REALISM VERSUS ANTI-REALISM: PHILOSOPHICAL PROBLEM OR SCIENTIFIC CONCERN?

Jesús Zamora-Bonilla
UNED (Madrid)
jpzb@fsof.uned.es

Forthcoming in *Synthese*

ABSTRACT

The decision whether to have a realist or an anti-realist attitude towards scientific hypotheses is interpreted in this paper as a choice that scientists themselves have to face in their work as scientists, rather than as a 'philosophical' problem. Scientists' choices between realism and instrumentalism (or other types of anti-realism) are interpreted in this paper with the help of two different conceptual tools: a deflationary semantics grounded in the inferentialist approach to linguistic practices developed by some authors (e.g., Sellars, Brandom), and an epistemic utility function that tries to represent the cognitive preferences of scientists. The first tool is applied to two different questions traditionally related to the problem of scientific realism: the non-miracle argument, and the continuity of reference. The second one is applied to the problem of unconceived alternatives, and to the distinction between realist and instrumentalist attitudes towards scientific hypotheses.

KEYWORDS

Scientific realism; anti-realism; instrumentalism; deflationism; pragmatics; inferentialism; non-miracle argument; explanation; predictive success; continuity of reference; theoretical concepts; anaphora; epistemic utility function; verisimilitude; coherence; falsificationism; unconceived alternatives.

To appear in *Synthese*
Published on-line, December 21th, 2015

most scientists are instrumentalists on Sunday and scientific realists the rest of the week
The Epicurean Dealmaker, "Our glassy essence"
http://epicureandealmaker.blogspot.com.es/2013/10/our-glassy-essence.html

1. Introduction: '∃ means ∃'.

The main thesis defended in this paper is that the choice between a realist or an anti-realist attitude towards scientific theories, concepts, etc., is a choice that scientists themselves have to face in their work, and whose rationality can be assessed 'from the inside' of science. I am not denying that there is something like 'the philosophical problem of scientific realism', which in very general terms could be described as the problem of whether the claims of science have to be 'interpreted' literally or non-literally from the point of view of some philosophical semantics or epistemology, but instead want to point that many important aspects traditionally associated to those 'philosophical' questions can profitably be understood as *scientific* problems, i.e., questions that scientists themselves have to consider and to answer within their own practice *as scientists*. After all, on reflection it is a little bizarre to assume that, as Arthur Fine famously described the debate between realism, anti-realism, and his own 'natural ontological attitude', we can make a neat cut between 'belief' and 'interpretation';[1] for, though we can imagine cases in which one believes something, or believes she believes it (say, the dogma of Trinity), without really knowing what it means, it seems that most of the time people have a more or less perspicacious understanding of what they believe and say; so, to have a belief is *already* to have an interpretation of that belief, it is to *understand* what one is believing. Of course we might say that this is usually not a 'philosophical' interpretation, even in the case of the scientists' beliefs and claims, but in most cases, and particularly in science, what these beliefs and claims are about will be *precisely* things such as what things are there ('ontology') and how can we manage to know them ('epistemology'). Furthermore, if it were true that scientific claims needed a

[1] "Realism requires two distinct elements. It requires belief and it also requires a particular interpretation of that belief", Fine (1986, 176).

‘philosophical interpretation’ in order for ‘us’ to determine whether ‘we’ are rationally justified to take those claims (e.g., that there are such and such molecules with such and such structure) as literally true, we could go one step further and ask whether ‘we’ must take the philosopher’s claims literally (say, perhaps an anti-realist philosopher doesn’t aim that we take him literally when he uses the philosophically-theoretical concept of ‘unobservable’); but this, obviously, would immediately lead us to an infinite regress. Lastly, looked from the perspective of scientific practice, it is not even clear what is meant in the arguments for and against realism when the expression ‘scientific theories’ or ‘scientific claims’ are used; for, after all, some scientists, especially during controversies, affirm some things while other affirm the opposite, some say that there are enough arguments to accept a hypothesis while other scientists deny it. For example, the history of science is full of cases where researchers have *rejected* the existence of some entities or the truth of some hypotheses: are we going as philosophers not to take *literally* what scientists are saying in these cases?

The main point of this paper, hence, is to consider the decision between having a realist or an anti-realist attitude towards a particular scientific hypothesis, model, concept, etc., not as a philosophical inquiry (without denying that there are deeply interesting things within *that* inquiry) , but as a decision that scientists themselves ordinarily face in their practice. I will not, however, engage here in a historical or ethnographic investigation about what scientists actually do, but I will try to develop a *philosophical* (or meta-scientific) explanation of their predicament, i.e., I will try to illuminate it from the point of view of several philosophical tools. This does not contradict what I said in the previous paragraph, because my goal is not to discuss whether the assertions of scientists are to be taken literally or not, but to understand why, e.g., they make some times affirmative existential assertions and other times they do the opposite, why they are concerned sometimes about the truth of some theories or models, while other times they are not.

In particular, I shall employ two different approaches. In section 2, I will show how a couple of traditional arguments for scientific realism (the ‘non-miracle argument’ and the ‘continuity of reference’) look like when considered from the point of view of a deflationary, pragmatically oriented semantics (i.e, one in which the primary role of semantics is mainly understood as that of explicating the use of semantic terms, like 'true' or 'refers', within natural languages). In section 3, I will employ a rational-choice model of scientists' behaviour, and in particular the assumption that their choices are guided by a certain type of 'epistemic utility' (one that in other papers I have called 'empirical verisimilitude'), in order to develop a tentative explanation of scientists' choice between 'realist' and 'instrumentalist' ways of evaluating theories, and in particular to the question of how relevant it is for that evaluation the possible existence of 'unconceived alternatives'. Section 2 will try to bring down, so to say, some aspects of the traditional philosophical debate about realism to the field of the scientists' own decisions about what to assert and what to deny, while section 3 will offer a more substantial account about what may justify those decisions from the point of view of scientists themselves.

