Synthese (2011) 180:157–172
DOI 10.1007/s11229-009-9599-4

Underdetermination, realism and empirical equivalence

John Worrall

Received: 15 October 2008 / Accepted: 1 April 2009 / Published online: 9 July 2009

Abstract Are theories 'underdetermined by the evidence' in any way that should worry the scientific realist? I argue that no convincing reason has been given for thinking so. A crucial distinction is drawn between data equivalence and empirical equivalence. Duhem showed that it is always possible to produce a data equivalent rival to any accepted scientific theory. But there is no reason to regard such a rival as equally well empirically supported and hence no threat to realism. Two theories are empirically equivalent if they share all consequences expressed in purely observational vocabulary. This is a much stronger requirement than has hitherto been recognised—two such 'rival' theories must in fact agree on many claims that are clearly theoretical in nature. Given this, it is unclear how much of an impact on realism a demonstration that there is always an empirically equivalent 'rival' to any accepted theory would have—even if such a demonstration could be produced. Certainly in the case of the version of realism that I defend—structural realism—such a demonstration would have precisely no impact: two empirically equivalent theories are, according to structural realism, cognitively indistinguishable.

Keywords Underdetermination · Realism · Empirical equivalence · Data equivalence

1 Introduction: the 'threat' to realism from underdetermination

It seems to be widely supposed that if scientific theories could be shown to be 'systematically underdetermined' by the evidence, then scientific realism would be in dire trouble. Why?

J. Worrall (✉)
Department of Philosophy, Logic and Scientific Method, London School of Economics and Political Science (LSE), Houghton Street, London, WC2A 2AE, UK
e-mail: j.worrall@lse.ac.uk

The appeal of scientific realism is chiefly based on the—staggering—empirical success of the theories currently accepted in science. The realist exhibits some currently accepted scientific theory (the General Theory of Relativity, say), points to its astounding empirical success (with the gravitational redshift, the precession of Mercury's perihelion, etc.) and suggests that it would be monumentally implausible to suppose that the theory could score such empirical successes and yet not reflect, *at least to some good approximation*, the underlying nature of reality. To hold that combination of beliefs would be, in Poincaré's celebrated phrase (1905/1952, p. 150), "to attribute an inadmissible role to chance".

In order, then, to produce a potential threat to scientific realism, theories would have to be shown to be 'underdetermined by the evidence' in a seemingly quite specific sense. It would have to be shown that no matter how empirically successful a given accepted theory T may have been, rivals T′ to T can always be constructed that are *equally empirically successful*, but that make claims quite different from those of T about the 'deep structure' of the universe. If (but only if) theories could be shown to be underdetermined in this sense, then the realist would indeed seem to be in some trouble. This is because, in that case, the realist presumably ought to regard any such T′ as 'equally good' as T in the light of the evidence, and therefore to stand equally ready to run her argument in favour of that rival. But this time the argument would conclude that it is monumentally implausible that the claims of T' about 'deep structure' are off-beam, given *its* empirical success. But T and T′ are, by supposition, rivals and cannot therefore both be true. Underdetermination in this sense (if established) threatens to reduce 'the master argument' for scientific realism to absurdity.

This understanding of 'underdetermination', although more specific than some, is not in fact as specific as might initially appear: it is by no means clear exactly what it takes for a rival T′ to some accepted theory T to 'share T's empirical success' and therefore to count as 'equally good' for the realist. Indeed clarification of this notion will form a central part of the current paper. (It is the topic of Sects. 3 and 4.) No sensible realist ought to accept a demonstration that two theories deductively entail the same data as showing that those two theories 'share the same empirical success'; and therefore that both are equally good candidates for her No Miracles 'Argument'. The chief reason for this is of course that mere accommodation of some piece of evidence *e* within a theoretical framework does not count as a genuine empirical success in the way that a real prediction of that piece of evidence does. (See again Sect. 3.)

There is however a prior question. Independently of what 'sharing the same empirical success' *really* means (or should mean), suppose that there is a scientific realist who as a matter of fact accepts that there are some particular pairs of contradictory theories T and T′ that do indeed share the same empirical success and therefore are 'equally good' candidates for feeding into the No Miracles 'Argument'. Is it as obvious as the above argument might make it seem that such a realist would be in trouble?

2 What would it take to hurt realism?[1]

Would a realist inevitably be disconcerted, as the argument sketched in the previous section suggests she would, by a demonstration that the choice between two or more fundamental theories is genuinely underdetermined by the data—a demonstration that each is, in whatever way she may understand the phrase, 'equally good' in the light of all the data?

Assuming that the two theories at issue are correctly considered as genuine rivals (and this would itself clearly need investigation in particular circumstances), then a very naïve realist would indeed be in trouble. Let's call such a philosopher a 'gung ho realist'. The gung ho realist holds that the rational position is always to believe in the *truth* of our accepted, empirically successful theories. And two rival, and therefore mutually contradictory, theories cannot, of course, both be true. However, quite independently of any consideration of 'underdetermination', no one should be a gung ho realist about any (let alone every) theory 'accepted' by science.

