Chapter 12
Re-inflating the Realism-Instrumentalism Controversy

Stathis Psillos

Abstract Compatibilist instrumentalism, aka irenic instrumentalism, exemplified in the work of Rudolf Carnap and Ernest Nagel, tended to underestimate the ontic differences between realism and instrumentalism making the claim that the two positions are merely different modes of speech (Nagel) of that they differ only qua languages (Carnap). Compatibilism as such did not last for a long time but it's taken two new leases of life. The first is in the very influential paper by Stein in the early 1990s. The second is in Kyle Stanford's recent work. Between them, the arguments by Stein and Stanford aim to deflate the realist-instrumentalism controversy and to make instrumentalism palatable as a view of science. The aim of this paper is to show that compatibilism fails. Instrumentalism and realism are genuine rivals. Hence, the difference between realism and instrumentalism (even when they are taken to be 'sophisticated') is deep and philosophically significant.

Keywords Realism · Instrumentalism · Carnap · Nagel · Stein · Stanford

12.1 Introduction

Ever since instrumentalism, based on Craig's theorem, was almost summarily dismissed by a series of arguments that favoured scientific realism, in the 1960s, denialism – the view that the so-called theoretical entities do not exist and that the only substantive dispute between realism and instrumentalism concerns whether theoretical terms have meaning or not – was split into two positions: agnosticism (exemplified in Bas van Fraassen's constructive empiricism) and compatibilism, called irenic instrumentalism, by Wilfrid Sellars (1965), (exemplified in the work of Rudolf Carnap and Ernest Nagel). The latter approach tended to underestimate the ontic differences between realism and instrumentalism making the claim that the two

S. Psillos (✉)
Department of History and Philosophy of Science, National and Kapodistrian University of Athens, Athens, Greece
e-mail: psillos@phs.uoa.gr

W. J. Gonzalez (ed.), *Current Trends in Philosophy of Science*, Synthese Library 462, https://doi.org/10.1007/978-3-031-01315-7_12

positions are merely different modes of speech (Nagel) of that they differ only qua languages (Carnap).

Compatibilism as such did not last for a long time but it's taken two new leases of life. The first is in the very influential paper by Stein in the early 1990s. In this Stein argued that "between a cogent and enlightened 'realism' and a sophisticated 'instrumentalism' there is no significant difference – no difference that makes a difference" (1989, 61). The second is in Kyle Stanford's recent work. After trying to give a new lease of life to instrumentalism by advancing the unconceived alternatives argument in his (2006), Stanford has been pushing the line of a new compatibilism between 'historicist' instrumentalism and those versions of realism, which, as he put it, have granted to instrumentalism "that we are in the midst of an ongoing historical process in which our theoretical conceptions of nature will continue to change just as profoundly and fundamentally as they have in the past" (2015a, 875). Between them, the arguments by Stein and Stanford aim to deflate the realist-instrumentalism controversy and to make instrumentalism palatable as a view of science. At the same time, however, instrumentalism as a genuine rival to realism is revamped by a younger generation of philosophers, notably Brad Wray (2018) and Darrell Rowbottom (2019). And while Wray ventures for agnosticism (see my 2020), Rowbottom goes for (a new) denialism (see Psillos & Zorzato, 2020).

The aim of this paper is to show that compatibilism fails. Instrumentalism and realism are genuine rivals. Hence, the difference between realism and instrumentalism (even when they are taken to be 'sophisticated') is deep and philosophically significant.

Here is the road map. Taking a cue from Rowbottom's claim that instrumentalism is a *movement*, it will be shown in Sect. 12.2 that the saving-the-phenomena tradition, which Pierre Duhem thought was instrumentalist, was not really part of the movement. Then in Sect. 12.3 I will identify two core commitments of the movement, viz., denialism and non-eliminativism. In Sect. 12.4 I will argue that there is a tension between these two commitments and in Sect. 12.5, it will be shown that Craig's theorem was an aborted attempt to relieve this tension by abandoning non-eliminativism. Sections 12.6, 12.7, 12.8 and 12.9 will examine the main forms and arguments for compatibilism: Carnap's, Nagel's, Stein's and Stanford's.

12.2 To Save the Appearances

Rowbottom (2018), one of the few outspoken current instrumentalists, takes it that instrumentalism treats science as an instrument; he also adds that instrumentalism takes it that science is about the observable (or the *phenomena*); scientific theories help us "orient" ourselves in the world around us. That's a very broad understanding of a "movement". What makes it "a form of anti-realism"? Surely, both of the points above are consistent with realism. The view that theories are useful instruments for prediction and control is actually a consequence of realism: if theories do succeed in uncovering the unseen world, then surely they succeed in what they assert about the

phenomena. In fact, the realist would argue that the latter success is fully due to the former. What then makes the instrumentalist movement *anti-realist*? It's precisely something added on to the two claims made above, and this more is typically captured by a 'nothing but' clause: theories are *nothing but* instruments for prediction and control. (Rowbottom puts "primarily" before "instruments", but the force is the same.) This *nothing-but-ness* is what realism denies; hence the incompatibility.

The *nothing but* is usually qualified by expressions such as 'in the final analysis' or 'the cash value is nothing but. . .'. But the key point should be clear: the *credo* of the instrumentalist movement is that any kinds of aspirations we might have that science goes beyond the phenomena is unwarranted and/or superfluous. This kind of block to science's cognitive aspirations is either because of ontological denialism (e.g., there is nothing beyond the appearances) or because of epistemological prudentialism (e.g., even if there is something beyond appearances, for instance, their unseen causes, either we are unable to find out what they are or it does not matter for doing science well that we try to map the unseen causes).

Instrumentalism is indeed a movement. Both Mach and Duhem flirted (to say the least) with it. But to be part of the movement you have to adopt the fundamental credo. This credo is captured nicely in the anonymous preface of the famous *De Revolutionibus Orbitum Coelestium*, in 1543, the posthumously published masterpiece of Copernicus, where it is written:

> For it is the duty of an astronomer to compose the history of the celestial motions through careful and expert study. Then he must conceive and devise the causes of these motions or hypotheses about them. Since he cannot in any way attain to the true causes, he will adopt whatever suppositions enable the motions to be computed correctly from the principles of geometry for the future as well as for the past. The present author has performed both these duties excellently. For these hypotheses need not be true nor even probable. On the contrary, if they provide a calculus consistent with the observations, that alone is enough. (. . .) For this art, it is quite clear, is completely and absolutely ignorant of the causes of the apparent nonuniform motions. And if any causes are devised by the imagination, as indeed very many are, they are not put forward to convince anyone that are true, but merely to provide a reliable basis for computation ([1543] 1992, xix).

As it turned out, the preface was written not by Copernicus himself but by Andreas Osiander, a Lutheran theologian. The intention was obvious: to avoid the wrath of the church. But leaving this aside for a minute, the key *argument* for instrumentalism is simply that the astronomer cannot find the truth about the causes of motion. Now, in this particular context, the truth is "divinely revealed"; hence neither the astronomer nor the philosopher will be able to attain it by evidence plus sound reasoning. But why should we take this argument seriously in a post-revelation context?

Pierre Duhem in his famous essay *ΣΩZEIN TA ΦAINOMENA* [*To Save the Phenomena*] (1908) secularised the argument by embedding into a long homonymous tradition which goes back to Plato. Before Copernicus, the dominant astronomical theory was Claudius Ptolemy's (ca. 85–ca.165). He had assumed, pretty much like Aristotle and Plato, a geocentric model of the universe. To save the appearances of planetary motions, Ptolemy had devised a system of deferents and

epicycles. There were alternative mathematical models of the motion of the planets (e.g., one based on a moving eccentric circle), but Ptolemy thought that since all these models were saving the appearances, they were good enough. The issue of their physical reality was not raised (though at least some medieval philosophers understood these models realistically). Astronomical hypotheses were "*certa instrumenta*", as Erasmus Reinhold put it in 1551, for the construction of astronomical tables. Copernicus's heliocentric system was achieving this goal more efficiently than Ptolemy's, and this accounted, according to Duhem, for the fact that it started to win the attention of astronomers in the saving-the-phenomena tradition.

