C.R.AIPS 3

Proceedings of *the Académie Internationale de Philosophie des Sciences*

Comptes Rendus de l'Académie Internationale de Philosophie des Sciences

Tome III

Models and Representations in Science

Éditeur

Hans-Peter Grosshans

Models, representation, and idealization
Revisiting the inferentialism debate

Jesús Zamora Bonilla

Lógica, Historia y Filosofía de la Ciencia, Facultad de Fliosofía, C. de Bravo Murillo, 38, Chamberí, 28015 Madrid, Spain

Abstract. The view of scientific models as 'inferential prostheses' is defended against some recent criticisms. In particular, I argue how the view can offer a reasonable answer to the problems of a general theory of scientific representation, and how it does not depend on ontological assumptions about denotation. A defense of the idea that models do actually represent the world, and that they can give us representations of the world increasingly closer to the truth, is offered against the radical artifactualism of Sanches de Oliveira and the anti-veritism of Potochnik.

> Ordo et connectio idearum idem est, ac ordo et connectio rerum.
>
> Spinoza

> God writes straight with crooked lines.
>
> Teresa of Ávila

1 Introduction: models as inferential prostheses

The Conference on 'Models and Representations in Science', held at the University of Münster in September 2023 under the auspices of the *Académie Internationale de Philosophie des Sciences*, was a perfect occasion for revisiting one topic which, though deeply essential in the evolution of my philosophical ideas, I had only discussed it in an explicit and detailed way in a couple of papers, co-authored with my colleague and friend Xavier de Donato, papers that already were more than one decade old. In preparing my participation in that conference, I realised that these papers, especially the one that had appeared in *Erkenntnis* in 2009, had received a number of interesting comments by other authors since its publication, but the fact was that we, due to the pressure of other tasks and projects in the meantime, had not had until then the chance of answering any of those comments. Of course, the following pages contain only my own opinions, and they must not necessarily coincide with Xavier's views on the topics I will discuss.

The most important claim of "Credibility, Idealization, and Model Building" was that scientific models have to be seen fundamentally as *'inferential prostheses'*. This idea fits rather coherently both with the views of scientific models as artifacts (e.g., Knuuttila 2011), since protheses are after all a kind of manmade tools, and with the inferentialist approaches to scientific models and representations (e.g., Suarez 2015), according to which the essential role of models is to facilitate surrogate inferences to their 'real world' targets.

Models and Representations in Science, edited by Hans-Peter Grosshans.
Comptes Rendus de l'Académie Internationale de Philosophie des Sciences 3 (2025).
J. Zamora Bonilla, Models, representation, and idealization, pp. 243–259.

We can summarise these similarities by saying that our approach combines Knuutila's *artifactualism* about scientific models' *nature* with Suarez's *inferentialism* about scientific models' *function.* In this respect, we do not claim that our view is an especially original one, of course; what was most innovative in our papers was the framing of the different abstract modelling strategies and functions (like idealization, optimisation, approximation, representation, understanding, etc.) within a more comprehensive inferentialist philosophical view, in particular, the normative-pragmatic-expressivist brand of inferentialism famously championed by Robert Brandom (e.g., Brandom 1994), as an extension of Wilfried Sellars (1963) idea of knowledge as the ability of "playing the game of giving and asking for reasons". It is reasonable, hence, that some of the comments our view has received have to do with the approach not being 'Brandomian' (or, as we shall explain, deflationist) enough, or with such an approach being able of answering the questions levelled to an inferentialist theory of scientific representation. The main goal of this new paper will be precisely to answer those criticisms, and spell out how the project fits with a deflationary view of scientific knowledge in general, and of scientific representations and idealizations in particular.

2 Brandomising scientific models

A possible way of putting the problem our approach attempted to tackle is offered by the following text from a recent book on scientific models as representations (Frigg and Nguyen 2020):

> Rather than attempting to investigate the conditions of epistemic representation by investigating the representational practices that establish it in every instance, one could instead take those conditions as foundational, and investigate how they give rise to representational practices, practices which themselves are explained by the inferentialist's conditions (rather than explaining them). Such an approach is inspired by Brandom's (...) inferentialism in the philosophy of language, where the central idea is to reverse the order of explanation from representational notions—like truth and reference—to inferential notions—such as the validity of argument. We are urged to begin from the inferential role of sentences (or propositions, or concepts, and so on)—that is, from the role that they play in providing reasons for other sentences (or propositions, etc.), and having such reasons provided for them—and from this reconstruct their representational aspects. So by analogy, rather than taking the representational practices (analogues of truth and reference) to explain the inferential capacity of carriers (the analogue of validity), we reconstruct the practices by taking the notion of surrogative reasoning as conceptually basic (...) Such an approach is outlined by de Donato Rodríguez and Zamora Bonilla (2009) and seems like a fruitful route for future research.[1]

[1] Frigg and Nguyen (2020), p. 92.