Of course, though I will put some simple examples of my arguments, a detailed historical study of 'realism-and-anti-realism-in-the-practice-of-science' under the lines of the philosophical ideas presented in this paper would be necessary, but, being myself not a historian, I will have to leave that work for other scholars. I'd be content if the ideas drafted in this paper serve as an inspiration for such type of work.

2. Scientific realism from a deflationary point of view.

2.1 Semantic deflationism.

A theory about semantics is deflationist if it claims that semantic notions can be sufficiently explained without committing ourselves to some or other position about metaphysical problems. For example, Paul Horwich's minimalist theory of truth is based on the trivial (Aristotelian, Traskian) point that proposition 'X' is true iff X, but adds that this is *all* that is needed to explain the role of the truth predicate.[2] Another interesting deflationary approach is Robert Brandom's expressivism, which explains the role of semantic terms like 'true' or 'refers' by the *expressive* capacity they give to those languages possessing them, i.e., by the things we can say thanks to those expressions and that we could not say without them (e.g., "all logical consequences of

[2] Horwich (1990).

true axioms are true").[3] I will not commit, however, to any detailed version of deflationary semantics, but will limit myself to using some simple ideas derivable from their more trivial claims; basically, in the first place, the idea that, for every proposition 'X', there can be no difference at all between a world of which we can assert that X, and a world of which we can assert that 'X' is true, and second, the idea that the most basic role of the notion of reference is that of establishing co-reference (or 'anaphora'), i.e., allowing us to determine when can one expression be substituted for another.

2.2. What are we explaining when we explain the success of science?

In the last decades, the most popular philosophical defence of scientific realism has been what is known as the 'no miracle argument' (NMA).[4] Though there is a range of different interpretations of NMA,[5] it typically asserts that scientific realism is 'the best explanation' (or perhaps, the only reasonable one) of the 'success of science', or more particularly, of the empirical success of modern scientific theories. I admit that the argument is intuitively compelling, and my own discussion will probably not contradict it strictly speaking, but I will try to show that, when considered from the point of view of a deflationary semantics, NMA transforms itself in something close to a trivial *scientific* claim.

Let's start by considering what is that NMA tries to be an explanation *for*. In the most compelling cases, what it tries to explain is not the 'general success' of modern science (i.e., how it is that *we* have managed to develop such a successful science), but, more specifically, the tremendous empirical success of *some* theories, especially in the natural sciences, and more particularly in physics. Defenders of NMA claim that it would be almost impossible that those theories made so many and so good predictions if they were not true, or at least, very approximately true. But, what are we exactly explaining when we explain 'the tremendous empirical success of a theory'? What kind of *fact* our *explanandum* is?

[3] Brandom (1994), ch. 5.

[4] Putnam (1981).

[5] See, e.g., Psillos (1999, ch. 10), Worrall (2007), and Frost-Arnold (2010).

Let T be the theory whose empirical success we want to explain, and let E be the proposition (or conjunction of propositions) that constitute the empirical evidence on which T is assessed. For simplicity, let's suppose that T explains and predicts E perfectly, i.e., that T logically entails E, and that E was deeply implausible before being derived from T and empirically tested.[6] The *fact* that T is 'predictively very successful' seems to consists in the *conjunction* of the following propositions (each of them logically independent of the rest):

(1) T entails E
(2) The truth of E was surprising at the moment E was derived from T
(3) E is true.

Hence, in order to explain the 'success' of T, we shall have to offer an explanation of these three facts. However, once we enter into the details, it's easy to see that (1) is just a 'logical fact': there is no difference between explaining 'why' T entails E and just *proving* that T entails E; there is nothing like a 'substantive' explanation in explaining (1), i.e., an explanation that has to do with how the world is, or anything we can conceive as related to 'the problem of scientific realism'; and, of course, it is not 'the truth of T' what explains the (logico-mathematical) fact that T entails E, for this is independent of whether T or E are true or false.

Regarding (2), it can in turn be decomposed into the following claims:

(2.a) E was deduced from T before knowing whether E was true
(2.b) The prior probability of E before testing it empirically was very low[7]

[6] Of course, there can be lots of discussions about what is to be a 'good explanation' and its connection to predictions, but they will distract us from the specific point I want to make, so I shall use in my argument the most naïve version of the nomologico-deductive schema.

[7] I employ this 'Bayesian' reconstruction of the idea of 'surprisingness', but each reader can replace it for his or her favourite one.

As in the case of (1), the fact (2.a) seems to have nothing to do with a substantive explanation in the sense that would be necessary for assessing the NMA; at most, it is a psychological or historical fact about the specific people who carried out the deduction of E from T, and about the historical evolution of the mathematical or logical technics that allowed to perform it. Its explanation seems also not to have anything to do with whether T is true, which is the *explanans* favorited by defenders of the NMA. On its turn, (2.b) simply asserts that E was (very) implausible at the time; we can explain (2.b) either by offering a kind of formal description of the amount of information E contains, or by giving details about why the contemporaries considered it implausible on the basis of their other beliefs, but at no place it seems that 'the truth of T' can serve as an explanation of (2.b), nor of the other facts contained in (2) or (1).

Lastly, we have the fact (3), i.e., the fact that E is true, as the only element in the *explanandum* of NMA for which 'the truth of T' might serve as an *explanans*. Here is where our semantic deflationism reminds us that 'the fact that E is true' is just another way of *expressing exactly the same fact* that proposition 'E' expresses. Hence, explaining (3) *is just the same thing* as explaining E. But *for explaining E we don't need any philosophical theory; what we need* (and scientists pursue) *is a scientific theory like T*, one of whose goals is, obviously, to explain E.

We find out something analogous when we reflect on the usual way realist philosophers express, not the *explanandum* of NMA, but its purported *explanas*: it is, they say, the *truth* of T, or the fact that T is true (or approximately so), what 'explains' its empirical or predictive success. But from our deflationary semantics, there is, again, *no difference at all* between 'the fact that T is true' and those facts about the world that we express in affirming T itself. Hence, 'that T is true' is just a different way of saying *that* T (when T is not taken as the *name* of the theory, but as the conjunction of its axioms, principles or hypotheses). Hence, *by saying that 'the truth of T explains its empirical success', or that 'the fact that T is true explains why T is empirically successful', we are just expressing in a slightly more complicated way what we can express just by saying that T explains E*, for 'T' asserts exactly the same as 'the fact that T is true', and 'E' asserts basically the same as 'the empirical predictions or T are true' (if, as we are assuming, they are).