So, for Duhem astronomy aimed merely at calculating the various astronomical phenomena; astronomical hypotheses were (by default) fictions and not realities, whose "whole purpose" was to save the phenomena; hence, it's only natural that astronomers use different hypotheses to do their business ([1908] 1969, 22). Then, Osiander's plea for instrumentalism was just the culmination of a long instrumentalist tradition.

It should be noted however that the key instrumentalist argument that Osiander put forward, viz. (roughly) that truth cannot be known, is not part and parcel of the saving-the-phenomena tradition. Hence this tradition was not outright instrumentalist. Rather it was a mixed bag. There are a number of important issues that ideally should be discussed here. One of them is the relation between physics and astronomy in the ancient Greek tradition and beyond. It was physics which was looking for causes (the essence and the power) of things, whereas astronomy (being in effect applied geometry) relied on principles of physics (e.g., that orbits are circular) but aimed at capturing the motions of the heavenly bodies and not, at least in the first instance, their causes. This is something that had to do with the nature of astronomy qua science: it relied heavily on mathematical (i.e., geometrical) models and there can be (in fact there are) more than one of them to save the appearances. Mathematics was taken to be an instrument that was transforming astronomy into an exact science, leading it to exact predictions with ever-increasing levels of accuracy. As Samuel Sambursky put it, in his classic study of ancient Greek science "Mathematics had transformed astronomy into what Iamblichus called a 'prognostic science'" (1962, 58). Still, it doesn't follow that the causes are unknown or unknowable. Nor does it follow that astronomy was an instrumentalist science. This kind of view, fostered by Duhem's reading of the saving-the-phenomena tradition, was challenged by GER Lloyd in his (1978).

A very interesting case was Geminus's account of Posidonius's *Meteorologica,* which is reported by Simplicius, Aristotle's commentator, in his Commentary on Physics, Book 2 (cf. 2014a). According to Geminus, despite the fact that physics and astronomy have different aims, they study the same entities (e.g., "that the sun is a sizeable body, that the earth is spherical") albeit from a different perspective and using different methodologies. Whereas physics treats of heavenly bodies qua substances and examines their powers, "the astronomer argues from the properties of their shapes and sizes, or from quantity of movement and the time that corresponds to it" (292, 7–8). So, astronomy, qua mathematical science, deals only with the geometrical properties of bodies and the quantitative aspects of local motion. These

are abstractions from the actual properties of bodies. Hence, unlike physics, which "often touches on causes" (292, 10) astronomy is not in the business of causal explanation. This happens at three levels.

First, astronomy doesn't "pay any great attention to causes" (292, 12), that is, to actual physical bodies, since the objects of astronomy are abstractions. That's clearly the case when the astronomer represents "the earth or the heavenly bodies as spherical" that is as perfect spheres.

Second, astronomy "does not even attempt to find the cause" (292, 12), For instance, astronomy is not concerned with *why* an eclipse happens but with *when*.

Third, occasionally astronomy relies on assuming "certain orbits by whose presence appearances will be saved" (292, 15). That's the territory of epicycles and eccentrics. This level is non-causal because the assumptions made (the orbits chosen) are constrained by the demand of saving the phenomena. In this sense, any assumption will do, provided that it saves the phenomena.

But that's only half of the story. The other half, so to speak, is in what Geminus, as reported by Simplicius, goes on to say, viz., that *the very idea* of saving the phenomena aims to produce hypotheses such that "the treatment of the planets is squared with the accepted method of causal explanation" (292, 19–20). That is, the various hypotheses should be in accord with the principles of physics so that a causal explanation of the motions of the planets is possible. In particular, this causal explanation should be based on principles of accepted physics, viz., that the motions of heavenly bodies are uniform, regular and always in the same direction. So, it turns out that the saving-the-phenomena tradition is quite far from being instrumentalist, pure and simple. Astronomy was taken not to aim, in and of itself, at causal explanations of the motions of heavenly bodies, since (a) it deals with abstractions and (b) it's not concerned with the reason why X happens, but with when it happens.

In achieving (b) astronomy relies on hypotheses which, though they do not offer causal explanations themselves, they should be constrained by principles which are causal-explanatory, viz., that the motions of heavenly bodies are uniform, regular and always in the same direction. Now, it turns out that there are more than one hypotheses which save the appearances consistently with the causal-explanatory principles of physics. Astronomy has to state and examine all of them; it's not part of its business to pick one of them, simply because of (b) above. Hence, astronomers will aim to examine all relevant mathematical models.

The problem of saving the appearances by relying on correct principles of physics was accentuated by the fact that the motions of the planets presented various irregularities, such as the retrograde motion of Mercury and Venus. These were apparent motions which had to be explained away by reference to real motions – real in the sense that they satisfied the principles noted above, viz. uniformity, regularity and uni-directionality. Which of the possible real motions were actual was not the business of the astronomer to find out. Geminus notes this very well when he points out that astronomers should "enquire closely in what ways it is possible for these appearances to be produced" (292, 18–19), and hence to save the appearances, whilst at the same time the treatment of the planets offered should be "squared with the

accepted method of causal explanation". This dual task, as it were, renders astronomy peculiar. On the one hand, it should investigate all possible geometrical arrangements of heavenly bodies which save the appearances. Hence Geminus reports the heliocentric model of Heraclides of Pontus adding that it's not the concern of astronomers to find out which heavenly body, by nature, is at rest and which by, nature, moves. Rather they should look for configurations of motion and rest among the planets, the sun and the earth consistent with the appearances. On the other hand, however, astronomers should rely on "basic" principles that identify the real motion and properties of the heavenly bodies, viz., that "the dance of the heavenly bodies is simple, regular and ordered" (292, 27).

Now, Duhem treats Geminus's views as instrumentalist. He writes that Posidonius as reported by Geminus appeals to equivalent hypotheses "in order to drive home the astronomer's inability to grasp the true nature of the heavenly motions ([1908] 1969, 11). But, there is no such point in Geminus's account. In fact, if we look at Simplicius's own views, we realise that this tradition was not instrumentalist. Simplicius was an Aristotelian. No wonder then that he quotes Aristotle with approval when he says that "it would be absurd for the natural scientist to know nothing of the essential properties of natural bodies" (292, 33).

So, what's the point of the saving-the-phenomena tradition? For Simplicius, the key to understanding this view lies in the irregularities there are in the motions of the heavenly bodies. These are merely appearances in the sense that though the motions of the planets are regular and uniform they *appear* to us differently. In his Commentary of Aristotle's *On the Heavens* (2.10–2.14) (cf. 2014b), Simplicius notes: "The true account does not accept that they [the planets] stand still or move backward, or that there is in addition or subtraction in the numbers of their motions [i.e., their speed changes] even if they are observed to move in this way, nor does it admit hypotheses of this kind; rather it demonstrates that the heavenly motions are simple, circular, uniform, and ordered, using as evidence their substance" (488, 10–15). Astronomers, he adds, were unable to "grasp with precision how what occurs in the heaven is only the appearance of their condition and not the truth"; hence, they "were content to find out on what hypotheses the phenomena concerning the stars which are said to wander could be preserved by means of uniform, ordered, circular motions" (488, 17–19).

In citing this passage, Duhem takes it that Simplicius considers the various astronomical hypotheses to be fictions and not realities ([1908] 1969, 22), And yet, the only fictions in Simplicius's account are the supposed irregularities in the motions of the planets. The contrast Simplicius draws is between the fictitious (i.e., apparent) irregularities and the true circular and uniform motions.

What's the point of all this? It tuns out that this textbook case of an 'instrumentalist movement' is not really (or purely) instrumentalist. Astronomical theories were never 'nothing but' mere instruments dealing in fictions. They were not offering causal explanations of the planetary motions and yet the geometrical models were constrained by causal principles. Hence the fact that astronomical methods, bound up as they were by geometrical models, could not map out the true motions of the planets implies noting about an inherent inability to find out the true motions. After

all, we know that the very issue of 'the system of the world' was settled by dynamical considerations after Newton showed that the centre of gravity of the solar system was close to the sun.