One of the main goals of an inferentialist conception of scientific models is, exactly, that of changing the typical order of explanation in philosophy of science, so that, instead of giving an account of how scientific inferences are possible thanks to the representational properties of the scientific models (or theories) that are employed in those inferences, what we do is to explain what scientific representations are, how they function, and how they are evaluated, in terms mainly of their inferential virtues. The fundamental idea of this inferentialist view is, indeed, the one captured by the expression *surrogative (or surrogate) reasoning*: scientists (and non-scientists alike, by the way) employ models as tools *on which* to perform some inferences that would be much more difficult (or directly impossible) to carry out 'directly on', or 'explicitly about', the real-world systems the models attempt to correctly represent (whatever this 'directness' actually may mean in each particular case). A model is a physical or mathematical system (or 'structure') some of whose properties and connections amongst its elements we know how to manipulate in order to perform some inferences within it, and we take profit of this inferential capacity by interpreting some of the model-system elements as 'representing' some analogue elements in the real-system that is our 'target'. The main difference with a 'representationalist' (and hence, non-inferentialist) account would be that the 'representing' part mentioned in the previous sentence is in itself explained in inferential terms: the fact that some elements in the model-system 'represent' some elements in the target-system reduces to the fact that model-users know how to perform some *inferences* from sentences that talk about the target-system to sentences about the model-system, and vice versa. This basically replicates Hughes (1997) famous "DDI-account", where the initials stand respectively for "denotation" (the inference from target to model), "demonstration" (inference within the model), and "interpretation" (inference from model to target), but the idea is at least as old a the German physicist Heinrich Hertz's (1894) description of how scientific reasoning is performed with the help of our "images in thought", and, if we allow ourselves a little bit of hermeneutical freedom, it may also be glimpsed in Baruch Spinoza's famous dicto according to which "the order and connection among ideas is the same as the order and connection among things" (*Ethics*, II.7), only that, as we shall see below, contemporary people cannot be as optimistic about that 'sameness' as the old rationalist philosopher was, because that equivalence has often to be seen as just a matter of more or less lucky *conjecture* and of bigger or smaller *approximation.*

Instead, according to representationalist approaches, like that of Frigg and Nguyen, an essential aspect of scientific models is their being a *representation* of some real systems, a notion that (besides creating some conceptual difficulties—to which we shall refer later—about models that do not have a specific real system as their target) presupposes the existence of some kind of *ontological* relation of correspondence between the model and the system. This relation can be interpreted both as the semantic relation of *reference* (the target being the Fregean 'reference' of the model, i.e., the object the model—as a symbol—'points to'), and as the also semantic property of *truth* (or 'accuracy': how well the model describes the target). But, contrarily to how it is at times (mis)understood, Brandomian inferentialism does not *deny* that we may reasonably talk about a model's reference or about a model's truth (or lack thereof); what this approach allows to do is to understand the meaning of the ideas of reference and truth in terms of their *expressive role*, i.e., in terms of what having those concepts permits to *say* to users of a language that contain the corresponding terms (as contrasting to users of some imaginary language that lacked any terms analogous to the concepts of truth and reference). According to Brandom, this expressive role is basically what he calls an *anaphoric* function ('anaphora' being the technical grammatical term for the function of *pronouns*). For example, the idea of 'reference' serves mainly to help speakers to determine when two expressions are co-referential, in the sense that one of them can substitute the other (as the pronoun 'substitutes' the name), and hence, when an inference from a sentence containing a name or description to a sentence containing another name or description is valid (if both names or descriptions are co-referential). Similarly, the expressive role of the concept of 'truth' is allowing speakers to assert, to deny, to question, to express doubts about, etc., some propositions that do not need to be repeated, or that are only indirectly or abstractly identified in the speech (as when I say "what this report contains is not totally true", or "everything that logically follows from true premises is true"). Hence, truth and reference are not primarily a kind of deep ontological stuff for the *philosopher* to discern, but a couple of mundane expressions that *ordinary speakers* employ in order to clarify to other ordinary speakers what they are saying or what they are talking about. I insist: this does not mean at all that a scientific proposition 'cannot be objectively true', or that typical scientific models 'do not refer to something in the world'. It only means that what the *philosopher* can say about it is not essentially different, nor more ontologically profound, than what *ordinary scientific speakers* tell when they say, for example, that the Crick-Watson model of DNA is a 'right' description of DNA molecules, or that its wires 'represented' electronic chemical bonds. All this makes sense of Frigg and Nguyen's correct description of the inferentialist approach as one not intrinsically *contrary* to the representationalist *claims* about which are the representational virtues and properties of scientific models, but only different in the *type of explanation* we offer of those properties and virtues: a good model is not good *primarily* in the sense that it 'rightly depicts the world', but is good primarily in the sense that it allows to carry out numerous, interesting and successful inferences about the world, and it is *because of this* that we *infer* that the model must probably be an accurate description or representation of the (relevant fragment of the) world ... an inference that, after all, it will be real practicing scientists (rather than philosophers) who will have to say if it is valid, or by how much it is, and under which circumstances.