In a nutshell, I have decomposed the *explanadum* of NMA into three different, logically independent facts. Of these facts, (1) requires not an 'explanation' at all, but a logico-mathematical proof, which is what scientists do when *they* derive E from T. (2) may require an explanation, but it will be a historical one (why some people did some things at some time), which is, in a sense, a scientific explanation, not a philosophical one. Furthermore, neither (1) nor (2) can be explained in any way by 'the truth of T'. And lastly, (3) needs indeed an explanation, but it is also a *scientific* explanation, the one for which T has been devised.

I want to insist in that I am not claiming that there is nothing like 'explaining the empirical success of a theory'. What I am saying is that, when we carefully analyse the *explanandum* of that explanation, we find that the only substantive part of it is just the one the theory itself explained, if it was empirically successful. Hence, in order to assess whether 'the truth of T' is a good explanation of 'the empirical success of T', the *only* thing we can do is to see whether T is a good, or appropriate, or acceptable explanation of E. But *this* is a scientific problem, i.e., a problem *for scientists* to solve, not a 'philosophical' one.[8] I am also *not* denying that there is something important to the intuition that 'it is very unlikely that T is not (at least approximately) true, given how empirically successful it is', nor to the idea that novel predictions are a better reason to accept the truth of the theory than *ex post* accommodations of already known facts; I will even say something else about this in the section 3. But I invite realists to identify within the *explananda* of NMA some elements that I might have missed in my own analysis, and to show how 'the truth of T' is actually a good explanation of that, in some way that goes further from the mere logical fact that T explains or predicts what it explains or predicts.[9]

[8] Of course, philosophers can discuss about what an explanation consists in, what is its relation to prediction, etc., etc. But this is (at best) an additional clarification of the work of scientists, whereas 'explaining the success of T' is, if my analysis is right, exactly *that* work, not a clarification *of* it.

[9] Actually, it seems even clear that the best explanations of the empirical success of some past scientific theories do not often come from philosophical arguments, but from the scientific theories that supersede the former ones. For example, Newton's gravitation theory offered a sensational explanation of why Kepler's theory of planetary movements was so accurate (though not perfectly so), like Einstein's

Note, finally, that my argument is not an attack to scientific realism as a *substantive* thesis; rather on the contrary, if T is empirically adequate and many of its predictions were highly implausible, this is of course a reason to accept the truth of T (what *other* better reason could we have, after all?). What I am saying does not go against *this* claim, only against the idea that there is something in the 'philosophical interpretation' of the relation between T, the observed facts, and 'the world', that goes beyond what scientists themselves need take care of.

2.3. The continuity of reference.

Another common topic in the realism debate is that about the reality of 'theoretical entities', and in particular, about the 'continuity' or 'discontinuity' of the reference of theoretical terms as scientific theories change and are replaced by others during the course of history. Typically, anti-realists have tried to show that many, if not most theoretical entities hypothesised by overcome scientific theories did not exist, according to our current knowledge, even if they were necessary to generate those theories' successful empirical predictions.[10] A common line of defence of realist philosophers has been to indicate that, even if past theories have been refuted and their theoretical terms have been shown to be non-denoting, or their meanings have considerably changed, there is nonetheless a strong *continuity* between the ontologies of the successive theories.[11] What is understood by this continuity is that something important is 'preserved' in the process of passing from one theory to another. Structural realists affirm that what is preserved is 'structure': things like equations, symmetries, or any other type of *abstract* forms. Entity realists, on the other hand, affirm that besides the conservation of some structural features, also the *reference* or the denotations of some theoretical terms are often preserved to some extent.

---

relativity explained in a superb way why Newtonian mechanics is so empirically adequate in describing the movements of not-too-fast and not-too-heavy bodies. Again, this shows that you need scientific arguments to explain the success of science, rather than philosophical ones

[10] This is the famous Laudan's 'pessimistic meta-induction'; cf. Laudan (1981).

[11] See, e.g., Psillos (1999, ch. 12) for a general discussion, and also Worrall (1989), Ladyman (2007),Bueno (2008) and Bartels (2010) for some different versions of the argument.

Moreover, it may be argued that many ‘overtaken’ scientific concepts can be reinterpreted as more or less erroneous (and hence, more or less right) descriptions of entities that we take as really existing.[12] For example, phlogiston has been interpreted as a kind of rudimentary account of electrons, and the caloric fluid has been taken as a naïve description of molecular heat. I will not enter here into historical discussions, but will instead approach the problem in a purely abstract way from the point of view of our semantic minimalism. This suggests that, instead of considering the question of whether theoretical terms ‘really’ refer or not, we should consider first of all what is what the speakers (in this case, scientists) want to express *by choosing one way of speaking or another* about those theoretical entities.

Imagine the following simplified, ‘Sesame Street’ situation. We have two scientists, Ernie and Bert, that employ several predicates P, R and S, with which they can make different assertions. Let ‘$C_i(X)$’ stand for ‘scientist *i* claims that *X*’ (for the sake of simplicity, I will not make any difference between ‘claiming’, ‘knowing’ or ‘believing’), and let ‘#xFx’ stand for the definite description ‘the x such that Fx’. Suppose that Ernie and Bert assert the following:

(4) $C_e(\ \exists x(Px \ \&\ Rx \ \&\ \neg Sx))$

(5) $C_b(\ \exists x(Px \ \&\ \neg Rx \ \&\ Sx))$

That is, Ernie affirms that there is something that has properties P and R, but not S, whereas Bert affirms that there is something that has properties P and S, but not R. The interesting question is, of course, whether those ‘things’ (the x’s) are ‘the same’ or not. From Ernie’s point of view there are two possibilities:

(6) $C_e[\exists x(Px \ \&\ Rx \ \&\ \neg Sx) \ \&\ C_b(\exists y(Py \ \&\ \neg Ry \ \&\ Sy)) \ \&$
$\&\ \#xC_e(Px \ \&\ Rx \ \&\ \neg Sx) = \#yC_b(Py \ \&\ \neg Ry \ \&\ Sy)]$

(7) $C_e[\exists x(Px \ \&\ Rx \ \&\ \neg Sx) \ \&\ C_b(\exists y(Py \ \&\ \neg Ry \ \&\ Sy)) \ \&$
$\&\ \#xC_e(Px \ \&\ Rx \ \&\ \neg Sx) \neq \#yC_b(Py \ \&\ \neg Ry \ \&\ Sy)]$

[12] Of course, anti-realist defend just the contrary; cf. Chang (2012).