'Accepted' of course means 'accepted as the currently best available', and a theory may certainly be the best available, and impressively predictively successful, while having problematic aspects. Kepler, Galileo and Newton, for example, all accepted (that is they were all realist about) the basic Copernican theory on the grounds of its predictive success (for example with planetary stations and retrogressions) but were *not* realist about, for example, Copernicus's 'third motion' (a conjectured conical motion of the earth's axis). This was because that 'third motion' was introduced by Copernicus entirely *ad hoc* to solve a problem of the theory's own creation. (Because Copernicus still believed that the planets, and therefore in particular the earth, were fixed in crystalline spheres whose motions carried them round the sun, his theory had a problem explaining the constant angle of inclination of the earth's axis relative to the sun.)[2]

Or take current science: no one should (as is widely recognised) be a gung ho realist about the two most powerful theories in contemporary physics—the General Theory of Relativity and the Quantum theory. GTR and QM are arguably not outright inconsistent but they are mutually incoherent—scientists often say 'incompatible': very roughly, QM says everything is quantized, spacetime, according to GTR, isn't; while GTR says all laws are covariant, but QM is not a covariant theory. (And of course the 'measurement problem' supplies a quite independent reason for being sceptical about a gung ho realist interpretation of QM.) QM and GTR do not perhaps present even a presumptive case of 'underdetermination' in any regular sense since they are not directly two rival theories based on the same range of data. Nonetheless they are two ill-fitting theories when we would like to have one unified theory.[3]

[1] This section was motivated by a brief discussion at the BSPS 2008 Annual Conference in St Andrews with Branden Fitelson—my thanks to him for making me face up to the problem dealt with in this section more directly than I had done before.

[2] See Kuhn (1957/1985, Chap. 5 and Technical Appendix); and Lakatos and Zahar (1976).

[3] One of the currently favoured approaches toward developing such a unified account is M theory.

But if no one seriously believes that QM and GTR are both strictly true, everyone accepts that they are amongst the most impressively empirically successful theories ever. It seems reasonable to believe therefore (or so the sensible realist will insist) that there is *something* about the overall theories (and not just their directly checkable empirical parts) that reflects the 'deep structure' of the universe; but this doesn't mean they are outright true, only that they will both live on as approximations to some future 'synthesis'. (This is exactly why physicists often talk of the search for a *synthesis* of the two rather than outright replacements for them.) Similarly Newton, as just remarked, was realist about Copernican theory—believing it had latched on to the underlying truth in some basic way, while at the same time seeking actively to modify it in important respects (that is, while clearly not believing it to be outright true).[4]

The sophisticated realist therefore claims *not* that belief in the truth of our best theories is rational but 'only' that belief in their *approximate* truth is. The fact that it is only this weaker claim that can seriously be endorsed is of course further underlined by the history of theory-change in science that forms the basis of the so-called pessimistic meta-induction. Pessimistic meta-inducers claim to supply a whole list of previously accepted theories, in the most threatening version of the argument *predictively* successful theories, that were eventually replaced by theories inconsistent with them.[5] No one can seriously argue that our currently accepted theories are definitely immune to similar replacement in the future by theories that are inconsistent with them.[6] The only plausible view, then, is that currently accepted theories are likely to prove 'merely' approximately true in the same sense as those earlier and now rejected theories count as approximately true from the vantage point of the current theories.

Moreover, as the case of Newton's attitude toward Copernican theory indicates, one can be realist about a theory without even expecting that *all of it* will be preserved as a limiting case in future theories. There is nothing in Newton's modification of Copernicus corresponding to the latter's 'third motion' of the earth. But this doesn't mean that it was unreasonable to have a realist view of the theory *overall*.[7]

The important point, then, is that while two mutually inconsistent theories cannot of course both be true, they *may* both be approximately true—that is, both may emerge as (of course different) limiting cases of some further, superior theory, just as current physicists expect GTR and QM both to emerge as limiting cases from the eventual

4 Accepting a theory is, as has often been pointed out, partly a question of deciding to dedicate one's efforts to working on it; but clearly this pragmatic element is not self-standing.

5 Laudan (1981).

6 See my (2000) criticism of Lipton (2000).

7 This is what is correct about the 'selective (or 'partial') realism' of Kitcher (1993) and Psillos (1999, 2004) but, aside from the fact that they provide no satisfactory demarcation between 'working' and merely 'idle posits' within a theory, they do not take on board the fact that the parts of theories that are preserved in 'revolutions' are preserved only structurally. Their selective realism is an *addendum* to structural realism not a rival to it.

'synthesis'. Given this fact, the apparent threat to realism from underdetermination becomes harder to specify.

Once sophistication is allowed, then, first, it becomes clear that the realist *need not* be troubled by particular instances of 'underdetermination' since she may, despite their rivalry, have a (reasonable) realist attitude to *each* of the theories left underdetermined by the data. Moreover, the sophisticated realist certainly *is not* troubled at all by some of the cases that have sometimes been cited against her. Consider the example made much of by van Fraassen amongst others: the case of Newton's 'hypothesis' that the centre of mass of the universe is at rest in absolute space.[8] Here there is a *specific* readily identifiable 'underdetermination'—a parameter λ can provably be adjusted at will without any loss of either empirical power *or theoretical unity*. This means that there is an infinity of different equally empirically powerful theories $T(\lambda)$ for a range of values of λ. Newton himself, while recommending the 'hypothesis' that $\lambda = 0$, demonstrated that all the appearances would be the same (and, importantly, the unity of the overall theory—of mechanics plus universal gravitation—would be retained) if that centre of mass had any uniform velocity relative to absolute space other than zero. In such a case the sensible, sophisticated realist surely says that (provided that, as here, *any* $T(\lambda)$ is predictively successful), there is something about that range of theories that accurately reflects the 'deep structure' of the universe, *but not with respect to the parameter* λ—about whose precise value there is no scientifically justified view. Newton's demonstration that any uniform velocity of the centre of mass would do just as well as the assumption of absolute rest leaves the sensible realist, being realist about the overall theory, but sceptical about any precise value of that velocity, and perhaps about the whole notion of an absolute velocity. The acknowledged underdetermination here does *not* challenge scientific realism. Newton's theory scored stunning successes, both early (with, for example, the precession of the equinoxes) and late (with, for example, the prediction and discovery of the existence of Neptune). It is therefore reasonable to think, insists the realist, that there is *something* 'right' about the overall structure of the theory; but Newton's own demonstration shows that this 'something' need not include the assumption he endorsed concerning the absolute velocity of the centre of mass of the whole system.