Naturally, representationalist philosophers like Frigg and Nguyen are rightly entitled to ask how well inferentialist views of models, such as our own, responds to what they see as the main questions a philosophical theory of scientific representations has to answer, and also how it overcomes some possible general difficulties they see in other inferentialist approaches.[2] This is what I shall try to do in the section 4. Before that, it is relevant to consider other comments our approach has received, in particular from Khalifa, Millson and Risjord (2022). These authors have developed what they call "a thoroughgoing inferentialism", or a "thoroughly deflationary account of scientific representation", and (as I mentioned) criticise our own view for not being 'deflationary enough'. By 'thoroughgoing inferentialism', Khalifa *et al.* understand a view that "makes no appeal to denotation nor to any non-linguistic representation relationship in its account of surrogative inference". Our approach, instead, would—according to them—suffer from what they call "the smuggling objection", i.e., from necessarily presupposing some 'substantive' *denotation* relation in order to explain how our explanation of the representational capacity of models works. I confess it is difficult for me to discern how Khalifa *et al.* reach exactly this diagnosis of our approach, for Donato and myself said basically nothing about denotation in our paper. My guess is that their diagnosis arises from our use of something similar to the already mentioned Hughes' DDI-account, though we called "immersion" the first of the three inferential steps (instead of "denotation", as Hughes does), i.e., the inference from target to model. But my view is that it is not right to interpret that first 'D' (or our 'immersion' step) in a representationalist way, i.e., as an (ontologically) 'substantive relation' that needs to be presupposed *by the philosopher* in order to make her explanatory account to work. Rather on the contrary, in my own view at least (I cannot speak for other inferentialists), that first 'step' *is intrinsically and primarily an inferential step*, an essential part of the inferential practice of the model users; in particular, it corresponds to the inferences they make from *claims about the target system* to their presumed equivalents *in the model system*. All the full DDI cycle (inference from target to model, from model to model, and from model to target) has to be seen, in our *surrogative* account, as an indirect way of constructing inferences *from target to target*, like, for

[2] Frigg (2023), p. 275.

example, in the case of prediction making: we obtain some empirical data *about the real material system*, transform them into the language of the model system, and perform some model-based calculations whose results are in their turn translated as (still unchecked) claims about the material system; hence, what we do with the help of the model is an inference from the available empirical data to new testable predictions. All of the steps of this process are *inferences*, even the steps from the material system to the scientific model. If there is something here like a 'substantive denotation *relation'*, it is just the (scientific, not philosophical) *conjecture* that the model will be useful in allowing successful predictions by following those kind of operations (i.e, of *inferences*). Hence, I do not think that our approach 'smuggles' in any non-inferentialist-enough kind of elements. But perhaps a clearer and more explicit argument by Khalifa *et al.* could make me see the objection in some more positive view.[3]

Related to this, I find something problematic in Khalifa *et al.*'s attempt to explicate the surrogative inferential use of models with something like (what they call) an *inferential pedigree.* As they formulate the issue, the question is *how to justify inferences from the model to the target*, i.e., how to justify that a conclusion we reach within the model can be applied to the target. The 'inferential pedigree' would consist in the set of all reasons that make *this kind of inference* a legitimate one. But this applies only to the *third* step in the surrogate inference machinery (Hughes' 'interpretation', or the 'I' in the 'DDI-account'), whereas, as far as I understand the very idea of surrogate reasoning, it is rather the *second* D (Hughes 'demonstration', or what Donato and I called 'derivation') which is a literal surrogate or *substitute* of the inference (much more difficult to do without the help of the model) from the target to the target itself (or, more precisely, from *some* claims about the target to *other* claims about the target).[4] The other two types of inferences in the DDI-account serve, obviously, to operate the 'translation' from claims about the target to claims about the model,

[3]A related criticism has been recently leveled by Suárez (2024), in which our application of Brandom's inferentialism is, rather surprisingly, described as "a study of science through its linguistic categories and syntax, which was valuable prior to the semantic conception and mediating models revolution in the late 1990s, (but) looks too restrictive now". This is particularly shocking because our interentialist view of models' function and functioning is essentially *the same one* as Suarez's, only that embedded into a Brandomian inferentialist framework. It goes without saying that Brandom's own theory is immensely more general than just 'linguistic categories and syntax', for it is actually a full-fledged *theory of rationality*, and in particular, a *pragmatist* explanation of some fundamental *semantic* categories, i.e., an explanation of what people *say* in terms of what they *do.* A Brandomian view, hence, is as far from the (Carnapian?) 'received-view of scientific theories' of the positivist age as any of the other semantic-plus-mediating-models approaches may be.