Both (6) and (7) coincide in affirming that Ernie believes the same as in (4), and furthermore, that he believes what (5) affirms about Bert. The difference between (6) and (7) is that, according to (6), Ernie thinks that the entity of which Bert believes what is expressed in (5) is the *same* entity of which Ernie believes what is expressed in (4), whereas, of course, in the case of (7) Ernie thinks that these are two *different* entities.

One simpler way of expressing the same as in (6) would be

(8) $C_e[\exists x(Px \,\&\, Rx \,\&\, \neg Sx \,\&\, C_b(Px \,\&\, \neg Rx \,\&\, Sx))]$

That is, Ernie thinks there is something that has properties P, R, and not S, and *of which* Bert believes that has properties P and S, but not R.[13] Co-reference of the 'x' and the 'y' of (6) is directly and 'transparently' expressed in (8) as the fact that now there is only one existential quantifier and one individual variable.[14] However, (7) does not admit such a simplification, but at most the following one

(9) $C_e[\exists x\{(Px \,\&\, Rx \,\&\, \neg Sx) \,\&\, C_b(\exists y(Py \,\&\, \neg Ry \,\&\, Sy)) \,\&\, x \neq \#y\, C_b(Py \,\&\, \neg Ry \,\&\, Sy)\}]$

The question is, nonetheless, that Ernie will have to choose between (8) and (9): does he believe that the entities of which Bert claims such and such are *the same* entities of which he (Ernie) claims different things, or not? Lets apply our schemata to a couple of examples. If P, R and S stand for 'is the minimal unit of a chemical element', 'is decomposable into smaller particles', and 'is perfectly spheric', both (8) and (9) would assert that Ernie (say, a contemporary scientist) believes that there are minimal units of chemical elements, that these units are decomposable into smaller particles, but that they are not perfectly spherical, whereas Ernie knows that Bert (say, a mid 19th-century chemist) thinks that there are indeed minimal units of chemical

[13] This (very simplified) analysis of the semantics of referential terms is inspired in Brandom (1994), ch. 5.

[14] Technically, (6) logically follows from (8), but not viceversa, but this is because of some technical reasons about doxastic logic that are not relevant for my argument.

elements, but that they are perfectly spherical though not decomposable into smaller units. The difference between (8) and (9) would be that, in the case of (8), Ernie thinks that the entities ('atoms') accepted by Bert *are the same* entities accepted by Ernie, whereas in (9), of course, Ernie thinks that the entities Bert accepts *are not* the same entities accepted by Ernie. Probably, most contemporary scientists would opt for an interpretation like that suggested by (8): the atoms of the atomists of the 19th century *are* the atoms 'we' (contemporary scientists) talk about, though those chemists *attributed* to these atoms some different properties from those we attribute to them.

But let's consider a different example, in which P stands for 'is an infectious agent causing Bovine Spongiform Encephalopathy (BSE)', R stands for 'is a virus', and S stands for 'is a prion'. In this case, according to (8) and (9), Ernie would think that what causes BSE is a virus, not a prion, whereas he knows that Bert thinks that what causes BSE is a prion, not a virus. Additionally, (8) asserts that Ernie thinks that the entities Bert takes as causing BSE are the same ones that Ernie takes as causing BSE, whereas according to (9), Ernie thinks they are not the same entities. In this case, it seems more natural to interpret the disagreement between Ernie and Bert through a proposition like (9): Ernie thinks that the prions imagined by Bert *just do not exist*, for, 'of course', prions cannot be 'identical' to viruses.

But, what is the difference between both historical cases? It cannot be a *formal, structural* difference, because our two examples are totally analogous from a formal point of view; you can add more detailed and complex predicates to formulae (4)-(9), but the point would still be the same: there is nothing in the logical structure of propositions (4) to (9) allowing to prove that it is justified to accept some times that you are talking about the same, and other times that you are not. However, it seems that it is not also a difference having to do with the '*causal* capacities' of the entities or systems under study, or with our capacity of '*manipulating'* those entities or systems: after all, there are procedures to isolate in a flask the infectious agents of BSE, procedures that are notoriously 'causal', and Ernie and Bert may agree that a particular flask contains such entities… only that each of them thinks that the entities the other believes that are contained in the flask do not really exist (e.g., there is not 'a BSE virus').[15]

I think there is simply no way to solve the dilemma of choosing between (8) or (9) by applying a logical or philosophical theory of 'reference' (though, of course, defenders of philosophical realism are invited to provide one), nor, by the way, by any philosophical theory of scientific confirmation that only takes into account the logical structure of the propositions involved. This, after all, is just a consequence of the 'ontological relativity' and 'inscrutability of reference' suggested by Quine half a century ago, which is particularly conspicuous in the type of second-order intensional contexts we are now examining (i.e., when one thinks about what other people think). But the undeniable fact is that real scientists *do* choose something like (8) in *some* cases and something like (9) in some *other* cases, and often in a pretty spontaneous way, without feeling the need of having a philosophical theory to guide their choice. My suspicion (which I will not explore here in detail) is that the decision about how much historically continuous the reference of scientific terms in a particular context is depends more on *pragmatic* reasons than on *semantic* ones. The belief of 'being talking about the same' is obviously necessary in any process of research, including conversation and controversy, no less than in the ordinary use of language: if you and I are going to rationally debate about something, I will have to assume that *some* of the entities you are talking about *are* entities I am also talking about (e.g., we both accept we are talking about the same flask). Scientific consensus will tend to enlarge the set of 'shared' entities, both as a consequence of the working of other consensus forming strategies, and as an efficient strategy to promote intersubjective discussion: in cases where there are strong controversies, it is wise to make a previous effort to determine 'what are we talking about'. But these strategies will work differently in different contexts, and they will also manifest a certain degree of hysteresis and contingency: once certain criteria or methods of 'identifying' entities have been established within a field, their application will lead to accepting some entities that might have not been