The conclusion of this section, then, is that for an underdetermination result to be truly threatening to the scientific realist, it would have to be much stronger than is often recognised. Not only must it be shown that (i) for any accepted scientific theory there is always another that is 'equally empirically successful'; it must *also* be shown (ii) that the alternative cannot plausibly be regarded as equally 'approximately true' as the accepted theory.

As we will see in the next section, it is easy to show (on lines laid down already by Duhem) that on a *very weak* construal of what it takes for two theories to be 'equally empirically successful', condition (i) can readily be established. On that weak construal, alternatives to accepted theories can readily be constructed for which the realist could not plausibly deny condition (ii)—that is, where the realist could not plausibly

8 van Fraassen (1980, Chap. 3).

claim that the 'equally successful' rival was in fact equally approximately true as the accepted theory. But again the sophisticated realist is in no real trouble: the notion of 'equally empirically successful' used to 'demonstrate' (i) is obviously inadequate.

3 Empirical success and 'data equivalence'

It might naively be thought that a rival T′ 'shares T's empirical success' if (and only if) for every empirical prediction e made by T, T′ also entails e. More precisely, I mean by the condition on the right hand side here *not* that every consequence expressible in empirical terms of one theory is also a consequence of the other, but rather only that every *directly checkable* observation result (about apparent planetary positions, the outcome of some experiment such as the two slit experiment in optics, etc.) entailed by T is also entailed by T′. (These are—very—significantly different notions as we will see in the next section.) Let us call this condition, as I intend it, *data equivalence*.

So the proposition that we are now considering is that two theories 'share the same empirical success' or are 'equally well supported by the evidence' (and hence the same realist case can be made for both of them) exactly if the two theories are data equivalent. It is however one of the major lessons of the past 40-odd years of philosophy of science (and indeed one that ought already to have been clear from Duhem's *Aim and structure* (1906/1954), if not still earlier) that this proposition is untenable.

First, we need to be clear about exactly which theoretical units are being considered. As Duhem pointed out (*op. cit.*, Part II, Chap. 6), assertions of the sort that we tend to think of as 'single' theories—Newton's theory (of mechanics plus universal gravitation), the wave theory of light, etc.—entail no empirically checkable results at all when considered 'in isolation': that is, without 'auxiliary' assumptions. Hence if we are considering such 'single theories', there is of course no problem in producing for any accepted theory a rival that is data equivalent to it. The negation of the accepted theory will do: so, for example, Newton's theory and its negation are of course rivals and they are trivially data equivalent since neither entails any datum.

Additionally, as Duhem also pointed out (*ibid*), many 'single' theories such as 'the wave theory of light' (say, to be specific, that developed by Fresnel by 1823) themselves break down naturally into a central or *core* theory—in this case the assertion that light consists of periodic vibrations transmitted through some sort of medium—together with *more specific* assumptions: about the mechanical characteristics of the medium, the types of wave corresponding to light of different colours and so on. It is this that allows for coherent talk about 'a' theory 'evolving' over time: in response to empirical and conceptual difficulties wave theorists rejected some specific assumptions and replaced them by others, while of course retaining the core theory that identified them as wave theorists.

Even once some particular set of specific assumptions for such a theory has (temporarily) been fixed, however, further auxiliary assumptions are still required before any datum can validly be deduced. It seems natural then to characterise the resulting 'theory' consisting in general of core, specific and auxiliary assumptions as in fact a *theoretical system*.

So theoretical systems, unlike 'core' or 'single' theories, *do* entail directly checkable observation results. And the upshot of Duhem's analysis was, of course, that rival theoretical systems based on rival 'core' theories can always be made data equivalent. Suppose we have two rival 'core' or 'central' theories C_1 and C_2 (the basic wave theory of light versus the basic corpuscular theory (light consists of material particles of *some sort/s*), for example, or Newton's theory versus the special theory of relativity). For any given set of data E, there must always be *some* sets of auxiliaries A_1 and A_2 which when added to C_1 and C_2, respectively, will produce rival theoretical systems T_1 and T_2 both of which entail E. (So, for example, $C_1 \rightarrow E$ and $C_2 \rightarrow E$ would do for A_1 and A_2 respectively.)

Of course there is, and can be, no guarantee that data equivalence will be preserved once the stock of data is expanded, via the discovery of some new datum e, into the set E'. It might turn out that only one (or perhaps neither) of the systems T_1 and T_2 entails e and therefore E'. Suppose that T_1 does entail e (and hence E') but E_2 does not. However, Duhem's point applies again to show that, by of course now invoking different auxiliaries, a framework T_2' based on the same core (C_2) can be constructed such that T_1 and T_2' are again data equivalent.

Equally clearly, and very significantly for this debate, 'data equivalent' does not entail 'equally empirically supported'. Any number of accounts of the confirmation of theory by evidence, starting with hints in Duhem and including my own detailed account,[9] entail a crucial difference between prediction and accommodation.