[4]I would call these inferences from-target-to-target 'material' if the term were not already associated to another very precise meaning in philosophy of language and philosophy of logic.

but what needs a 'pedigree' is the *whole cycle* of inferences, not only the ones from model to target. In fact, some of the content of Khalifa *et al.*'s 'inferential pedigree' (for example, measurement) necessarily consists in fact in inferences *from the 'real' world to the model* (in this example, transforming empirical observations into numbers that can be expressed in the model's language). Khalifa *et al.*'s way of expressing it seems to present the role of measurement, and of other related arguments, as something that merely serves to 'justify', 'entitle', or 'support' the 'final' inference from model to target, but I tend to see measurement as simply one *part* of the whole process of scientific inference, and hence, as in need of an 'inferential pedigree' as any other part of the process.

3 A short interlude: do we learn from models?

The idea according to which the fundamental question of a theory of scientific models is that of what 'justifies' the inferences from models to targets seems to be equivalent to another assumption I find regrettably common in the relevant literature: the opinion that the main philosophical problem in this field is "how we can learn from models" (e.g., Morgan 1999). Actually, I think the answer cannot be simpler: *we just do not learn (about the world) from models.* What we learn about the world is *that* some models work and others do not (or which models work better and which models work worse), but *this* is something that we obviously do not learn *from the models themselves*, but that is learnt *from the world*, i.e., from empirical observations, or, more exactly, from repeated applications of the DDI cycle. *That* one model will work is simply a *conjecture* (at least at the beginning), and it is only *after* this conjecture gets enough empirical support that we can use the model as a 'sufficiently enough good representation' of its target. Obviously, *I myself* can learn a lot about chemical elements 'from' the model we call 'the periodic table', but this is because generations of chemists before myself learnt that the *empirical facts* support very strongly the hypotheses on which the table is grounded, and this is not something they could have learnt 'from the table itself'.[5] It is also true that models can *suggest* 'connections and order' *in the target* that we would not have envisaged without their help; after all this is the main reason why (at least according to an inferentialist viewpoint) we wanted models *for* to begin with: to draw consequences not easy to derive without the models. But these new consequences are in principle *as conjectural* as the conjecture that the model will be a good enough representation of its target (in fact, they are simply a

[5]There can be cases in which a model gets its support by the fact that it naturally derives from other models of theories *already well confirmed*, but this does not go against the general claim that *it is experience* what help us to ultimately learn *whether* models work better or worse.

part of *this* conjecture), and we need empirical knowledge *to learn whether* those inferences are materially correct or not, or to what extent they are.

In summary, we learn *about the world* from scientific models simply in the same way and in the same sense in which we 'learn' *from any other scientific hypothesis*: by applying the good-old-fashioned hypothetico-deductive method in any of its many variants. We do not learn *from the model* that the model is a useful representation of its target: we learn *this* from the experience about the target, even if that experience has been obtained in big part by following the specific conjectures suggested by the model. Stated otherwise: one model might be very fruitful in 'teaching' us lots of new interesting consequences from its internal structure, but if empirical data strongly *falsify* these consequences once they are applied to the real target systems (even if those consequences are stated in some loose approximate way), we would not say that this model 'teaches' us something *about the world*, except the possibly interesting fact that the world is *not* well represented by it.

4 Sketch of an inferential theory of scientific representation

As I said a few pages before, the most important test for an inferentialist theory of scientific (models as) representations would be to show whether and how it answers the main problems of a general theory of scientific representation. To avoid much speculation from my part, I will directly make use of Frigg and Nguyen own list of such fundamental problems for a theory of scientific representation:[6]

1. The representation problem: what makes of something a representation of something else.
2. The demarcation problem: what makes of something an epistemic/scientific representation.
3. The accuracy problem: under what conditions is scientific representation accurate.
4. The problem of carriers: what kinds of objects carriers are, and how are they handled.
5. The problem of targetless representations.

I think that our view of models as inferential prostheses allows to illuminate all these questions in a rather straightforward way. As for the first problem, this is just what 'surrogate inference' consists in, to begin with: the model is used to perform *indirectly* inferences about the target

[6] Frigg and Nguyen (2020), ch. 1.

system. The model's *being* a representation of the target just consists in its being *used* as a surrogate inference mechanism. There is no need of any substantive philosophical general explanation of how is this possible (our explanation is, hence, deflationary in Suarez's sense), for different models will work (better or worse) thanks to different 'physical', 'mathematical' or 'logical' reasons, and not because of some universal property like 'correspondence', 'isomorphism', etc. Furthermore, as I explained above, *that* a model 'rightly represents' its target is not a philosophical presupposition, but just a scientific conjecture that may end being corroborated or contradicted by the empirical facts.