[15] Furthermore, those 'causal capacities' would be after all nothing but additional predicates to be included in formulas (4)-(9), and so *logic* itself would be as powerless to allow us to derive any relevant difference from causal predicates as it is regarding any other type of predicates.

accepted if different methods had been established.[16] On the other hand, competition will in some cases lead some researchers to emphasize the *differences* between their concepts and those of their rivals, more than the similarities. All these reasons may explain why we have cases in which scientists end accepting (8) and cases in which they accept (9), in spite of there being *no substantial* difference from a formal, empirical, or causal point of view between both cases. This does not mean that 'ontological continuity' is not an important issue in the history of science. It only means that it is a problem that we must not think we can solve thanks to a philosophical theory, or at least, thanks to an aprioristic metaphysics, semantics or epistemology. The most similar thing to a 'philosophical' explanation of why certain continuities are *accepted* in the history of science and why others are not, could probably be something like an applied theory in philosophical pragmatics, but I propose this only as a suggestion for further research. The most important conclusion of this subsection for my general argument is that the 'continuity of reference' is something that scientists themselves have to decide (and *do* decide) just in choosing one way of talking or another when comparing some theories to others, and for doing that they don't need philosophical arguments, just the common linguistic or argumentative strategies that might have been developed within their disciplines.[17]

---

[16] See again Chan (2012) for a nice historical illustration.

[17] One referee has pointed out that the continuity of reference is not so much something that philosophical realism has to explain, as something that, if true, will count as an argument in favour of realism. I do not deny this, though I doubt that it reflects a unanimous philosophical view about the connection between the problem of continuity and the problem of realism. But, assuming the referee's point, it does not count against my argument, for I'm not criticising *realism*, but the idea that the arguments in favour of realism are fundamentally 'philosophical', rather than scientific. So, if the continuity of reference is a datum supporting realism, my point is that it will be scientists, nor philosophers, the ones that have to decide whether continuity exists and in what cases.

By the way, an analysis of scientists' choice between (8) and (9) under the pragmatic approach suggested here would be in line with the view of 'rhetoric as the strategic use of language' I have

3. Realism and the aims of scientists.

3.1. The scientist's utility function.

Having a realist or an instrumentalist *attitude* towards a particular scientific theory, model, hypothesis, etc., (i.e., accepting whether some theory is true, whether some entities exist...) is a matter of concern for flesh and bone scientists, as we have seen in the previous subsection. In order to understand how scientists may *decide* to do one thing or the other, we need to have some insight about *their* preferences, or, as economists and rational choice theorists say, about their 'utility functions'. Instead of trying to derive some intuition about what the epistemic preferences 'should' be, I will take for granted that scientists are society's best experts in knowledge (latin: *scientia*), as cyclists are the best experts in cycling, or fishermen in fishing, and I will try to develop a very simplified model of why scientists do what they do in their pursuit of knowledge, in an analogous way as how economists produce simplified models that try to understand why economic agents do what they do, and what consequences follow from that, without assuming that they, the economists, 'know better' that producers or customers whether, say, one car is better than another. Hence, the 'utility functions' I am going to present here are just to be taken as a *hypothetical, empirical, and simplified* reconstruction of the *real* preferences scientists have on epistemic matters.[18]

My suggested definition of the concept of *empirical verisimilitude of a hypothesis or theory T on the light of evidence E* consists in the combination of a high degree of coherence or similarity between t and E, and a high degree of coherence or

advanced in Zamora Bonilla (2006), though applied to a totally different case (there, the topic was the choice of one interpretation or another of an experimental result).

[18] Elsewhere(e.g., Zamora Bonilla, 2002a) I have used the hypotheses that the scientists' utility functions combine two elements, an epistemic and a social one, which I identify basically with recognition from their colleagues. This 'social' part will be irrelevant for the rest of my argument here.

similarity between E and the whole truth (W), i.e., how similar T resembles to what we empirically know about the truth, weighted by how similar this part of the truth seems to the whole truth (this weight will be shown below to be equivalent to how informative E seems to be). This notion of coherence between two propositions X and Y is modelled as p(X&Y)/p(XvY),[19] where *p* is assumed to be a typical subjective (prior) probability function, which can be different for different scientists. Hence,

(10) $Vs(T,E) = [p(T\&E)/p(TvE)]\ [p(E\&W)/p(EvW)]$

$= [p(T\&E)/p(TvE)]\ [p(W)/p(E)]$ (since E is assumed to be true)

$= [p(T\&E)/p(E)]\ p(TvE)p(W)$ (re-arranging)

$\propto p(T,E)/p(TvE)$ (since p(W) is a constant)[20]

For simplicity, it will be the latter expression what I will take as the definition of empirical verisimilitude. In the particular case when T correctly explains or predicts E (i.e., if T entails E), this leads to:

(11) $Vs(T,E) = p(T,E)/p(E)$ (since T entails E)

$= [p(T\&E)/p(E)]/p(E)$

$= p(T)/p(E)^2$ (since T entails E)

However, in the case when T is fully confirmed by the evidence E (i.e., if E entails T), the value is:

[19] I presented this definition in Zamora Bonilla (1996). For more details, see Zamora Bonilla (2013). (independently, the formula *p(X&Y)/p(XvY)* became later a standard definition of coherence, after Olsson (2002); similar results could be proved with other definitions of coherence or similarity, but I think the combination of power and simplicity of that one justifies its choice). My definition is close in spirit to other approaches to verisimilitude (the notion that different theories can be farther or closer to the truth; see Niiniluoto 1987), but combines in a single measure of 'how closer to the whole truth a theory seems to be' what in other approaches, like Ilkka Niiniluoto's, is decomposed into a notion of 'objective distance to the truth' and a notion of 'estimated distance'. My own point of view is that any relevant notion of 'similarity' contains always some subjective factors, which in my definition are partly taken into account by the fact that *p* represents a subjective probability function.