'Prediction', as I have argued following Lakatos and Zahar,[10] has, when properly understood, no (necessary) temporal connotations—whether or not the evidence was known before a theory was discovered to entail it is, by itself, irrelevant. Prediction properly understood is simply the opposite of accommodation. A piece of evidence e is accommodated within a theoretical system T based on a core theory C by tailoring specific and/or auxiliary assumptions exactly so as to produce such a system that entails e. A datum e′ is predicted by a theoretical system just in case it is deductively entailed by that system but was not accommodated within it.

A classic case of accommodation is that of the 'fossil' evidence within the framework of 'special creation' by using what I sometimes call the 'Gosse dodge'. This was invented by Philip Gosse in his book *Omphalos*. There seem to be the impressions of the skeletons of previously existing but now extinct species in certain rocks, and fossilised bones of such creatures underground in bits of earth—which, had those creatures really existed, would have (long) predated 4004BC. No problem, said Gosse: God obviously chose, when creating the whole universe in 4004 BC, to make those particular rocks or those particular pieces of earth with that 'engraving' or that bone-*like* 'fossilised' structure already in them.

Another classic case is the accommodation by Ptolemy of the evidence of planetary stations and retrogressions within theoretical systems based on the core claim of a fixed and central earth. The planets as observed from the earth seem to have a combined motion consisting of two components—a westward diurnal rotation

9 Worrall (2006).

10 Lakatos (1970) and Lakatos and Zahar (1976).

with the fixed stars and a generally eastward motion *against* the fixed stars. However that second 'proper' motion of the planet is periodically interrupted by its gradually coming to a halt (station)—so that it now instantaneously has exactly the same diurnal rotation as the stars—and 'retrogressing' for a while against the background of the fixed stars before again halting and then assuming its more normal eastward motion. This phenomenon is straightforwardly accounted for (predicted!—even though the phenomenon had been known for centuries before Copernicus) by the basic Copernican model. The planets have their own regular orbit; but we observe them from our moving observatory on Earth. The stations and retrogressions are the (merely apparent and inevitable) results of the earth either overtaking (the superior) or being overtaken by (the inferior) planets: during the overtaking the planet (when viewed against the background of the fixed stars) will automatically *appear* from our moving observatory on Earth to retrogress. On the other hand, in order to produce a theoretical framework based on the geostatic core that dealt with stations and retrogressions, Ptolemy, as is well known, had to introduce a special device—the epicycle—and the relative velocities of the epicyclic and deferent rotations had to be adjusted precisely in the light of the known observations.[11]

In both the Darwin vs. Creationism and Copernicus vs. Ptolemy cases, the two theories, or rather latest theoretical systems based on them, end up as data equivalent. (And indeed in the second case the two theories were demonstrably data equivalent relative to *all* the data known at the time of the publication of *De Revolutionibus*.) But it is surely clear that any serious account of empirical support will need to entail a difference in the empirical support leant to Darwin and Creationism by the 'fossils', and in the empirical support leant to Copernicus and Ptolemy by the observation of planetary stations and retrogressions.

Some accounts invoke simplicity to underwrite the accommodation/prediction distinction—both Copernicus and Ptolemy entail the correct data concerning stations and retrogressions and that is all one can require empirically, but Copernican theory is the simpler. However even sticking to the intuitive level (it has of course proved notoriously difficult to characterise simplicity formally), it seems clear that this is to underrate the role of the phenomena, which drop out of Copernicus in a completely natural way, but which *force* the complexity in Ptolemy. The accurate judgement—delivered by the account of confirmation I endorse (*op. cit.*)—is that, despite the fact that fossils are accounted for both by Darwin and by Creationism and planetary stations and retrogressions follow from theoretical systems built around the two rival core claims of helio- and geo-centrism, the phenomena in both cases give more empirical support to (and hence supply an empirical reason to prefer) the first theories in these pairs.

Hence, returning to the vague notion of 'sharing empirical success' that I started from, it is not true on this account of confirmation (or indeed on any that seems half-way adequate) that the fact that two theories (or rather theoretical systems) are data equivalent entails that those two systems (and more pertinently) their respective core theories 'share the same empirical success'.

[11] See Kuhn *op. cit.*, Chap. 5.

The Duhemian way into 'underdetermination' seems to be the only one that really arises historically in cases of theory-change. Or at least it is the only such way that can be invoked generally across all theories. Kuhn's account and in particular his claims about 'elderly hold-outs' for what turns out to be an older paradigm in a revolution are in essence just (rather confusing) paraphrases of Duhem's analysis.[12] And, as we just saw, this Duhemian way does not underwrite any notion of underdetermination that should trouble the realist. Such a realist certainly does not want to adopt the sort of realist attitude suggested in Sect. 2 toward, say, *both* Darwinism *and* Creationism, *both* Copernicus *and* Ptolemy; but there is no argument to suggest that she is obliged to do so. Independently of any consideration about realism, the fact that data equivalent theoretical systems can be produced based on either the first or second of either of these two pairs of core theory does not commit the realist to holding that both theories in either pair are equally empirically successful (even with respect to the range of phenomena to which they have so far been shown to data equivalent). And hence there is no suggestion that the No Miracles 'Argument', if it applies at all, should apply to both theories equally.