What this case shows, I think, is that the case for compatibilism is not brought out by historical facts. On the contrary, the instrumentalist credo, captured nicely by the Osiander quote above, sets severe limits to any claim of compatibility. As noted already, the 'nothing but' view is based on an argument, viz., (roughly) that truth cannot be known, that no realist would endorse and which we have every reason to believe it's incorrect.

12.3 Varieties of Instrumentalism

What is instrumentalism? Here are some ways to describe it:

> The instrumentalistic fictionalist (or "instrumentalist" as we shall call him) does not propose to eliminate the non-E [i.e, the theoretical] portion of science, but simply to treat it differently from the E-portion [i.e., the observational]. Unlike the latter, which he holds significant and therefore true or false, confirmable or disconfirmable, and possibly expressive of belief, he treats the non-E portion as machinery, not to be qualified in any of these ways, though useful in the scientist's work. (. . .) Though, like all fictionalists, the instrumentalist holds that a uniformly significant E is not capable of absorbing all of science as a *de facto* body of doctrine, he is able to make a weaker affirmation, i.e,, he can say that E as he construes it *is* capable of expressing all *genuine assertions or beliefs* contained in the body of science (Scheffler, 1963, 186).

> [Instrumentalists] are those who claim that scientific theories are not true accounts of the world; theories are not so much true as *useful*. Since the task of scientific theories is to organise the data of our experience in such a way that predictions about, and eventual control of, the future are possible, theories are simply *instruments*; theoretical objects are by and large convenient fictions (Lambert & Brittan, 1987, 149).

> [S]cientific instrumentalism doubts that theoretical entities (. . .) literally exist, or at least that we can never know of their existence. Rather we are only justified in assuming the existence of *observable* entities or properties. Theoretical concepts possess only an instrumental function or the purpose of a 'most economical representation' of empirical knowledge, as Ernst Mach put it (Schurz, 2014, 291).

Formulations such as the above are quite typical. They are not equivalent with each other but this need not detain us here. What matters is that instrumentalism is a rival of scientific realism; and hence a kind of anti-realism. Two are the main tenets.

The first is *denialism*. Instrumentalism denies the existence of a certain kind of entities. It denies those entities which are required for the truth of a theory and they do not belong to an elite set of entities which satisfy a certain characteristic, typically observability. Hence, instrumentalism takes it that there are entities which are real and are, or could be, given to us in experience. Let's call them OK-entities. But instrumentalists also recognise that the putative content of scientific theories exceeds whatever can be asserted of the OK-entities. Of this excess content, they claim that it

is a *mere* instrument, albeit useful, which facilitates predictions about, and classification of, the OK-entities.

Hence, the second tenet of instrumentalism is *non-eliminativism*: theoretical discourse is not eliminable and hence its content is not fully captured in a language which refers only to OK-entities. Given denialism, the excess content of theories cannot be about not-OK entities; hence, it's about (useful) fictions.

Let's call the advocate of both tenets, fictionalist instrumentalist, or simply fictionalist. Not all self-proclaimed instrumentalists are denialists; nor non-eliminativists. Hence, not all instrumentalists are fictionalists. Take John Dewey, for instance, who invented the term "Instrumentalism". He took it to be a view akin to pragmatism; a view about the "proper" objects of science and their relations to the things of ordinary experience. For Dewey the objects of science are tools or means for supplying connections among otherwise disconnected appearances or experiences. He has that vivid image of a series of mountains which are submerged except for their tops. These tops, which correspond to appearances, would be totally unrelated and disconnected, if it was not accepted that there are submerged – hence unseen – mountain parts underneath the visible peaks. In a similar fashion, science, says Dewey, posits theoretical objects whose constant relationships explain the order of appearances. These instrumentalities, as Dewey called the objects that scientific theory posits, are more 'real' than the objects of common sense only in the sense that they fulfil "the function of instituting connections" among the latter: they weave together otherwise disconnected objects of appearances into "a consecutive history" (1958, 139). Dewey's instrumentalism denies that there is anything more to the reality of the "conceptual objects" of science.

However, Dewey-instrumentalism is not denialist. It does not deny the reality of the objects of science. Rather it puts a gloss on what it is for them to be real. Dewey-instrumentalism is a kind of *contextualism*: reality is not an absolute category firmly attributed to some entities and firmly denied to others. Rather, what counts as real depends on the context. Speaking in an auto-biographical way, he noted: "That the table *as* a perceived table is an object of knowledge in one context as truly as the physical atoms, molecules, etc. are in another situational context and with reference to another *problem* is a position I have given considerable space to developing" (1939, 537). The context and the problem are determined, at least partly, by the things one does with an entity and by the role an entity plays within a system – for instance, one can put books on a table but one cannot put books on a swarm of molecules. One may well question the motivation for this view, since it is not clear, to say the least – how contexts are separated. But here, in any case, is a variant of instrumentalism that does not reduce theories to calculating devices and does not deny that (in a sense, at least) explanatory posits are real.

Take Philipp Frank, who's been one of the very few outspoken instrumentalists. In his (1932), he started his treatment of causality by stating that science is an instrument and that scientific formulas are not propositions about the real world. Science, he argued, aimed at prediction and theories are merely tools for this aim – in particular, symbolic tools that do not (aim to) represent the real world. Theories, for

Frank, are neither true nor false; they have only instrumental value that is cashed in terms of predictions of future observations on the basis of present ones. As he characteristically put it, a theory can be called 'true' or 'false' in the sense that '"a badly sharpened knife is a false instrument for cutting" (1932, 19) and not in the sense in which statements about ordinary objects of experience are false (or true).

Let us call this *non-cognitivist instrumentalism*. It is non-eliminativist but only in the sense that theories (the not-E portion of theories, as Scheffler has put it) lacks assertoric content and is only a means to facilitate moving from claims about observables to predictions about observables. In this sense, theoretical premises are merely inference-tickets. And it is clearly denialist since it takes it that all this theoretical discourse comes ontically to nothing, since there is nothing to be about. As Frank put it, a scientific theory is not a *replica of the real world*.

Non-cognitivist instrumentalism is not taken seriously any more. The reason is simply that, judged by its fruits, it does a very poor job even when it comes to explaining how prediction and control are possible. One important argument against non-cognitivism (of the sort espoused by Frank) is that it is a reconstruction of science that turns a perfectly meaningful practice – where there is communication and understanding – into a meaningless manipulation of symbols underlied by problematic and context-dependent rules that connect some of the symbols with experience (and hence give them some partial meaning). More importantly, however, non-cognitivist instrumentalism offers a causally disconnected image of the world. Given its way of reading the theoretical superstructure of theories, it's not even possible that the observable phenomena are connected to each other via processes and mechanisms that involve unobservable entities. The problem here is not that causal laws become very complicated by being meant to connect only observable magnitudes. Rather, the problem is that causal laws become arbitrary, since they are *after all* formulated by means of symbols such that, were they to be taken literally would refer to unobservable entities, but, being not-taken literally, there is simply no fact of the matter as to whether these or other (incompatible) causal laws describe the workings of nature, provided that they are all co-ordinated in the appropriate manner with observables.

12.4 A Tension for Fictionalism

The instrumentalist "movement" is too diverse. I argued that the core position involves denialism and non-eliminativism, aka fictionalism.

Fictionalism is a view with a long pedigree. It was defended for the first time by Hans Vaihinger in 1911. It is the view that some entities whose existence is implied by the truth of a theory are not real, but useful fictions. Hence, on the fictionalist approach, scientific theories which are prima facie committed to the existence of unobersvable entities are false, simply because there are no such entities for the theories to be committed. On this view, to say that one accepts the proposition that *p* *as if* it were true is to say that *p* is false but that it is useful to accept whatever p

asserts as a fiction. Introducing this view in *The Philosophy of As If*, Vaihinger noted that what is meant by saying that matter consists of atoms is that matter must be treated *as if* it consisted of atoms. But what does it mean that "matter must be treated as if it consisted of atoms?" As he said: "It can only mean that empirically given matter must be treated as it would be treated *if* it consisted of atoms" (1911, 93). Though it is false that matter has atomic structure, Vaihinger argued that the as-if operator implies a decision to maintain formally the assumption that matter has atomic structure *as a useful fiction*. Hence, we may willingly accept falsehoods or fictions if this is useful for practical purposes or if we thereby avoid conceptual perplexities. We then act *as if* they were true or real. It should be noted that Vaihinger's fictions are not *just* false assumptions – they are *knowingly* false and, in their stronger version, impossible to be true. Vaihinger distinguished between fictions and hypotheses – the latter can be true or false and it is an open issue what they are; the former *cannot* be true.