[20] For simplicity, the symbol '$\propto$' will be replaced by '=' in the reminder of the paper.

$$(12)\ Vs(T,E) = [p(T\&E)/p(TvE)]/p(E)$$
$$= [p(E)/p(T)p(E)]$$
$$= 1/p(T)$$

In the case when the empirical evidence contradicts the theory (and hence Vs(T,E) = 0), there are however some ways of using Vs to represent the epistemic preferences of scientists. For example, the epistemic value of H can be given by Vs(T,E(T)), where E(T) is the conjunction of established empirical laws explained by T, or the conjunction of the most approximate version of each empirical law such that t still entails it.

3.2. Unconceived alternatives.

According to our model, scientists intuitively assess their hypotheses on the basis of a function like *Vs*. Since usually the theories are attempted to explain the relevant empirical data, rather than being logically confirmed by then,[21] formula (11) will be more useful. Let's now consider by means of it the typical case in the discussion about the problem of *unconceived alternatives*:[22] although we may have one or several theories correctly explaining or predicting the known empirical facts, it can still be the case that these theories are false, and not only false, but very 'far' from the truth. We can derive from (11) some interesting lessons in connection to this problem. In the first place, the formula entails that, of several theories rightly explaining E, the one with the highest prior probability (which in this case is equivalent to having a higher posterior probability) will be most valuable; so, a theory correctly explaining the facts will be better for a scientist, the more plausible a priori it seems to her. Note that it is the

[21] Actually, as John Norton (2014) has argued, scientists often *try to prove* their theories from previously known empirical facts or laws (for example, Newton offered a 'demonstration' of the law of gravity taking Kepler's laws as premises), but if this were true at face value, it would entail that theories are not falsifiable (at least, while the empirical laws from which they are mathematically deduced are not rejected), nor can do more predictions than those derivable by the previous empirical laws alone. I will offer below a different interpretation of this kind of 'demonstrative' arguments by scientists.

[22] Stanford (2006).

personal, subjective estimate of the probability of each theory being true what counts in our model; hence, assuming that the scientists working about the problems T is trying to solve are the *best* positioned people to judge those probabilities, there will be nothing that we philosophers may *add* to the probability judgments scientists make about this question. Suppose that T is the only *conceived* theory successfully explaining E; this means that p(¬T,E) will literally be the probability that, according to the particular scientist whose subjective probability function is given by p, there is some *unconceived* theory which is true. To what extent will this be a problem for the *acceptability* of T from that scientist's point of view? It will depend in part on the relative magnitude of p(T,E) and p(¬T,E). The philosophers that have criticized scientific realism on the basis of the 'pessimistic induction' argument seem to base their criticism on the idea that T is 'very' probably false, but scientists can indeed think that p(¬T,E), though 'high', is not *so* high as to preclude the acceptance of T; this will depend, in any case, on the scientists' judgments, not on the philosophers'. Furthermore, (11) does not say anything in principle about how probable it is for a scientist that t is *approximately* true (i.e., that the 'world' corresponding to proposition W is 'close' to the set of worlds consistent with T): she can accept that p(¬T,E) is 'very' high, but also that, so to say, p('W is close to T',E) is also high. So, there is nothing in the arguments about the pessimistic induction and unconceived alternatives that *forces* scientists to conclude that the posterior probability of an empirically successful theory is 'too low' to make its acceptance irrational, or impeding to take an empirically successful theory as a good approximation to the truth. Rather, what follows from my suggested strategy of deferring to scientists' judgments of probability regarding the theories they are working on, is that if they have accepted T on the basis of empirical evidence E, it is *because they estimate* that the probability of an 'unconceived alternative' totally different from T being the right solution to that scientific question is *not high enough* as to preclude the acceptance of T.

Furthermore, and perhaps more important for our discussion on realism: what formula (11) also allows to see is that, of those theories successfully explaining the available empirical evidence, scientists will tend to prefer those that have a *higher* probability, i.e, a higher probability *of being true*, and hence, *truth* must be taken as one *goal* of scientists, at least in the sense that it is an element of their epistemic preferences (the more probable the truth of a theory seems to them, the more likely it is that they accept it, *ceteris paribus*). Hence, even in cases where scientists are sceptic (and hence, anti-realist) about the literal truth of a particular theory because in that case they judge that p(¬T, E) is 'too high', even if they recognise that T is the best available theory, they can still be *realist* in the sense of considering *that it would be good* to have a theory whose probability of being true were much higher. Anti-realism about *specific theories* is, hence, compatible with realism as a general attitude towards the *aims of science*. I will come back to this point in the last subsection.

Second, (11) also let us see that the epistemic value of T given E does not only depend on T's probability: it also depends on *how unlikely E is*. This means that, given two theories (T and T') and two different bodies of empirical evidence (E and E'), such that T correctly explains E, and T' correctly explains E', it *can* be the case that T is judged *better* on the light of E than T' on the light of E', even if p(T,E) < p(T',E') and p(T) < p(T'). This is reasonable, because in order to explain an increasingly more and more exhaustive body of empirical evidence, we will tend to need theories that are stronger and stronger, and hence, more and more improbable a priori. Hence, if E is very informative, so that p(E) is very low, the verisimilitude of a theory T explaining E can be 'very' high even if both p(T) and p(T,E) are 'very' low.

In the third and last place, our formulas allow to calculate something like a *minimal threshold of acceptability* for theories: the worst 'right solution' we might give to a scientific problem would be to answer it with a tautology, i.e., with a proposition that does not assert absolutely anything about the world.[23] Since any tautology Taut is entailed by any body of empirical evidence E, the verisimilitude of the former will be, according to (12):

(13) Vs(Taut,E) = 1/p(Taut) = 1

[23] I do not deny that logico-mathematical truths play an important role in scientific argumentation and discovery, but I don't think they can be identified with 'theories' in any relevant sense, at least when they are taken in isolation. They are, at most, important elements of theories or research programmes, and above all, rules or principles for mathematical deduction.