And indeed the point is strongly underwritten exactly by concentrating on that argument. There is at least some intuitive bite to the idea that it is, for example, implausible that Copernican theory could get the phenomena of planetary stations and retrogressions correct as directly as it does unless it has latched on, at least approximately, to the 'way things really are'. But we *know* the explanation for Ptolemy's 'success' with those same phenomena; and it has nothing to do with the world, but rather with the ingenuity of Ptolemaic astronomers in solving the problem of engineering post hoc a geostatic accommodation of the already known phenomena—a problem for which, as Duhem's analysis assures us in advance, there must be any number of solutions.

4 Data equivalence and empirical equivalence

I argued in Sect. 2 that in order to trouble the scientific realist an underdetermination argument would have to establish *not only* (i) for any accepted scientific theory there is always another that is 'equally empirically successful' *but also* (ii) that there is some reason why the realist could not reasonably regard the alternative as 'approximately true' just like the accepted theory. In Sect. 3 I argued that the standard Duhem way into 'underdetermination' (the only one that seems really to arise in the development of science as a general issue) fails even to establish condition (i). That argument was based on the claim (fact!) that it is mistake to equate 'equal empirical success' with 'data equivalence'. Is there some other explication of the notion of 'equal empirical success' that might form the basis for a genuine challenge to scientific realism from 'underdetermination'?

In a much-discussed (1991) paper, Laudan and Leplin have already presented an argument to the effect that underdetermination is less troubling for the realist than might meet the eye. And that argument bears at least some superficial similarities to the one being developed here. Laudan and Leplin's argument proceeds as follows.

[12] Worrall (2003).

First they identify what they explicitly take to be the 'traditional' notion of *empirical equivalence*. *Secondly*, they argue (a) that there is in fact no general guarantee that for any given theory we can always construct empirically equivalent rivals and (b) that even if there are some cases where empirically equivalent rivals can be produced, it would be a mistake to infer automatically that those rivals are equally empirically successful or equally well supported by the evidence.

There are problems with Laudan and Leplin's argument for (a)—some pointed out in the subsequent literature (e.g. Okasha 1997); and, as for (b), while this may seem superficially to be related to the argument in the previous section, their own version of it is in fact very different (and very problematic) as we shall now see.

Laudan and Leplin's characterisation of empirical equivalence (which, as just remarked, they take—perhaps with some justification—to be 'traditional' in the literature) is as follows. First, divide the vocabulary of the common language within which any two theories T and T′ are expressed into the purely empirical (or observational) vocabulary and the theoretical vocabulary. T and T′ are, then, *empirically equivalent* just in case the sets of their deductive consequences that are expressible purely in the observational vocabulary are identical.[13] (I take it here that purely logical and mathematical vocabulary is shared: we want to say, for example, that 'there are two planets in that portion of the sky' is in the observation language, while 'there were two electrons in that section of the bubble chamber' is in the theoretical language.)

It might be thought (and it seems hitherto to have been assumed in the literature) that this notion of empirical equivalence is itself equivalent to the notion of data equivalence introduced earlier. However this is far from being the case.

We saw in Sect. 3 that Ptolemaic theory and Copernican theory are data equivalent relative to the (apparent) motions of the sun, fixed stars and planets known at the time of *De Revolutionibus;* and that Darwinian theory and 'Gossefied Creationism' are data equivalent with respect to the fossil record. But, contrary perhaps to immediate impressions, this by no means entails that Ptolemy and Copernicus or Darwin and Creationism are (or, rather, can be made to be—that is, can be embedded within suitable theoretical systems that are) empirically equivalent in Laudan and Leplin's sense. Nor is this to do with the possibility (of course actualised in both these cases) of extensions of the data sets.

This is especially clear in the case of the second pair of theories. Darwinian theory (D) and Gossefied creationism (G) (relative, remember, just to the 'fossils') yield all the same data, but there are any number of statements that are in purely observational vocabulary over which they differ. For example, G entails that no observable element

[13] Of course, Laudan and Leplin, like everyone else, are aware that, at least if ordinary usage is our guide, the distinction between theory and observation—and hence the division into theoretical and observational vocabulary—is extraordinarily vague. (Indeed they explicitly attempt to exploit this vagueness in arguing for one of their central theses). But obviously some such distinction must be presupposed in order even to raise the underdetermination issue: if there is no distinction between statements about data and theoretical claims, then the question of whether or not theories are 'underdetermined by data' cannot even be raised. For the purposes of the present paper we can operate, as Laudan and Leplin implicitly do, with some intuitive distinction that yields gluons, quarks, electrons, spacetime curvatures, and light waves, for example, as theoretical, and planets, people, tracks on cloud chamber photographs, and interference fringes, for example, as observable.

of the universe has existed for more than approximately 6000 years—this is expressed purely in observational language (I assume) and yet is at odds with D. Notice then that this is definitely not a question of two theories that are 'equivalent' with respect to one set of data, becoming non-equivalent when that data set is extended through new types of result. With respect to this dispute, the claim that nothing is older than 6000 years old can never be a datum—it is an *observational claim that is subject to theoretical dispute*.

One reaction to this, exploiting the vagueness of the ordinary usage of 'observational', would be to deny that statements about an object's age can count as observational: only statements about an object's '*apparent* age' can count. And of course the two theories D and G *agree* that there are lots of denizens of the universe whose *apparent* ages are more than 6000 years. But this simply complicates the situation without affecting the point: the assertion that there was a time (roughly 4004BC) before which none of the current constituents of the 'material' universe had an apparent age (because nothing, except presumably God, existed) is (i) unambiguously in the observation language even on this more demanding construal; (ii) entailed by G; and (iii) contradicted by D (which of course identifies apparent and real ages).