In contemporary philosophy, fictionalism was famously introduced by Hartry Field as an alternative to mathematical realism. On this view, there are no numbers (or other mathematical entities), but mathematics is still useful being a conservative extension of mathematics-free (that is nominalistic) scientific theories: it facilitates deductions; all that is required for this is that a mathematical theory M be consistent.

Now, it should be stressed that fictionalism allows for an important distinction to be drawn between what the theory says that the world is like and what the world is *really* like. It leaves a gap between the claim (a) that *T* is false and the claim (b) that everything in experience is *as if T* were actually true. It is fully consistent to argue, according to fictionalism, there are no electrons and yet everything in experience is *as if* there were actually electrons. This combination of claims is logically consistent precisely because fictionalism takes theories at face-value: it does not re-interpret them; nor does it claim that they refer to entities other than those implied by a literal understanding of them. In other words, fictionalism is a non-eliminativist position: theoretical entities, were they to exist at all, they would be irreducible to observable entities; but (the fictionalist says) they do *not* exist.

Putnam, who was a major critic of fictionalism in the 1970s, offered some credit to fictionalism, and rightly so, precisely because fictionalism allows for a distinction between a theory's being true and everything in experience being such that the theory is actually false. This leaves it entirely open that the theory is empirically adequate and yet false. Why then should fictionalism be rejected? Because it does not make sense to have a *merely* fictionalist stance towards a theory that has been accepted and employed in the explanation and prediction of observable phenomena. The fictionalist would typically read the theory literally, would treat the theoretical concepts as indispensable and would accept a theory 'for scientific purposes' but would refrain from commitment to the reality of the entities implied by the theory since she would take it that theory – though perhaps empirically adequate – is *false*. But what possibly could show to a fictionalist that the theory is true? Putnam (1971) takes it that the fictionalist would demand a deductive proof of the theory and rightly objects that if this were the golden standard for acceptance as true, no non-trivial observational statements would be accepted as true either. The fictionalist would end

up with scepticism. So Putnam challenges the fictionalist to draw and motivate a robust distinction between rationally accepting a theory T (but treating its supposed entities as useful fictions) and rationally accepting that T is true. As Putnam (1971, 354) put it, if one rationally accepts a theory for scientific purposes, "what further reasons could one want before one regarded it as rational to *believe* a theory?" His answer was that these reasons are good enough! Being a fallibilist, Putnam adds that belief in a theory is belief in that the theory is "an approximation to the truth which can probably be bettered".

In light of this, we can see that there is a tension between denialism and non-eliminativism. If t-terms (which taken literally putatively refer to not-OK entities) cannot be dispensed with when a theory is presented (non-eliminativism) and if, on top of this, they do play a useful role in prediction and classification, is it *reasonable* to adopt denialism? In other words, if theories are indispensable for achieving what instrumentalists are after, viz. classifying and predicting facts about the world of experience, why should we deny the reality of the alleged not-OK entities whose existence is implied by the theories?

The non-eliminativist tenet implies that its advocate takes it that the theory has excess content which, on the face of it, is about the not-OK entities. But denialism takes it that there are no not-OK entities for theory to be about. But if reasons to accept a theory are *ipso facto* reasons to believe it is true, non-eliminativism is not great bedfellow for denialism.

12.5 Craig to the Rescue

Can fictionalism be improved upon? Naturally, if it could be shown that t-terms are dispensable, that they can be eliminated without loss of OK-content, then fictionalism would avoid the foregoing tension. Even if theories *appear* to have excess content over their OK-content, this is illusory since, ultimately the vehicle for talking about the alleged not-OK entities, ie., theoretical vocabulary, is dispensable. And that's the reason why Craig's theorem fell like manna from heaven to all aspiring instrumentalists. This theorem was taken as proof that a certain theory T which implies – if literally understood and true – the reality of certain entities can be replaced by another theory T' which does not imply commitment to the reality of the 'suspicious entities'.

In his Ph.D Thesis in 1951, logician William Craig constructed a general method according to which given any first-order theory T and given any effectively specified sub-vocabulary O of T, we can construct another theory T' whose theorems are exactly those theorems of T which contain no constants other than those in the sub-vocabulary O. But as Hempel noted and Craig (1956) himself proved, the application to philosophy of science of Craig's theorem yields the following powerful result: given any first-order theory T and any effectively specified observational sub-vocabulary V_O of T, one can construct another theory T' whose theorems are exactly those theorems of T which contain no terms other than those already in the

sub-vocabulary V_O. What came to be known as Craig's Theorem is the following: for any scientific theory T, T is replaceable by another (axiomatisable) theory Craig (T), consisting of all and only the theorems of T which are formulated in terms of the observational vocabulary V_O. Craig showed how to construct these axioms of the new theory Craig(T). There will be an infinite set of axioms (no matter how simple the set of axioms of the original theory T is), but there is an effective procedure that specifies them. The new theory Craig(T) which replaces the original theory T is 'functionally equivalent' to T, in that all observational consequences of T also follow from Craig(T): the latter establishes all those deductive connections between observation sentences that the initial theory T establishes.

The gist of Craig's theorem is that a theory is a conservative extension of the deductive systematisation of its observational consequences. As Hempel put it in a note on Craig's theorem in 1954: "The method then yields a theory in terms of observables only which has, in Craig's (or Feigl's) terms, the same 'cash value' at the original theory with its constructs" (Feigl Archive, University of Pittsburgh. All rights reserved. Doc. HF 09-38-04). In a Conference on Physicalism in Princeton, April 1954, where Carnap was present, too, Hempel presented Craig's result as a dilemma: "There is also a theorem of Craig proved in an unpublished doctoral thesis which is damaging to the idea that a theory is necessary at all or to the idea that explanation is deduction" (Carnap Archive, University of Pittsburgh. All rights reserved. Doc. 090-62-06, p.

And yet, this promising route to fictionalism by means of Craig-eliminability was short-lived. One key reason was that Craig(T) as a replacement of the original theory T fell short of the very instrumentalist standards for a scientific theory. This was pointed out by Nelson Goodman – who was no friend of realism – in his review of Craig's (1956) in JSL in 1957. Goodman's point was that Craig(T) lacks a key virtue that theories should possess even by instrumentalist standards: that's not truth but *integration* – aka economy of thought. But Craig(T) will have an infinity of axioms, no matter how simple the original theory T might be. Hence, Goodman says:

> Evidently preservation of all theorems in non-suspect [i.e., observational] language is far from enough. What more, then, is required [for a good theory]? For one thing, that the replacing system [i.e., Craig(T)] have an appreciable degree of deductive coherence or economy. The chief purpose of proof in a philosophical system is less to convince of truth than to integrate. If every theorem has its own postulate, no integration is achieved (1957, 318).

Putnam (1965) mounted a formidable attack on the philosophical significance of Craig's theorem arguing that a) theoretical terms are meaningful, taking their meaning from the theories in which they feature and b) scientists employ terms like 'electron', 'virus', 'spacetime curvature' and so on – and advance relevant theories – because they wish to *talk about* electrons, viruses, the curvature of spacetime and so on; that is scientists want to find out about the *unobservable* world. Theoretical terms provide scientists with the necessary linguistic tools for talking about things they want to talk about. Besides, the very idea of Craig-eliminability requires the clear separation of the theoretical vocabulary from the

observational one but this is wishful thinking since there is no such sharp distinction. Rather, this distinction is vague and context-dependent.