A similarly so simple (or even simplistic) answer can be given to the second problem: what makes of a representation a *scientific* representation is just that it is used as such in the scientific process. There is no bigger mystery in this case than in the question of what is what makes of a laboratory, a measuring instrument, or a journal, a 'scientific' one. Of course, what is far from easy is to state what is the general difference (if there is one) between 'science' and 'non-science'. Generations of schollars have dismayed about trying to solve this old 'demarcation problem', or trying to determine if there is a solution at all, and it will not be me who pretends to have an answer to any of those questions. But, assuming that we have at least a minimal *pragmatic* understanding of when it is appropriate to use the adjective 'scientific' in numerous everyday contexts, I do not think we need more than this 'ordinary speaker' lexical knowledge to answer Frigg and Nguyen's second question.

The answer to the third problem comes also implicit with our answer to the first one: there is no general philosophical explication of what an 'accurate scientific representation' is, but, instead, how good one specific model is will depend on the contingent reasons that specific scientists will have for using that model in particular. The most general answer an inferentialist account can give is that a fundamental criterion to determine the value of a model will be how well it works in allowing to make numerous, useful and successful inferences about its targets. Donato and I summarised these types of reasons into what we considered the two most general categories applicable to the evaluation of models: *credibility* (or 'realisticness': how well scientists consider in the end that the model 'describes' the target) and *enlightening* (or 'understanding': how 'fluent' the process of inference-making is made, cognitively speaking, thanks to the model); but we doubt that there is something like a universal algorithm that can transform these two rather vague and context-dependent values into a precise philosophical theory of epistemic virtues.[7]

[7]Some may rightly point to a possible inconsistency between what I have just said in this paragraph and my own extended work on verisimilitude as a mathematical function of

As for the carriers problem, our approach sympathizes with Knuutila's artifactualism, as I already said. Models are inferential prostheses, and there are no limits as to the kinds of 'stuff' those prostheses can be 'made of', as long as they allow to make appropriate inferences in the way desired by their users. Models can be fully material (as plastic-and-wire models of organic molecules, or as Phillips hydraulic model of the British economy), or they can be totally abstract (at least, as abstract as mathematical equations can be, like in the case of Lotka-Volterra prey-predator model, or like the first cosmological relativistic models), or they can contain any mix of material-plus-computing machinery (as in the innumerable cases of mathematical models aided by graphic diagrams, or as the equally countless models that combine the use of computer programmes and of diverse hardware-processing and interface-devices). It is not even necessary that the model-'makers' know in detail how *it* performs the inferences it is supposed to make; this is clearly the case in the use of organism models in biology, but also in the case of deep-learning models in computer science: in these examples, the inferences the models make are not anything like 'mental' or 'abstract' operations, but are really physical *causal* processes whose *physical effects* are taken as the *consequences* of the relevant inferences.

Lastly, the problem of targetless representations is answered in the surrogate inference view just by 'switching-off' the target-model and model-target links (the first and third steps of the DDI-account), leaving them 'open' to possible future applications. Targetless models are just inferential prostheses that have not (yet) been 'attached' to a 'real system'. Their denotation-interpretation is just an open function that can be filled with a real system if and when appropriate. We can 'play' with them just to test their inferential capacities, with a pedagogical function (like 'finger exercises'), or as a representation of a non-existent by somehow 'possible' entity (like the discarded scale model of a building project, or the map of an imaginary land).

The account of models as inferential prostheses allows to understand, hence, the function of scientific models as representations in a way which is non-problematic from the philosophical point of view, and is even quasi-trivial in the sense that it shows that 'representation' is not an obscure and deep ontological relation between models and the world, but just a name for an important part of what model users *do* when they use the models: employing some elements of the model, and their formal configuration, in

the epistemic value of scientific theories (e.g., Zamora Bonilla 2013). The answer to that concern is that my work is not intended to be taken as a metaphysical speculation on the essential goal of scientific knowledge (or something like that), but only as a (quasi-)scientific *model* of scientists' epistemic preferences. Viewed this way, the model has to be assessed as a simplified, idealized, approximate, limited, and conjectural explanation of a small set of stylized empirical facts about how scientists evaluate their own hypotheses.

order to carry out surrogate reasonings about some real systems. This can be read as a *deflationary* theory of representation, both in the more general sense of not needing a *substantive definition* of what a representation is (beyond the fact that something is *used* to represent—i.e., to facilitate surrogate reasoning about—something else), and in the Brandomian sense of offering a merely pragmatical explication of the use of *representational vocabulary* in ordinary language (like Brandom did with the semantic concepts of truth and reference).

