Cambridge Univ. Press.

20. Quine, W.V. (1960). Carnap and logical truth. *Synthese, 12*(4), 350–374.

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

22. Severo, R.P. (2012). The intelligibility objection against underdetermination. *Principia: An International Journal of Epistemology, 16*(1), 121–146.

23. Shelah, S. (1971). Every two elementarily equivalent models have isomorphic ultrapowers. *Israel Journal of Mathematics, 10*(2), 224–233.

24. Sprenger, J. (2016). The probabilistic no miracles argument. *European Journal for Philosophy of Science, 6*(2), 173–189.

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

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

27. 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.