If abandoning non-eliminativism is not open to aspiring fictionalists, would abandoning denialism be? Strictly speaking one cannot be a fictionalist without denialism. Yet, more loosely speaking, one can combine the thought that not-OK, i.e., theoretical, entities are useful fictions with agnosticism, as opposed to denialism. On this combination of views, one need not deny the reality of the entities assumed by a literal reading of a theory – but instead be *agnostic*. This is the position articulated by van Fraassen (1980). On this view, the failures of fictionalism (or denialist instrumentalism) does not make scientific realism the only rational option. An agnostic variety of instrumentalism is not, *ipso facto*, ruled out: one can always remain agnostic as to the truth-value of the particular theoretical descriptions of the world offered by a theory, even though one uses the theory for all scientific purposes. Notably, van Fraassen capitalised on the very fact that Putnam credited to fictionalism, viz., that there is a gap between empirical adequacy and truth, and argued that this very gap allows for a position such that "theories could agree in empirical content and differ in truth-value" (1980, 36). This position, once occupied, would have scientific realism going for a "leap of faith" to close the gap between empirical adequacy and truth. And as van Fraassen stressed: "The decision to leap is subject to rational scrutiny, but not *dictated* by reason and evidence" (1980, 37). But given that I've dealt with van Fraassen extensively elsewhere (e.g., my 2012), I will say no more here.

12.6 Compatibilism I

Is denialist instrumentalism dead then? The answer that became possible in the 1960s was that if it is, so is realism. Why? Because the two views are compatible with each other, after all. In its very, influential *The Structure of Science*, in 1961, Ernest Nagel famously argued that realism and instrumentalism are merely *different languages* about theories and the choice between them is only a choice of the preferred mode of speech. Here is how he put it:

> It is therefore difficult to escape the conclusion that when the two apparently opposing views on the cognitive status of theories [realism and instrumentalism] are each stated with some circumspection, each can assimilate into its formulations not only the facts concerning the primary subject matter explored by experimental inquiry but also all the relevant facts concerning the logic and procedure of science. In brief, the opposition between these views is a conflict over preferred modes of speech (1961, 152).

Nagel takes instrumentalism to be the view that theoretical propositions are, in effect, inference tickets, which licence transitions from observational premises, which are truth-evaluable to observational conclusions. Being disguised rules of inference, theoretical claims lack truth-values. Hence, they are not in the business of assertion. Here is a striking passage: "a theory is held to be a rule or a principle for analysing and symbolically representing certain materials of gross experience, and at

the same time an instrument in a technique for inferring observation statements from other such statements" (1961, 129). This passage is striking because it aligns Nagel with Frank's non-cognitivism. But then the argument against non-cognitivism, viz., that it makes a mockery of a perfectly meaningful practice, holds against Nagel too. But still one may wonder: how can it be that realism and instrumentalism are compatible?

Nagel's argument capitalises on the difference between treating a statement as a rule or as a premise. Take, for example the statement 'All As are B'. On the face of it, it looks like it's a universal generalisation. It can be a premise in the following argument: a is A; All As are B; therefore a is B. However, it can be the case that we can move directly from 'a is A' to 'a is B", via the material rule of inference: from 'x is A' infer 'x is B'. This rule is material and not formal in the sense that it implicates a specific subject-matter. For instance, from 'copper is a metal and 'copper is heated', we can infer 'copper expands' using the rule: from 'x is metal' and 'x is heated' infer 'x expands'. Now, Nagel takes this distinction between premises and rules of inference to imply that "a given statement may function as a premise in one context but may in effect be used as a leading principle in another context, and vice versa" (1961, 138). And from this he infers that "there is on the whole only a verbal difference between asking whether a theory is satisfactory (as a technique of inference) and asking whether a theory is true (as a premise)" (1961, 139).

This 'irenic instrumentalism', as Sellars (1965) put it, faces two important difficulties. The first is that a theory, on the face of it, is a putative *map* of reality and not a set of principles for *constructing maps*. This is acknowledged by Nagel (1961, 139) but its implications are not fully appreciated. The theory has a part (the E-portion, as Scheffler put it) which is a map of reality, according to instrumentalists. This part offers the premises upon which the rules operate and also contains the conclusions of the rules. The question then is: how is the distinction between rules and premises to be drawn? Why, in other words, assume that some of the statements of the theory have a fact-reporting role, while others do not? This problem is far from trivial, since it amounts to the problem of whether there is a sharp line between observational statements (for which the issue of factual truth is open) and theoretical statements (for which there is no issue of factual truth). And we all know that there is no sharp line to be drawn.

The second difficulty relates to Nagel's idea that the theory should be "satisfactory" as a rule of inference. As Sellars pointed out, what else could 'satisfactory' mean other than truth-preserving? Suppose that observational statement O1 entails O2, via statements T1&T2&T3. The truth of T1&T2&T3 explains why the inference O1 |- O2 is truth preserving. Lacking this explanation, that O1→O2 is true is totally ungrounded.

This difficulty becomes a serious problem if we think of the role theories play in the confirmation of observational statements, the idea being that theories are often necessary for the establishment of inductive connections between seemingly unrelated observational statements. Here is Putnam's (1963) own example. Consider the prediction *H*: 'When two subcritical masses of U_{235} are slammed together to form a supercritical mass, there will be a nuclear explosion'. *H* could be re-written in

an observational language – that is without the term 'Uranium $_{235}$' – as O_1: 'When two particular rocks are slammed together, an explosion will happen'. Consider now the available evidence, namely O_2: 'Up to now, when two rocks were put together nothing happened'. Given this, it follows that prob(O_1/O_2) is very low, if it can be determined at all. But consider the posterior probability of O_1 given the past evidence and the whole of atomic theory T which entails that the uranium rocks would explode if critical mass were attained quickly enough. It is obvious that prob (O_1/O_2&T) is now determined and much greater than prob(O_1/O_2).

All this speaks directly against instrumentalism. But there is still the following worry: how could it *possibly* be that realism and instrumentalism are compatible, given that realism affirms the existence of t-entities while instrumentalism denies it? Nagel claims that the very idea of the physical existence or reality of entities is vague. He takes it that "real" in an "honorific" term which is used "to express a value judgment and to attribute a 'superior' status to the things asserted to be real". And he adds: "it would be desirable to ban the use of the word altogether" (1961, 151). But the point is not to find a way for the instrumentalist to understand 'being real' such that they can assert that e.g., atoms are *real*. Rather the point is that the reality attributed by an instrumentalist to atoms is bound to be always 'second-class' vis-à-vis the reality attributed to tables and pointer readings. And that's exactly the point Nagel *missed* when he stressed:

> to assert that in this sense atoms exist is to claim that available empirical evidence is sufficient to establish the adequacy of the theory as a leading principle [i.e., rule of inference] for an extensive domain of inquiry. But (. . .) this is in effect only verbally different from saying that the theory is so well confirmed by the evidence that the theory can be tentatively accepted as true (Nagel, 1961, 151).

12.7 Compatibilism II

A few years after the publication of *The Structure of Science*, Carnap advocated compatibilism in his *Philosophical Foundations of Physics*, referring with approval to Nagel. He said:

> It is obvious that there is a difference between the meanings of the instrumentalist and the realist ways of speaking. My own view, which I shall not elaborate here, is that the conflict between the two approaches is essentially linguistic. It is a question of which way of speaking is to be preferred under a given set of circumstances. To say that a theory is a reliable instrument-that is, that the predictions of observable events that it yields will be confirmed-is essentially the same as saying that the theory is true and that the theoretical, unobservable entities it speaks about exist. Thus, there is no incompatibility between the thesis of the instrumentalist and that of the realist. At least, there is no incompatibility so long as the former avoids such negative assertions as, '... but the theory does not consist of sentences which are either true or false, and the atoms, electrons, and the like do not really exist' (1966, 256).

And in a footnote he added: "An illuminating discussion of the two or three points of view on this controversy is given by Ernest Nagel, The Structure of Science

(New York: Harcourt, Brace & World, 1961), Chapter 6, 'The Cognitive Status of Theories'."

Carnap's point is striking, since as he emphatically notes, realism and instrumentalism are compatible (they can both be true at the same time) provided instrumentalism abandons Denialism! Insofar as instrumentalist avoids negative claims of the kind 't-entities do not exist' or 't-propositions are not assertions', then instrumentalism is compatible with realism. But can there be such an instrumentalist?