The near triviality with which an inferentialist theory of scientific representation permits to understand the use of representational vocabulary in the context of science makes me being more surprised of attempts aimed, not to offer an alternative theory of representation, but to dispense with all representational concepts whatsoever, like in the case of what Sanches de Oliveira calls 'radical artifactualism',[8] i.e., the project of explaining the use of scientific models as artifacts without any kind of representational function (that is, without assuming that models are models *of* something, or *about* something), but limiting ourselves to a purely 'enactive' description of the *material use* the agents make of models. I do not deny that there are lots of interesting things we can learn from the study of scientific activities from the point of view of enactivist approaches, and even from the study of the innumerable scientific practices that clearly are 'operational-but-not-representational', but I simply fail to see the point of a philosophical project that forces itself to interpret the pervasiveness (or rampantness) of representational vocabulary in science as just a misleading 'way of speaking' the poor scientists are led to use by the confounding influence of some nefarious philosophical dogmas. For me, this is as unintelligible as an attempt of explaining in enactive terms the material practices of luthiers and musicians avoiding all possible use of musical vocabulary, and interpreting the own musicians' use of that vocabulary as just the careless adoption of unsound metaphysical concepts. After all, the fact that pianos are most often made and used in order to play music with them is (at least for me) something as blatant and straightforward as the fact that the Crick-Watson model of DNA was a (better or worse) representation of the real DNA molecules of real cells, that Kepler's drawing of the elliptical orbits of the planets was a representation of their real trajectories around the sun, or that the periodic table is a representation of the types and mutual relations between the real types of chemical elements. If some *philosophical theory* of scientific representation leads its supporters to commit themselves to some outrageous or implausible metaphysical claims when interpreting this type of trivial facts, that would be a reason to doubt of the soundness of *that* theory, but not to react with the *still more implausible* opinion like that 'the Rutherford's

[8] Sanches de Oliveira (2022).

model of the atom was not a model *of the atom*, because nothing is a model *of anything'*. And I think that the view of scientific models and representations as instruments for surrogate reasoning allows to make philosophical sense of those trivial facts without forcing us to choose between any kind of controversial philosophical explication (either 'ontological' like Frigg and Nguyen's, or 'nihilist' like Sanches de Oliveira's) about the 'ultimate nature' of scientific representation.

5 Idealization and truth

In this last section I turn to what is perhaps the most philosophically controversial issue regarding scientific models and representations: their relation to the idea of truth. This is particularly problematic because of the also obvious fact that scientific models tend to *clearly deviate from being an accurate description* of the real systems we try to represent with them. Hence, they are in most many cases *literally false and distorted* descriptions of the world, or what we can call, following Angela Potochnik recent work on this question, 'rampant and unchecked' *idealizations*.[9] In fact, I think that one of the reasons why in the last decades both scientists and philosophers of science speak much more of 'models', rather than of 'theories' and 'laws' as their grandparents used to do, is because we have become much more aware of the fact that scientific representations tend to be ephemeral caricatures much more often than marble-engraved decrees. "Of course the real-world target systems", every scientist worth her salt would unhesitantly acknowledge, "are not *literally* like our models say they are; models are most often *extremely distorted* and *very partial* representations of their targets!". The question is, can we derive from this platitudinous fact the conclusion that, as Potochnik claims, "science isn't after truth" at all? According to her, the prevalence of un-truth in science would immediately prove that "science is not in a lockstep pursuit of truth. Instead, there are a variety of scientific aims that are in tension with one another, and the ultimate epistemic aim of science is not truth but understanding".[10]

I cannot enter here into a full discussion of the very detailed and interesting work Potochnik does on the presence in science of different types of idealizations, and of the roles of most of the 'epistemic aims' she mentions, and so my comments will probably be much more abstract, limited and general than what an exhaustive criticism of her work would demand. My main argument is that neither the widespread presence (and use) of blatant falsities in science, nor the existence of other goals different from 'literal truth', entail in any way that the pursuit of truth has to be discarded as one essential goal of the kind of practice we call 'scientific research', and

[9] Potochnik (2017).

[10] See esp. Potochnik (2017), pp. 90–91.

more seriously, that forcing ourselves to see science as an endeavour totally disconnected from the pursuit of true knowledge about the world prevents us to rightly understand the methods and the accomplishments of science. First of all, the inference from the premise that ‘most scientific models contain radically false assumptions’ to the conclusion that ‘being in some relevant sense closer to the truth is not an important goal of science’ is prima facie as doubtful as the inference from ‘almost all human beings through history have lived in misery’ to ‘the pursuit of economic wealth is not an important goal for humans’. Perhaps most scientific models are indeed very far from the truth, but this does not mean that scientists would not often be happier if they knew how to replace them with some models that were substantially closer to the relevant truths. It is even conceivable that in some cases scientists may consider that the passing from some old model to a new one that is recognisably ‘less true’ *in all the relevant aspects* is a case of scientific progress, because other values different from ‘truth’ are better exemplified in the second model; but in order to show that this makes truth an *irrelevant* value in science one should have to demonstrate that this type of examples are not something occasional, but systematic, or at least, that we cannot just explain them as cases in which one of the multiples values is given preference over another value without entailing that this second value is ‘unimportant’. Let’s illustrate this argument with a different goal in mind, one that (though non-epistemic in nature) is obviously very important in scientific practice: *cheapness*. A research team may opt for using a calculator that is known to commit more mistakes than another one, if the second is extraordinarily more expensive than the first. From this we should not infer that exactitude in the calculations is ‘not relevant at all as a scientific goal’, only that, as most kind of goals human beings have, there may be trade-offs between them. Hence, we may also say that truth and understanding (or, in the terms employed a few pages above: credibility and enlightening) can be in a trade-off relation, without this entailing that some of the two goals is *irrelevant* just because the other happens to have more weight in some, or even in most cases.