Hence, for a theory T that *entails* the evidence E, having a verisimilitude *lower* than 1 will be a sufficient reason to discard it, for a non-answer, like Taut, would be epistemically preferable to T. According to (11), the condition for T having a verisimilitude *higher* than one, if T entails E, is:

(14) $Vs(T,E) > 1$ iff $p(T,E)/p(E) > 1$

iff $p(T,E) > p(E)$

iff $p(T) > p(E)^2$

This result has a nice and straight interpretation: in order to be acceptable, a necessary (but by no means sufficient) condition a theory that successfully explains the empirical evidence must fulfil is that its own *posterior* probability must be higher that the *prior* probability of the evidence; or, stated differently, we must not accept an explanation which is *so unlikely* that, even taking into account the evidence, its truth would be *less* probable than the prior probability of having found that evidence.[24] According to this result, scientists would be interested in employing arguments that can show two things: first, that *the empirical facts their theories manage to explain or predict are very unexpected* (i.e., that p(E) is very low, as we saw in section 2.2); and second, that *their theories are relatively plausible* (i.e., that p(T), and hence p(T,E) is sufficiently high). And of course, each researcher can also try to show the opposite in connection with the theories of her competitors. These judgments of plausibility can take any form, from quantitative estimations of probability, to mathematical 'proofs' of the theories' principles from some empirical laws (plus some more or less 'innocent' assumptions)[25], and also to mere 'rhetorical' or 'philosophical' arguments trying to persuade her colleagues of the likelihood or naturalness of some principles.[26] In any

[24] This is a rewording of the idea that extraordinary claims demand extraordinary evidence.

[25] This would be my suggested interpretation of Norton's claim I have referred to above, according to which scientists often use the empirical evidence to 'prove' their theories, and not only to test them.

[26] Hence, in connection to our discussion in section 2, it's not only false that 'scientific discourse' is ontologically or epistemologically dull, so to say, and is in need of a 'philosophical interpretation' in order to derive from it rational answers about what there is, what is true, and how to know it; the case

case, all these arguments can be interpreted as attempts to establish that it is not very likely that the researcher's preferred theory is false, or very *far* from the truth, even though some 'unconceived alternatives' could still be *closer* to the truth.

3.3. Why to be an instrumentalist?

From our previous discussion, it seems that the scientists' (epistemic) utility function entails that they will *necessarily* have a realist attitude towards the goal of science: after all, what they want to prove (if Vs rightly represents their epistemic preferences) is that the theories they defend are probably *true* or close to the truth, and that the theories of their competitors are false or far from the truth. Is it, hence, always irrational for them to have an anti-realist or instrumentalist attitude? And by the way, what would 'an instrumentalist attitude' consist in? A plausible answer to the latter question is that, if we have identified 'a realist attitude' with being concerned for the (probability of the) truth (or approximate truth) of the competing theories, then having an instrumentalist attitude would amount to *not being concerned* by that; i.e., one would reveal an instrumentalist attitude towards a theory, model or hypothesis if one is willing to give it a *high* epistemic value because it explains or predicts well the available evidence, *even if one acknowledges that its probability of being true or approximately true is very low*.[27] The question is, hence, are there circumstances where a researcher whose epistemic utility function is represented by Vs would assess theories in an instrumentalist way? Luckily, it is easy to show that there are.

---

is also that philosophical arguments can play a role *within* scientific discourse in order to argue about the plausibility or naturalness of certain claims. I think this is specially true in the case of scientific revolutions, but of course, a detailed historical study would be necessary to assess this idea.

[27] One referee has suggested that an instrumentalist would be someone who does not even assign a probability to theories. I see no much difference in practice between that and what I am proposing: that not taking into account the probability of a theory, even if it is very low, would count as having an instrumentalist attitude. Whether you don't take probability into account because you think there is a probability but ignore it, or because you think there is no probability, would lead to the same choices.

In the previous subsection I have assumed that the relevant empirical evidence E is given and fixed, but obviously very often this is not really so. There may be situations where scientists consider that they already have all the necessary data to decide on the acceptability of the competing theories, i.e., that it is unlikely that the possible addition of new empirical findings is going to force them to reverse the judgments they have made on their theories. However, in many other situations this is not the case, and the relevant empirical data are just being *searched* for, or at least awaited or desired. Of course, this is more common in the first stages of a new research programme (to say it in a Lakatosian way), or when competition between different programmes is very strong. In cases like these, scientists will be *uncertain* about what the future empirical discoveries will be, and it seems reasonable that theories will not be directly judged according to a function like Vs, but according to its *expected value*, i.e., the mean value Vs(T,X) seems to have, given evidence E, averaging on the possible additions the available empirical evidence might have in the future. This expected value is easy to calculate:[28]

$$
\begin{aligned}
(15)\ EVs(T,E) &= \sum(w \in E)p(w,E)Vs(T,w) \\
&= \sum(w \in E\&T)p(w,E)Vs(T,w) + \sum(w \in E\&\neg T)p(w,E)Vs(T,w) \\
&= \sum(w \in E\&T)p(w,E)p(T,w)/p(Tvw) + 0 \\
&= \sum(w \in E\&T)[p(w)/p(E)][1/p(T)] \\
&= [\sum(w \in E\&T)p(w)]/[p(E)p(T)] \\
&= p(E\&T)/[p(E)p(T)] = p(E,T)/p(E) = p(T,E)/p(T)
\end{aligned}
$$

Hence, the estimated value of the empirical verisimilitude of theory T in the light of the empirical evidence E, is identical to the classical 'ratio measure of confirmation' (Horwich, 1982). An immediate result deriving from (15) is:

(16) If T and T' both entail E, then EVs(T,E) = EVs(T',E) = 1/p(E)

[28] For simplicity, I will use the expression "w $\in$ E" as an abbreviation of "a point in the logical space that satisfies E"). "Vs(T,w)" would refer, then, to the empirical verisimilitude T would have if the future evidence uniquely selects w as the real world.