A few years later in the paperback edition of the 1966 book, Carnap replaced the previous paragraph with the following and dropped the reference to Nagel. The paragraph now reads as follows:

> It is obvious that there is a difference between the meanings of the instrumentalist and the realist ways of speaking. My own view, which I shall not elaborate here, is that the conflict between the two approaches is essentially this. I believe that the question should not be discussed in the form: "Are theoretical entities real?" but rather in the form" "Shall we prefer a language of physics (and of science in general) that contains theoretical terms, or a language without such terms?" For this point of view the question becomes one of preference and practical decision (1974, 256).

The footnote now reads: "In my view greater clarity often results if discussions of whether certain entities are real are replaced by discussions of preference of language forms. This view is defended in detail in my *Empiricism, Semantics and Ontology* (1950)".

The change is rather deep and interesting. Carnap moved from the strong compatibilist claim that there is no need to choose among Realism and Instrumentalism because 'they say the same thing' to the claim that commitment to entities is a matter of a choice of language (that is a choice to use of t-terms or not); and hence it's a practical issue in line with the views in his *Empiricism, Semantics and Ontology* (1950). So do or don't Realism and Instrumentalism say the same thing? And given that the famous (1950) paper predates both Nagel's account as well as Carnap's strong compatibility view, what accounts for the change in the two editions and what does this change imply?

Between 1950 and 1966, Carnap had re-invented the Ramsey-sentence approach to scientific theories. I have told this story elsewhere in great detail (see my 1999, chapter 3). So I will be brief. Not only did Carnap know of Craig's theorem, but he also extended Craig's results to "type theory, (involving introducing theoretical terms as auxiliary constants standing for existentially generalised functional variables in 'long' sentences containing only observational terms as true constants)". (As it is reported in the protocol of the Los Angeles Conference in August 1955 – Feigl Archive 04–172–02:14). In the same protocol, Carnap is also reported to have shown that '[a]n observational theory can be formed which will have the same deductive observational content as any given theory using non-observational terms. (Namely, *by existentially generalising non-observation terms*)' (ibid., 19).

What's special about Ramsey-sentences? What are the advantages over Craig's method? Three points are relevant:

- RS is finitely axiomatizable.
- RS retains the logical structure of the original theory.
- RS lays a claim of satisfaction by certain entities (those in the range of the existentially bound variables).

As Carnap put it: "the Ramsey sentence has precisely the same explanatory and predictive power as the original system of postulates" (1966, 252). And it achieves this in a way that is tempting for an empiricist since it delivers us from metaphysics. There are no t-terms and predicates in the RS of a theory; hence the issue of their reference evaporates. Still, however, the RS implies existential commitments to "whatever it is in the external world that is symbolized by" the t-terms. *The vehicle of non-observational commitment is the existential quantifier.* Here is a long quotation characteristic of Carnap's enthusiasm:

> In Ramsey's way of talking about the external world, a term such as 'electron' vanishes. This does not in any way imply that electrons vanish, or, more precisely, that whatever it is in the external world that is symbolized by the word 'electron' vanishes. The Ramsey sentence continues to assert, through its existential quantifiers, that there is something in the external world that has all those properties that physicists assign to the electron. It does not question the existence-the 'reality'-of this something. It merely proposes a different way of talking about that something. The troublesome question it avoids is not, 'Do electrons exist' but, 'What is the exact meaning of the term 'electron'?' In Ramsey's way of speaking about the world, this question does not arise. It is no longer necessary to inquire about the meaning of 'electron', because the term itself does not appear in Ramsey's language (1966, 252).

Now, it's not surprising that the Ramsey-sentence RS(T) of theory T is consistent with T. RS(T) logically follows from T; it cannot be the case that RS(T) is false and T true. So is the point that RS-instrumentalism is compatible with realism? Is the Ramsey way instrumentalism? Grover Maxwell was quick to point out that it is not. Commenting on Carnap's use of the Ramsey sentence, Maxwell wrote to Carnap: "I disagree that thinking theoretical entities 'in the Ramsey way' should be associated with instrumentalism' (Maxwell to Carnap, 24 June 1966; Carnap Archive 027–33–29). Carnap replied on 9 December 1967 conceding the point and putting the blame to carelessness and confusion on his part:

> You are quite right in the one critical remark you make, that the Ramsey way should not be associated with instrumentalism. In an earlier version of the manuscript I had distinguished three instead of two views on the question of the reality of entities, by splitting off instrumentalism into two forms, a negativistic one and a neutral one which I identified with the Ramsey way. Then a reader of the manuscript pointed out that the distinctions were not in agreement with the customary terminology; in particular that the term 'instrumentalism' is always used in the negativistic sense. Then I made a radical change, distinguishing only two points of view. This I did in great haste and so I mixed things up. For a future edition of the book I have decided on a reformulation which you see on the enclosed sheet (Carnap Archive 027–33–28).

If Carnap can be confused, anyone else can be. Certainly confusion did play a role. On the face of it, the Ramsey way is neither realist, nor instrumentalism; yet it

appears to be a halfway house. But there is a deeper reason why Carnap took the Ramsey way to be instrumentalist.

In his reply to Hempel in the Schilpp volume, Carnap says:

> the Ramsey-sentence does indeed refer to theoretical entities by the use of abstract variables". But he immediately adds that "these entities are not unobservable physical objects like atoms, electrons, etc., but rather (*at least in the form of the theoretical language which I have chosen in [MCTC] §VII*] purely logico-mathematical entities, e.g. natural numbers, classes of such, classes of classes, etc. (1963, 963).

On this reading, the Ramsey sentence says that "the observable events in the world are such that *there are numbers*, classes of such, etc., which are correlated with the events in a prescribed way and which have among themselves certain relations; and this assertion is clearly a factual statement about the world' (ibid.).

Here again I invite the reader to look at my (2007) for the details of Carnap's views about the language of science and his claim that every legitimate scientific concept is extensionally equivalent with a certain mathematical function. For our purposes, it's enough to say that it is this particular reading of RS(T) which makes it consistent with instrumentalism. To sum up, RS(T) does not deny the existence of unobservable realisers of the theory; it's not denialist; yet it's fully consistence with the claim that the realisers of the theory is not physical stuff (i.e., physical unobservables) but instead mathematical entities. And to this extent, the Ramsey way can be seen as a form of instrumentalism. It's a *kind* of instrumentalism since RS(T) can be satisfied without having any commerce with those parts of the physical world, if there are any, that are not given to us I experience. But it's a *strange* kind of instrumentalism, since the satisfiers of RS(T), being mathematicalia, are not given to us in experience.

Carnap's neo-Pythagoreanism, as Sellars (1965, 177) put it, might secure some kind of instrumentalist reading of Ramsey sentences but it creates more problems than it solves. A rather severe one is that, on the face it, saddles instrumentalism with a rampant commitment to abstracta. So an instrumentalist might be accused of swallowing a camel and straining at a gnat, the camel being absracta and the gnat being physical unobservable entities. I think that's why Carnap in the end abandoned the strong Nagel compatibility and opted for the weaker claim that the choice between instrumentalism and realist is a matter of choice of language, where picking a language is a matter of practical expedience. Hence the shift from Nagel's *Structure* to his own *ESO*.

12.8 Compatibilism III

In a seminal piece that appeared in 1989 Stein aimed to save compatibilism by noting that "between a cogent and enlightened 'realism' and a sophisticated 'instrumentalism' there is no significant difference – no difference that makes a difference" (1989, 61). But what is this 'sophisticated instrumentalism'? Let's call 'philosophical type

X' the view, or the philosopher who holds the view, that these is no difference between sophisticated instrumentalism and enlightened realism. The basis for X is the claim that theories are instruments. Yet, it's not the case that they are nothing but instruments. According to Stein, a 'sophisticated instrumentalist' should extend further their conception of what the theories are instruments *for*. Going beyond the view that a theory is a *mere* instrument for calculating the outcomes of experiments, X must also encompass the claims that a theory is an instrument for *representing* the phenomena of nature (1989, 50) and for *guiding* the inquiry (1989, 52).