Even aside from particular examples, the fact that there are many claims that are expressed purely in observational vocabulary but that ought to count as theoretical in anyone's book should hardly come as a surprise. A much used example[14] is the assertion that there are unobservable objects—i.e. objects with no (directly) observable property. Carl Hempel in his famous paper 'The Theoretician's Dilemma' (1958/1965) provided the following further example (p. 197):

> Let Sx, y, z hold iff x is farther away from y than from z, then $\mathrm{Pa} \equiv \exists x \forall y[\neg(x = y) \rightarrow Sa, x, y]$ states that there is an object such that a is further away from that object than it is from any other object.

Pa is clearly an expression of the observation language on any reasonable construal and yet, as Hempel points out, it surely counts as theoretical: since no finite set of observation statements can either verify or refute it.

This second example in particular underwrites the important point (much emphasised also by Popper) that whether or not a sentence counts as a theory is not just a question of the vocabulary in which it is expressed but also of its quantificational structure.[15]

These general facts indicate that the failure of full empirical equivalence for pairs of data equivalent theories is not an accidental feature of the particular examples I have cited (nor of others that are often cited such as Reichenbach's flat space plus universal forces versus the General Theory of Relativity). Any theory has consequences that are (i) expressible in the observation language and yet (ii) cannot be decided on the basis of observation or experiment but (iii) rival theories deny. This seems clearly to categorize such consequences as *observationally expressed theoretical statements*. Genuinely rival theories, even if they can be made data equivalent, will (by (iii)) continue to conflict over a range of such statements and hence will automatically fail to be

---

14 I think I learnt it as an undergraduate from a lecture by Imre Lakatos.

15 See Worrall and Zahar (2001).

empirically equivalent in Laudan and Leplin's sense. Take the classical wave theory of light, for example. It entails that there is some medium with no directly observable properties which plays a role (mathematically specifiable) in optical effects. This claim is purely in the observation language, and the corpuscular theory contradicts it.[16]

Hence two theories that are really empirically equivalent in Laudan and Leplin's 'traditional' sense will in particular have to be equivalent with respect to a range of *theoretical* assertions—those theoretical assertions expressible in purely observational vocabulary. Hence any two 'rival' theories that are empirically equivalent in this (as it now transpires very strong) sense will at least have to agree, not only about the data, but also over a wide range of claims that everyone should take to be theoretical. Once again the threat that is posed to realism by underdetermination becomes altogether less clear cut than it at first appears.

Certainly for the form of realism that I advocate—namely structural realism[17]*—any demonstration that for any accepted theory there is another that is empirically equivalent to it would pose no problem at all.* This is because structural realism entails that any two such theories are, by virtue of their empirical equivalence, fully cognitively equivalent.

This perhaps initially surprising result is in fact easily proved:

1. The claim that the full 'cognitive content' of a theory T is captured by its Ramsey sentence R(T) is a defining characteristic of structural realism—at least as Poincaré, Zahar and myself have understood it.[18]
2. R(T), by construction, is expressed purely in the observation language (all the theoretical predicates having been 'quantified away').
3. Moreover R(T) is of course a (second-order) deductive consequence of T.
4. Hence any two theories that have the same set of empirical consequences (remember *all* consequences expressible in purely observational vocabulary) automatically entail equivalent Ramsey sentences and are therefore, according to the claim in 1, cognitively equivalent.

Premises 2 and 3 of this argument are trivial (3 is directly underwritten by the second-order version of the rule of existential generalisation, but if you prefer first order logic, just assume first order set theory and identify properties with sets); and the inference from premises 1, 2 and 3 to the conclusion at 4 is valid. The only part of the argument that can be questioned, therefore, is premise 1.

The detailed defence of this premise, as being both characteristic of structural realism and the only sustainable view, is developed in a separate paper (Worrall forthcoming) that investigates (and rebuts) the so-called 'Newman Objection' to structural realism. But let me briefly outline the argument here, lest the premise's claim appear absurd.

---

[16] At least the 'pure' corpuscular theory denies it. It says that light consists simply of particles subject to a variety of forces. There were versions of the theory—such as the one that Newton himself seems clearly to have believed while explicitly denying that he did—that stated that, while the light emitted by sources such as the sun consisted only of particles, those particles then moved through a medium and created waves in it which played a role in optical phenomena such as that of Newton's rings. (For references see Worrall 2001.)

[17] Worrall (1989 and forthcoming).

[18] Russell too, on my understanding of him, but for a dissenting opinion see Votsis (2005).