Hence, scientists whose epistemic utility function can be represented by Vs, but that are still relatively uncertain about how the new empirical evidence will affect the acceptability of the theories under discussion, will tend not to value these theories according to how plausible they currently are, but *only according to how good their predictions have been*. This result is consistent with Lakatos' thesis that, in the first stages of a research programme, scientists only care about confirmations, and not about falsifications;[29] i.e., they don't discard a programme because of its failure to explain or anticipate some empirical results, but assess it only in function of its empirical successes, or, in other terms, *they have an instrumentalist attitude towards it*. Instead, once the empirical data are considered stable enough, i.e., not foreseeably changing the judgements over theories, these will be evaluated according to the actual, not the expected value of Vs, and hence considerations about *the plausible truth* of the theories will become important, or, to state it in a different way, scientists will start to manifest a realist attitude in their epistemic judgments of those theories. Furthermore, in the same way as the function Vs may be different for scientists with different subjective probability functions, it can also be the case that different scientists have different expectations about the evolution of the empirical data, and hence some tend to apply Vs or EVs under somehow different circumstances. But again, it is *their* decision to do one thing or the other; there is nothing like a 'philosophical solution' to the question of what *must* they decide.

Lastly, the model presented here has a virtue which is not very common within philosophy: it allows to make an empirical prediction. For what it claims is that scientists will tend to be more instrumentalist (in the sense of accepting theories even when recognising that they are probably far from the truth) when they expect big increments in the availability of empirical data, and more realist (in the sense of evaluating theories by giving a bigger weight to considerations of plausibility) when they do not expect such increments (other arguments justifying why it is Vs or EVs more reasonably used will produce their own empirical predictions). From my own

[29] Especially if we consider the sophisticated version of our utility function briefly discussed a few pages above, i.e., replacing Vs(T,E) with Vs(T,E(T)). Cf. Zamora Bonilla (2002b).

point of view, Lakatos' mentioned discussion about the different ways of judging research programmes when they are 'young' or when they are 'mature' would count as a (very) provisional confirmation of my prediction, but I invite historians, sociologists and philosophers of science to test it more thoroughly against real data.[30]

REFERENCES

Bartels, A., 2010, "Explaining Referential Stability of Physics Concepts: The Semantic Embedding Approach", *Synthese*, 41:267-81.

Brandom, R. 1994, *Making it Explicit,* Cambridge (Mass.): Harvard University Press.

Bueno, O., 2008, "Structural Realism, Scientific Change, and Partial Structures", *Studia Logica*, 89:213-235.

Chang, H., 2012, *Is Water H_2O?*, Dordrecht: Springer.

Douglas, H., and P.D. Magnus, 2013, "State of the Field: Why novel prediction matters", *Studies in History and Philosophy of Science*, 44:580–589.

Fine, A., 1986, *The Shaky Game*, Chicago: University of Chicago Press.

Frost-Arnold, G., 2010, "The No-Miracles Argument for Realism: Inference to an Unacceptable Explanation", *Philosophy of Science*, 77: 35–58.

Horwich, P., 1982, *Probability and Evidence*, Cambridge: Cambridge University Press.

Horwich, P., 1990. *Truth*, Oxford: Blackwell.

[30] Though it is not as directly relevant to the discussion about realism as the other questions examined so far in this paper, I would also like to mention that the hypothesis that Vs more or less correctly represents the epistemic preferences of real scientists can be used to make sense of the fact (mentioned in section 2.2) that prediction of unknown empirical results is more valuable than mere 'accommodation' of those previously known. The reason is that scientists would prefer, as we have seen, to develop new theories by starting with those whose assumptions that seem most likely to them. Hence, accommodating an empirical fact will demand to replace some of the theory's assumptions with another which is less likely, hence reducing the maximum value of p, and thence of Vs, that the theory can get from E. This explanation is coherent with some other recent arguments about the superiority of prediction over accommodation, in particular those linking the former to the expectation of future empirical success (see Douglas and Magnus, 2013, for a survey).

Ladyman, J., 2007, “On the Identity and Diversity of Individuals”, *The Proceedings of the Aristotelian Society*, Supplementary Volume 81:23-43.

Laudan, L., 1981. “A Confutation of Convergent Realism”. *Philosophy of Science* 48:19-49.

Niiniluoto, I., 1987, *Truthlikeness*, Dordrecht: Reidel.

Norton, J.D., 2014, “A material dissolution of the problem of induction”, *Synthese*, 191:671-690.

Olsson, E.J., 2002. “What is the problem of coherence and truth?”, *The Journal of Philosophy*, 99:246-278.

Psillos, S., 1999. *Scientific Realism: How Science Tracks Truth*, London: Routledge.

Putnam, H., 1981, *Reason, Truth and History*, Cambridge: Cambridge University Press.

Stanford, P.K., 2006, *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives.* Oxford, Oxford University Press.

Worrall, J., 1989, “Structural Realism: The Best of Both Worlds?”, *Dialectica*, 43: 99–124.

Worrall, J., 2007, “Miracles and Models: Why Reports of the Death of Structural Realism May Be Exaggerated” *Royal Institute of Philosophy Supplements*, 82(61): 125–154.

Zamora Bonilla, J. P., 1996, “Verisimilitude, Structuralism and Scientific Progress”, *Erkenntnis*, 44:25-47.

Zamora Bonilla, J.P., 2002a, “Scientific Inference and the Pursuit of Fame: A Contractarian Approach”, *Philosophy of Science*, 69:300-23.

Zamora Bonilla, J.P., 2002b, “Verisimilitude and the Dynamics of Scientific Research Programmes”, *Journal for General Philosophy of Science*, 33:349-68.

Zamora Bonilla, J.P., 2006, “Rhetoric, Induction, and the Free Speech Dilemma”, *Philosophy of Science*, 73, 175-193.

Zamora Bonilla, J.P., 2013, “Why are good theories good? reflections on epistemic values, confirmation, and formal epistemology”, *Synthese*, 190:1533–1553.