Second, and more importantly, it is not even that the pursuit of (closeness to the) truth can in principle be taken as an important *goal* of science even in spite of most scientific models containing blatant falsities: I think that we can argue for the much stronger thesis that science has actually been considerably (and often spectacularly) *successful* in providing us with knowledge of the world that is substantially close to the truth, and that history shows, without the need of any kind of whiggism, that in many areas we have made a lot of progress in getting more and more detailed knowledge of the furniture and working of the world. In some cases, this may have been done even at the cost of having *less* understanding as we (thought we) had before: often what happens is that we transit from a vision of some segment of nature that provides both a neat small collection of elements and a simple explanation of its mutual interconnections, to a view that recognises the existence of a plethora of very different entities but simultaneously a much messier and less intelligible causal or taxonomic network between them (think, for example, in the evolution of the catalogues of astronomical entities, or of the groups of living beings at different levels). In cases like this, it is absurd to require scientists that they renounce to the big amounts of new mundane truths they have discovered, just because the previous vision of the field gave them a stronger feeling of 'understanding'. But, of course, in many other cases the progress in truth has fortunately gone hand in hand with a parallel progress in understanding, and we end both knowing much more things about the world, and understanding them in a more efficient way.

In the third place, I think that the (for me, rather bizarre) anti-veritist attempt of dispensing with the basically trivial claim that science has very often been considerably successful in the pursuit of truth has a similar explanation to the one I have just offered of Sanches de Oliveira anti-representationism: the confusion of the possible shortcomings of some *philosophical theories* about the nature of truth or representation, with failures in the run-of-the-mill understanding that *ordinary scientists* may have of the properties and virtues of their models *when they themselves use representational or veritistic language* to discuss a lot of things about those models and their connection with the world. In the case of Potochnik, the confusion probably derives from the supposition that the concept of 'truth' must refer to something like an absolute point-by-point metaphysical correspondence between our statements and an absolutely precise ontological scafolding of the world in itself, or something like that, and hence, that the scientific acceptance of anything that fails to be exactly identical to such a 'literal, absolute, and eternally unchanging truth' should be considered as a refutation of the idea that scientists pursue in some interesting sense 'true knowledge about the world'. But if we understand the concept of truth *and the concept of approximate truth* in a deflationary sense,[11] as just expressive tools of ordinary scientific language (rather than as a philosophical relation—whatever that could mean—between language and the world), we can easily see that a scientific model *being successful* in the sense of being approximately true (or 'close enough to the truth' for the relevant purposes) is not something requiring an ontological analysis (probably doomed to be engulfed by conceptual paradoxes), but just one of the things real scientists *say of their models* when they evaluate them: employing a model usually consists (as we

[11] The simplest deflationary definition of 'approximate truth' is given by Smith (1998): *a proposition 'X' is approximately true if and only if approximately X*. Of course, it is not the philosopher, but the practicing scientist, who has to decide in each case what senses and degrees of 'approximation' are relevant.

saw above) in the *conjecture* that the causal structure of the target system is 'close enough' to the inference-permitting structure of the model, so that the inferences made with the help of the model will be 'accurate enough' when applied to the target, and the model *being successful* usually consists in the fact that this conjecture (i.e., *the conjecture that the world is 'approximately like' the model in some relevant aspects*) being 'confirmed enough' by the empirical data. Once this success has been established strongly enough in the course of empirical research, the fact that some aspects of the model are not exactly, nor even remotely, 'like' their possible 'analogues' in the target systems is in many cases no argument against the conclusion, for the original conjecture did not affirm that the world had to be *literally or exactly* like the model system in all respects, only that it was 'similar enough in the relevant ways'. Changing the discussion from whether science tries to discover a true description of the world, to whether it tries to understand real causal patterns (as Potochnik defends) does not move a millimetre the argument in favour of Potochnik's anti-veritism, for our deflationary view of truth helps us to be agnostic about the 'right transcendental stuff' the world may be made of, inviting us to concentrate just on scientists' assertions or claims, taking 'truth' as just another expressive tool with which to formulate those same assertions: if scientists *claim* that one model captures better the causal patterns of a target system than another model, then what a deflationist infers from this is that scientists consider that *it is true* that the first model captures better those causal patterns than the second, and *that's all the truth that is relevant* in the discussion about whether scientists pursue the truth or not.[12]