At least as far as theoretical talk in science is concerned we are, I suggest, stuck with ‘global descriptivism’ and obviously so: all of our knowledge of electrons, protons, gluons and the rest of the rich stock of theoretical notions in current science is through description. To suggest anything else would be to indulge in clearly fantastical talk about being able to ‘stand outside’ the whole of our knowledge, and have some non-theory-mediated access to the world—one that allows us to compare the things that we thus extra-linguistically apprehend with our linguistically-formulated theories about them. (An *apparently* different alternative would be to invoke one version or other of the ‘causal theory of reference’ but, so I argue in [forthcoming], *either* this accepts that our knowledge of causes too is theoretical (in which case the causal theory of reference is ‘just more theory’ and so disguised descriptivism) *or* it is a disguised version of the above fantasy, essentially relying on some mystical ‘semantic glue’ between theoretical terms and theoretical entities and on our somehow being able to ‘apprehend’ that glue.) But if all our knowledge of theoretical entities is descriptive, then it follows that if you are asked what, say, the term ‘gluon’ refers to all you can do is reiterate our current best (total!) theories of gluons: that is, a gluon is a ‘whatever it is’ that structures the phenomena in certain complex ways through specific intricate relationships with the phenomena and with other, similarly characterised, theoretical notions. This characterisation, however, is just an informal statement of the Ramsey sentence for our theory of gluons, in which the theoretical predicates have been replaced by second-order quantifiers. (The primitive theoretical predicates in the initial un-Ramseyfied theory name (or attempt to name) theoretical entities in the same way that the ambiguous names involved in some systems of predicate logic do—that is, not directly in the way that we think of regular individual constants naming individuals but through the sentences we assert using them. And of course in such systems of first order logic, where α is any ambiguous name, $P\alpha$ and $\exists xPx$ are inter-derivable and so ‘cognitively equivalent’.)

In other words, once you have accepted global descriptivism concerning all our theoretical notions then, as Russell and Poincaré both clearly saw, you have automatically accepted the ‘Ramsey view’ that the full cognitive content of a theory is captured by its Ramsey sentence. To claim that we have epistemic access to something beyond R(T) would, in Russellian terms, involve the claim that we have some sort of *acquaintance* with the theoretical notions designated by the theoretical terms—and this is just a version of the ‘out of theory’ fantasy identified above.

There is one response to this argument that is given so often that I ought to at least indicate here how to deal with it (though again details will be found in Worrall [forthcoming]). This response is that it is no news that once you have adopted the ‘Ramsey view’ there is no problem of underdetermination; because adopting that view is in effect to reject realism in favour of an empiricist, instrumentalist view of theories; and nobody, of course, ever believed that there is a problem of underdetermination if you are an instrumentalist. If structural realism is committed to the ‘Ramsey view’ then it is not really realism, and so it is no wonder that it does not face the problem of underdetermination—that problem arises only for *real* realism.[19]

[19] See for example Psillos (1999, Chap. 7).

Well, structural realism is certainly committed *in a sense* to the claim that there is no difference without an observable difference—since it holds that the Ramsey sentence of any theory carries its full 'cognitive' content and that Ramsey sentence is itself purely, by construction, in the observation language. And this might suggest to the unwary that it does indeed collapse into some form of instrumentalism or positivism. But to follow that suggestion would again be to fail to recognise the data equivalence/empirical equivalence distinction articulated *above*. Structural realism is *not* committed, via its acceptance of the 'Ramsey view', to regarding, for example, Copernican and Ptolemaic theory as 'cognitively equivalent' at the time of Copernicus (even on the counterfactual supposition that the only evidence ever available will be that available to Copernicus). The notion of 'no observable difference', when understood as meaning no difference over any sentence expressible in the observation language, is an extremely powerful one, as I have tried to demonstrate. Many sentences expressed purely in observational vocabulary should count as theoretical in anyone's book. Hence if two 'different' theories are observationally equivalent then they will, in general, agree on much that is clearly theoretical.

Along the same lines, it might seem tempting to infer that by quantifying over theoretical terms, the Ramsey sentence must eliminate the 'real' theoretical content of its parent theory. But surely to stand ready to assert a sentence that quantifies over theoretical terms involves asserting (not denying) their existence. (This is just a second-order mirroring of Quine 1961 on ontological commitment.) And I have argued that, whatever your position on the realism/anti-realism issue, you just have to accept that some sentences expressed in purely observational vocabulary are theoretical—the Ramsey sentence of any complex scientific theory is a prime example. Carl Hempel, indeed, already made it clear in the 'Theoretician's Dilemma' that the Ramsey sentence does not 'do away with' theoretical notions. Hempel, recall, was attempting to find a way of eliminating theoretical terms—as a means of resolving the theoretician's dilemma. He notes that some philosophers have thought that Ramsey-fying provides exactly such a way, but emphatically rejects their view (*op. cit.* p. 216):

> …the Ramsey-sentence associated with an interpreted theory T avoids reference to hypothetical entities only in letter—replacing Latin constants by Greek variables—rather than in spirit. For it still asserts the existence of certain entities of the kind postulated by T, without guaranteeing any more than does T that those entities are observable or at least fully characterizable in terms of observables. Hence, Ramsey-sentences provide no satisfactory way of avoiding theoretical entities.

Bad news for Hempel, since it means that Ramsey sentences fail to resolve his 'dilemma', but good news for those of us who accept that Ramsey sentences capture the full cognitive content of scientific theories but still insist on being counted as realists about such (successful) theories. To accept, as structural realism does, that a theory in effect 'reduces' to (that is, carries no further epistemically accessible content than)[20] its Ramsey sentence is *not* to 'eliminate' theory. And hence to endorse

[20] Of course the Ramsey sentence is logically distinct from its 'full' theory—the theory is strictly logically stronger than its Ramsey sentence, but the (surely correct) claim of the (epistemic) structural realist is

the claims of the Ramsey sentences of our currently accepted theories to reflect the theoretically-described 'deep structure' of the universe is to advocate a version of scientific realism. Indeed it is to advocate what is, in my view, the only tenable version of scientific realism. Only those who assert that any two *data* equivalent theories are cognitively equivalent have abandoned realism for anti-realism.