Let us briefly see the concessions realists should make to be X. Stein's key complaint is that the unreconstructed realist is committed to a semantic theory of the relation between theory and the world such that t-terms have factual reference. To be more precise, Stein seems to be driven by an epistemic worry, which he puts as follows: 'How can you *know* that things are as you say they are? If the claimed "reference" of the theory is something beyond its correctness and adequacy in representing phenomena' (1989, 50). The implication is clear: any realist claim of truth and factual reference of a theory that goes beyond whatever is involved in rendering a theory a correct and adequate representation of the phenomena is a difference that makes no difference.

A whole paper could be written on the issue of the correct and adequate representation of the phenomena. Hence, I will restrict myself to a general argument, which starts with a question: what more, or other, than a correct and accurate representation of the phenomena should or would a realist demand? The obvious answer is: nothing. When a realist takes theories to be, roughly put, maps of the world, the aim is a correct and accurate representation of the world and hence of the phenomena (narrowly understood to refer to *observable* phenomena). This map won't be a mirror of reality, for reasons that have to do with the fact that representation involves idealisations, abstractions and perspectives. Still, it can be more or less accurate and more or less correct if it is world-involving; that is, if it identifies relevant causal factors and, more broadly, ways in which the phenomena are brought about. In the case discussed in the second section, for instance, most (all?) of astronomers knew the limitations of astronomical representations of the planetary motion and yet they were trying to make them world-involving by basing them on what they had taken to be correct general principles.

If the realist doesn't have to concede much to become Stein's Philosophical Type X, how about the instrumentalist? For one, an instrumentalist would have to abandon denialism. If the correct and accurate representation of the phenomena requires commitment to theoretical entities, then anyone of philosophical type X would not be a denialist. For another, if instrumentalism takes 'representation' of the phenomena to involve only OK-entities, or to be such that whatever goes beyond the OK-entities are mere fictions, then theories end up being *nothing but* instruments.

If going for a notion of representation does not help the compatibilist cause how about the extra requirement that theories should play a role as resources to inquiry? This requirement has been stressed by Richard Boyd as offering a methodological advantage to realism. Stein's position is the following:

> Either the methodological principles in question, and the fact of their success thus far in the history of inquiry, are susceptible of clear formulation in terms of the relations of theories to phenomena (and this, I think, is what Boyd intends) – in this case, the instrumentalist ought to adopt those principles, on the grounds of the evidence for their instrumental success; or they are not susceptible of such formulation – but then, where is the argument? In short, the instrumentalist's own principles entail no prohibitions: what the realist can do, the instrumentalist can do also (1989, 52).

Here again the keyword is "the relations of theories to phenomena". For Boyd and other realists like myself, the success of a theory is best explained by the theory's being approximately true. So theories are related to the phenomena by offering an approximately true account of them. Stein does not recognise approximate truth among the attributes that theories must have to serve as resources of inquiry. Instead he talks about the logical structure of the theories and their relation to phenomena and notes that "the attributes our theories must possess for them to serve as resources for inquiry are all attributes that concern the logical structure of the theories" (1989, 53).

What exactly is the logical structure of the theory and what are its attributes? Stein does not say anything directly on it. He does talk about mathematical structures, and gestures (uncharacteristically vaguely) that improved understanding consists in "structural deepening". But it's not clear at all how he understands the relation between mathematical structures and the physical world. What he is clear about is his insistence on the futility of "hypostatizing" (unobservable) entities. Now, to use Stein's ironic comment "it would be possible to do a lengthy dialectical number on this; but in brief": the long debates over structural realism have taught us that (a) it's hard to understand structures without entities and (b) hypostatising structures would not be less of a sin for an instrumentalist than hypostatising entities simply because they would end up with abstracta.

In any case, it might be that the logical structure of a theory is captured by the Ramsey-sentence RS(T) of a theory T. The chief relevant attributes of the Ramsey-sentence approach are (1) that T and RS(T) have the same logical-deductive structure and (2) they have exactly the same observational consequences. At the same time, RS(T) dispenses with the hypostatization of entities, since it replaces all theoretical predicates with existentially bound variables.

But if *that* is the 'sophisticated instrumentalism' that Stein has in mind, then there is not much progress since Carnap's compatibilism. To sum up: Stein claims that there is what I called a philosophical type X which going beyond the nothing-but-ness of instrumentalism by incorporating the two following roles at science:

- The theories represent the phenomena of nature.
- The theories are guides to inquiry.

According to Stein's main thesis, this philosophical type X is all we need to accept in order to understand science and its relation to the world, to reality. All the rest that the realist may add (genuine reference of the theoretical terms hence hypostatisation of the unobservable entities) is a difference that makes no difference. In other words, an instrumentalist can do whatever a realist does and because of that the claimed

difference between them does not exist. But it turns out that the sought after philosophical type X is, at best, out old fellow: Ramsey-sentences.

12.9 Compatibilism IV

It might be surprising that a leading current advocate of instrumentalism has been flirting with compatibilism. Indeed, Kyle Stanford has recently published an article titled: "'Atoms exist' is probably true, and other facts that shouldn't comfort scientific realists" (2015b). In this he says quite emphatically that the anti-realist need not and should not deny any particular existential commitment to unobservables issued by scientific theories. As he put it: "we should decline to saddle the scientific realist's historicist opponent with the belief that there are no such things as genes or atoms or that the terms 'gene' and 'atom' do not refer" (2015b, 411).

To motivate, and see the limits of, this irenic stance, let us define two philosophical types: let's call them R and I. Both types, presumably, have learned the lesson from theory-change in the history of science. Both types acknowledge that a number of current scientific beliefs will be overthrown in the future. Both types acknowledge that a number of particular claims made by current science, and the concomitant existential commitments, are part and parcel of any future science. Perhaps surprisingly, type I takes it that atoms and genes and other such stuff are real and that at least some of the claims made about them are true or roughly so. And yet, type I emphasises the discontinuities in theory-change, whereas type R the continuities. Is there a big difference between R and I? Or is the difference really a matter of where the emphasis is placed?

Note that type I is not an outright denialist. So type I is not a fictionalist. More interestingly, type I does not endorse the views of past compatibilists. So, if we think of type I as instrumentalist, what kind of instrumentalism is this? Before I answer this question, let me say something about scientific anti-realism in general.

Traditionally, scientific anti-realism has been an all-or-nothing stance towards theories. Better put, it's been a dichotomous position. The content of a theory is split into two principled parts – let's call them the OK-part and the not-OK-part. The dichotomy (OK/not-OK) has an epistemic motivation. The OK-part is about things (or that part of the world) we can have epistemic access to; the not-OK part is 'about' those things that were they existed they would cognitively impenetrable by us. The criteria of cognitive penetrability, and hence, of epistemic accessibility, differ. It might be perception and hence observability; or something more sophisticated. But whatever the line of distinction is, commitment is an all-or-nothing matter (see my 2009). All not-OK entities are suspicious: unreal, fictions etc. Instrumentalism is a species of antirealism; it too is dichotomous. Once the distinction between OK and not-OK parts is drawn, denialism is an attitude towards the not-OK part.

So back to the question above: what kind of instrumentalism is type I? Strictly speaking, it's not instrumentalism. There is no denialism; nor any sharp dichotomy between the OK and not-OK parts of the theory. Besides, there is no way to tell

ahead of time what unobservable (putative) entities will be accepted as real. Hence, the list of entities that will be accepted as real is open-ended. Stanford stresses:

> Even when we have found one of a theory's central existential commitments overturned in the course of further inquiry, many otherwise similar commitments have remained in place, so even in cases in which a successful scientific theory does have central existential commitments that will ultimately be judged to be false and/or central theoretical terms that will ultimately be judged non-referential, we should not imagine that we can specify in advance which existential commitments and/or theoretical terms these will be (2015b, 402).

Insofar as type I is instrumentalist, it's wait-and-see instrumentalism, to use John Earman's (1978) apt expression. Any putative entity that, were it to exist, it would be talked about in the not-OK part of the theory according to traditional instrumentalism, might be end up being talked about in the OK-part.

This last point is critical: type I, *aka* wait-and-see instrumentalism, leaves it up to science and the evidence-based consensus-forming mechanisms, to tell us what there is. There is no longer a principled philosophical argument to the effect that some kind of entities (e.g., unobservable entities) are useful fictions and the like. In effect, only persistence in time, i.e., withstanding major revolutions and breaks, can tell us what there is. In this sense, wait-and-see instrumentalism is compatible with type R, viz., realism. Actually, wait-and-see instrumentalism is more or less identical with an extra cautious realism.