Hence, scientific models and scientific theories being filled with idealizations 'that radically depart from the truth' is no reason at all to put into doubt science's capacity of getting an increasingly approximate knowledge of the truth about the systems it studies, for many of these idealizations are, on the one hand, not 'mere falsities', but approximately accurate descriptions of some real things (like point-masses in astronomy may be 'accurate enough' for many purposes), and on the other hand, because even if some idealizations are not justifiable as 'approximations' in this loose sense, this does not go against the fact that the models containing them can succeed in *saying many right things about the world* thanks in part to the working of those fictional elements. Paraphrasing Teresa of Ávila, we can say that very often *science discovers the truth by means of false idealizations.*[13]

[12]For a more detailed argument between the connection between scientific realism and deflationism, see Zamora Bonilla (2019).

[13]I cannot finish these comments on Potochnik's book without mentioning the surprise it caused me to realize that she failed to even mention the author that has been probably most influential in promoting the idea that all interesting scientific hypotheses are basically false (and in introducing the debate on whether this fact can be nevertheless coherent

Acknowledgement. Research for this paper has benefited from the Ministerio de Ciencia e Innovación (España) research projects PID2021-123938NB-I00, PID2021-125936NB-I00 and PRX22/00154.

References

Brandom, R., 1994, *Making It Explicit: Reasoning, Representing, and Discursive Commitment.* Harvard University Press.

Brzeziński, J., *et al.*, 1990, *Idealization I: General Problems*, Rodopi.

Donato Rodríguez, X, and J. Zamora Bonilla, 2009, "Credibility, Idealization, and Model Building: An Inferential Approach", *Erkenntnis*, 70:101–118.

Donato Rodríguez, X, and J. Zamora Bonilla, 2011, "Explanation and Modelization in a Comprehensive Inferential Account", in H. W. de Regt *et al.*, *EPSA Philosophy of Science: Amsterdam 2009*, Springer, pp. 33–42.

Frigg, R., 2023, *Models and Theories: A Philosophical Inquiry*, Routledge.

Frigg, R., and J. Nguyen, 2020, *Modelling Nature: An Opinionated Introduction to Scientific Representation*, Springer.

Hertz, H., 1894, *Die Prinzipien der Mechanik in neuem Zusammenhange dargestellt*, Johann Ambrosius Barth.

Hughes, R. I. G., 1997, 'Models and Representation', *Philosophy of Science*, 64:325–336.

Khalifa, K., J. Millson, and M. Risjord, 2022, "Scientific Representation: An Inferentialist-Expressivist Manifesto", *Philosophical Topics*, 50:263–291

Knuuttila, T., 2011, "Modelling and representing: An artefactual approach to model-based representation", *Studies in History and Philosophy of Science, Part A*, 42: 262–271.

with the idea that scientists try to discover theories that are closer and closer to the truth). I'm referring, of course, to Karl Popper. Potochnik did not also refer at all to the philosophical literature on truthlikeness or verisimilitude stemming from Popper, probably disdaining it by assuming that, for those philosophers, scientific progress is "a lockstep pursuit of the truth" (which of course *it is not* what they assume). Lastly, it is equally shocking for me Potochnik's failure to even recognize the existence of the large philosophical literature on idealization elaborated from a formal point of view (see, e.g., Brzeziński *et al.* (1990), and all the subsequent series of volumes on idealization in the *Poznan Studies on Philosophy of Science*). It seems as if Potochnik were denying to philosophers of science what she acknowledges that scientist constantly do in a "rampant and unchecked" way: using idealized models to understand their object of study.

Morgan, M., 1999, “Learning from Models”, in M. Morgan and M. Morrison (eds.), *Models as Mediators: Perspectives on Natural and Social Science*, Cambridge University Press, pp. 347–388.

Potochnik, A., 2017, *Idealization and the Aims of Science*, The University of Chicago Press.

Sanches de Oliveira, G., 2022, “Radical Artifactualism”, *European Journal for Philosophy of Science*, 12:1–33.

Sellars, W., 1963, *Science, Perception and Reality*, Routledge & Kegan Paul Ltd.

Smith, P., 1998, “Approximate Truth and Dynamical Theories”, *The British Journal for the Philosophy of Science*, 49:253–277.

Suárez, M., 2015, “Deflationary representation, inference, and practice”, *Studies in History and Philosophy of Science, Part A*, 49:36–47.

Suárez, M., 2024, *Inference and Representation: A Study in Modelling Science*, The University of Chicago Press.

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

Zamora Bonilla, J., 2019, “Realism versus anti-realism: philosophical problem or scientific concern?”, *Synthese*, 196:3961–3977.