5 Conclusion

Like Laudan and Leplin, though for notably different reasons as we have seen, I have argued that the alleged threat to orthodox epistemologies and in particular to scientific realism from 'underdetermination' has (to put it conservatively) yet to be substantiated. As pointed out in Sect. 3, Duhem already showed that theoretical systems based on rival core theories can always be developed that are *data equivalent*. But, as Duhem himself suggested—Duhemian 'natural classifications' proclaim themselves by being genuinely and successfully predictive—and has been developed in detail in other more recent accounts of confirmation, it by no means follows that theoretical systems based on rival core theories can always be developed that are equally 'empirically successful' or equally empirically supported. There may be—indeed there standardly are—good empirical reasons for preferring one of two data equivalent theoretical systems to the other.

In Sect. 4, I showed that quite different considerations apply to the notion of *empirical equivalence*—which has not been sufficiently clearly differentiated from data equivalence in the literature. According to at least one version of scientific realism—the one that seems to me most (indeed uniquely) defensible—there is no threat at all from the possibility of theories 'rival' to accepted ones that are empirically equivalent to them. This is because structural realism entails that there is no effective difference between two such 'rivals'. According to structural realism (and now definitely contrary to Laudan and Leplin), there can indeed be no empirical reason to prefer one of two 'rival' theories that are empirically equivalent in the sense discussed; but this is because there is no significant difference between them—they are not genuinely rivals.

There may be other accounts of what it takes for two 'significantly different' theories to 'share the same empirical success' but unless and until one such is developed it seems that scientific realism has nothing to fear from 'underdetermination'.

Acknowledgments I am grateful for comments made by those attending the April 2008 Dusseldorf conference where a version of this paper was first given. I am especially grateful to Jeff Ketland and to Roman Frigg who read an earlier full draft and have saved me from a number of errors (though several 'errors', as they see it, proudly remain). I am also grateful to Dean Peters for research assistance and helpful comments. Finally I am indebted to an anonymous referee for *this journal* for a number of very detailed comments and criticisms.

there is, even in principle, no epistemic difference between. The difference between the two—the so-called Carnap sentence—is an in principle entirely untestable, completely metaphysical assertion. There is no 'cognitive' difference between a theory and its Ramsey sentence, not simply no difference of any *present* epistemic moment, but none of any conceivable epistemic moment. (See Worrall and Zahar 2001.)

References

Duhem, P. (1906/1954). *The aim and structure of physical theory*. Princeton, NJ: Princeton University Press.

Hempel, C. G. (1958/1965). The theoretician's Dilemma: A study in the logic of theory construction. In *Aspects of scientific explanation* (pp. 173–228). New York: The Free Press.

Kitcher, P. (1993). *The advancement of science*. Oxford: Oxford University Press.

Kuhn, T. S. (1957/1985). *The Copernican revolution: Planetary astronomy in the development of western thought*. Cambridge, MA: Harvard University Press.

Laudan, L. (1981). A confutation of convergent realism. *Philosophy of Science, 48*(1), 19–49.

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

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. E. Musgrave (Eds.), *Criticism and the growth of knowledge*. Cambridge, MA: Cambridge University Press.

Lakatos, I., & Zahar, E. (1976). Why did Copernicus' research program Supersede Ptolemy's? In R. S. Westman (Ed.), *The Copernican achievement. UCLA Center for Medieval and Renaissance Studies Contributions* (Vol. VII, p. 354). Berkeley, CA: University of California Press.

Lipton, P. (2000). Tracking track records I. In *Proceedings of the Aristotelian Society*, (Supplementary Vols. 74, pp. 179–205).

Okasha, S. (1997). Laudan and Leplin on empirical equivalence. *British Journal for the Philosophy of Science, 48*, 251–256.

Poincaré, H. (1905/1952). *Science and hypothesis*. New York: Dover Publications, Inc.

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

Psillos, S. (2004). Tracking the real: Through thick and thin. *British Journal for the Philosophy of Science, 55*, 393–409.

Quine, W. V. O. (1961). On what there is. In *From a logical point of view: 9 Logico-philosophical essays* (pp. 1–19). Cambridge, MA: Harvard University Press.

van Fraassen, B. (1980). *The scientific image*. Oxford: Clarendon Press.

Votsis, I. (2005). The upward path to structural realism. *Philosophy of Science, 72*(5), 1361–1372.

Worrall, J. (1989). Structural realism: The best of both worlds. *Dialectica 43*(1–2), 99–124.

Worrall, J. (2000). Tracking track records II. In *Proceedings of the Aristotelian Society*, (Supplementary Vols. 74, pp. 207–235).

Worrall, J. (2001). The scope, limits, and distinctiveness of the method of 'deduction from the phenomena': Some lessons from Newton's 'demonstrations' in optics. *British Journal for the Philosophy of Science, 51*, 45–80.

Worrall, J. (2003). Normal science and Dogmatism, Paradigms and Progress: Kuhn 'versus' Popper and Lakatos. In T. Nickles (Ed.), *Thomas Kuhn* (pp. 65–100). Cambridge, MA: Cambridge University Press.

Worrall, J. (2006). Theory-confirmation and history. In C. Cheyne, J. Worrall, & A. Musgrave (Eds.), *Rationality and reality: Conversations with Alan Musgrave* (pp. 31–62). Dordrecht: Springer.

Worrall, J. (forthcoming). Defending structural realism; or: The 'Newman objection' what objection?

Worrall, J., & Zahar, E. (2001). Appendix IV: Ramseyfication and structural realism. In E. Zahar (Ed.), *Poincaré's philosophy: From conventionalism to phenomenology* (pp. 236–251). Chicago: Open Court.