12.10 A Coda

So what's the point of the realism-instrumentalism controversy? Stanford draws on the debate between catastrophists and uniformitarianism in geology and argues that the debate is about large-scale patterns in theory-change in the history of science. Instrumentalists claim that "the central commitments of future theoretical orthodoxy will (or would) ultimately be separated from those of the present by differences as fundamental, profound, far-reaching, and unpredictable as those that separate our own theories from their historical predecessors" (2015b, 415). While realists take it that at least the most central and fundamental claims of our most successful scientific theories as firmly established in ways that may be supplemented or modified but are quite unlikely to be overturned in the course of further inquiry" (2015b, 413).

If *that's* the difference, then there is not much progress made vis-à-vis type I or wait-and-see instrumentalism. Though showing this would require a careful look at the history of theory-change so far, there are two considerations that strongly support it.

The first consideration is that there are no clear-cut criteria about what's more likely to happen in the future, that is: radical wholesale changes or extensive continuity with some piecemeal changes? The historical record doesn't speak with the voice of an angel; so what patterns have prevailed so far is a matter of interpretation. We can certainly say that the history of theory-change is a history

of discontinuities and continuities. Overall, I would say that there is a pattern of retention; and in particular a pattern in which more and more unobservable entities are accepted as real (see my 2021). Others might disagree. But that's precisely the point. In more or less any future time t, the situation will be the same as it is now: the past will be a history of discontinuities and continuities. Faced with *this* history, future realists and instrumentalists will still argue about the prevailing patterns. This would give an air of futility to Stanford's suggestion about the point of the realism-instrumentalism controversy.

The second consideration is that history is made by human subjects – in this case, scientists. So whether or not changes in the future science will be *fundamental*, *profound*, *far-reaching*, and *unpredictable* depends on how science will be practised in the future. If scientists use past experience as a guide, they can learn to do better science, e.g., to look for more and varied evidence, to formulate and test alternatives etc. And if this the case, we have every reason to believe that as science grows the changes will be less and less *fundamental*, *profound*, *far-reaching*, and *unpredictable*.

Where does all this leave us? Compatibilism fails: instrumentalism and realism are rival views of science; the differences are deep and interesting. And yet, insofar as instrumentalism is a cautious wait-and-see attitude towards current science, emphasising the discontinuities, the raptures, in theory-change, insofar as instrumentalism is captured by type I above, it becomes a view that scientific realists can coexist with.

Acknowledgements Work which led to this paper has been presented in the Philosophy of Science reading group in the University of Athens. Many thanks to the Usual Suspects for comments; and especially to Maria Panagiotatou for extensive discussions of Howard Stein's paper 'Yes, But. . .'. Thanks are also due to Wenceslao J. Gonzalez and an anonymous reader for encouragement. I want to dedicate this paper to the memory of Richard Boyd who taught me how to be a realist.

References

Carnap, R. (1950). Empiricism, semantics and ontology. *Revue Intérnationale de Philosophie, 4*, 20–40.

Carnap, R. (1963). Replies and systematic expositions. In P. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 859–1013). Open Court.

Carnap, R. (1966). *Philosophical foundations of physics*. Basic Books.

Carnap, R. (1974). *An introduction to the philosophy of science*. Dover Publications.

Copernicus, Nicolaus. ([1543] 1992). On the revolutions (Translation and Commentary by Edward Rosen). The Johns Hopkins University Press, 1992.

Craig, W. (1956). Replacements of auxiliary assumptions. *Philosophical Review, 65*, 38–55.

Dewey, J. (1925). *Experience and nature*. Open Court Publishing Company/Dover Publications 1958.

Dewey, J. (1939). Experience, knowledge and value: A rejoinder. In P. A. Schilpp & L. E. Hahn (Eds.), *The philosophy of John Dewey* (pp. 517–608). Open Court.

Duhem, P. (1908). *ΣΩΖΕΙΝ ΤΑ ΦΑΙΝΟΜΕΝΑ. Essai sur la notion de théorie physique de Platon à Galilée*. Paris: Librairie scientifique A. Hermann. Translation: *To Save the Phenomena: An essay on the idea of physical theory from Plato to Galileo*. The University of Chicago Press, 1969.

Earman, J. (1978). Fairy tales vs an ongoing story: Ramsey's neglected argument for scientific realism. *Philosophical Studies, 33*, 195–202.

Frank, P. (1932). *The law of causality and its limits* (M. Neurath & R. S. Cohen, Trans.). Kluwer.

Goodman, N. (1957). Review: William Craig, replacement of auxiliary expressions. *Journal of Symbolic Logic, 22*, 317–318.

Lambert, K., & Brittan, G. G. (1987). *An introduction to the philosophy of science* (3rd ed.). Ridgeview Publishing Company.

Lloyd, G. E. R. (1978). Saving the appearances. *Classical Quarterly, 28*, 202–217.

Nagel, E. (1961). *The structure of science*. Harcourt, Brace & World.

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

Psillos, S. (2007). Carnap and incommensurability. *Philosophical Inquiry, 30*, 135–156.

Psillos, S. (2009). *Knowing the structure of nature*. Palgrave-MacMillan.

Psillos, S. (2012). One cannot be just a little bit realist: Putnam and van Fraassen. In J. R. Brown (Ed.), *Philosophy of science: The key thinkers* (pp. 183–206). Continuum Books.

Psillos, S. (2020). Resisting scientific anti-realism: Review of K. Brad Wray. Resisting scientific realism. *Metascience, 29*, 17–24.

Psillos, S. (2021). From the evidence of history to the history of evidence: Descartes, Newton and beyond. In T. Lyons & P. Vickers (Eds.), *Contemporary scientific realism and the challenge from the history of science* (pp. 70–98). Oxford University Press.

Psillos, S., & Zorzato, L. (2020). Against cognitive instrumentalism. *International Studies in Philosophy of Science, 33*, 247–257.

Putnam, H. (1963). "Degree of confirmation" and inductive logic. In P. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 761–783). Open Court.

Putnam, H. (1965). Craig's theorem. *Journal of Philosophy, 62*, 251–260.

Putnam, H. (1971). *Philosophy of Logic*. G Allen & Unwin Ltd.

Rowbottom, D. P. (2019). *The instrument of science: Scientific anti-realism revitalised*. Routledge.

Rowbottom, D. P. (2018). In J. Saatsi (Ed.), *The Routledge handbook of scientific realism* (pp. 84–95). Routledge.

Sambursky, S. (1962). *The physical world of late antiquity*. Routledge and Kegan Paul.

Scheffler, I. (1963). *The anatomy of inquiry*. Bobbs-Merrill.

Schurz, G. (2014). *Philosophy of science: A unified approach*. Routledge.

Sellars, W. (1965). Scientific realism or irenic instrumentalism: A critique of Nagel and Feyerabend on theoretical explanation. In R. C. M. Wartofsky (Ed.), *Boston studies in the philosophy of science* (Vol. II, pp. 171–204). Humanities Press.

Simplicius. (2014a) On Aristotle Physics 2 (Trans. Barrie Fleet). Bloomsbury.

Simplicius. (2014b). On Aristotle On the heavens 3.7–4.6 (I. Mueller, Trans.). Bloomsbury.

Stanford, P. K. (2015a). Catastrophism, uniformitarianism, and a scientific realism debate that makes a difference. *Philosophy of Science, 82*, 867–878.

Stanford, P. K. (2015b). "Atoms exist" is probably true, and other facts that should not comfort scientific realists. *Journal of Philosophy, 112*, 397–416.

Stein, H. (1989). Yes, but ... some skeptical remarks on realism and anti-realism. *Dialectica, 43*, 47–65.

Vaihinger, H. 1911. *The philosophy of 'As if'* (C. K. Ogden, Trans.). Kegan Paul, Trench, Trubner and Co Ltd (1935)

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

Wray, K. B. (2018). *Resisting scientific realism*. Cambridge University Press.