

As-if models and scientific realism: a response to Moscati

Antoine Brandelet¹ and Jeremy Attard^{1,2}

¹Department of Philosophy and History of Science, University of Mons, Belgium

²Department of Sciences, Philosophies and Societies, University of Namur,
Belgium

Abstract

In a recent paper, Moscati provides an analysis of as-if modelling in economics, more precisely in decision theory. According to him, most of decision theoretical models—be it neoclassical, behavioural or heuristic—are as-if, and that should be no worry at all. That observation is then taken to be a reason to move forward a realist understanding of modelling and embrace a form of instrumentalism.

Here, we wish to support Moscati's claim of the positive epistemic value of as-if modelling while keeping a realist interpretation. We show that the core of the instrumentalist critique he provides is based on a misunderstanding of the very mechanism of as-if modelling. We then propose a realist framework in which the variety of modelling strategies, in economics as well as in physics, can be described.

1 Introduction

In a recent paper [Moscati, 2023], Ivan Moscati addresses the question of as-if modelling in economics, more precisely that of (anti)realism in decision theoretical models. He first distinguishes three classes of such models: neo-classical, behavioral and heuristic. The major representative of the first class is rational choice theory, or expected utility theory, in which agents are assumed to attach a utility value (a real number) to each outcome of possible risky choices, and to make the choice which maximizes the expected utility. The second class contains attempts to improve neo-classical models by various means, including more complex assumptions about the form of utility function or accounting for biased perceived probabilities of different outcomes by the decision makers. The third class also aims at amending neo-classical models but unlike the two first ones does not assume that agents maximize anything, but only that they make choices among a finite set of possibilities following some heuristic (simple) rules. Behavioral and heuristic models historically developed in reaction to epistemological criticisms addressed to neo-classical models. The most salient is the lack of realisticness of their basic assumptions due to the accumulating amount of evidence that these basic assumptions seem to be systematically violated in decision theory experiments. That is to say, the first class of models was originally criticized for it seems to commit to an instrumentalist philosophical

position known as the “as-if” account of models in economics, mainstream in this field from the seminal work of Milton Friedman [1953]. Moscati gives the following definition of the as-if account of modelling in decision theory [Moscati, 2023, p. 2]:

Broadly speaking, as-if models attempt to account for the observable choices that individuals make, but do not pretend to capture the underlying psychological mechanisms that might generate those choices. Some underlying choice-generating mechanism, such as utility maximization, is attached to the model. However, in the as-if approach the decision theorist is agnostic about whether this mechanism actually operates in the mind of the decision maker. She may even deem, and explicitly acknowledge, that the posited mechanism and its components (such as the utility function, the preference relation or the heuristic rules), are only fictional constructs. Nonetheless, the decision theorist explains, describes or predicts the decision maker’s choices as if they were generated by the posited psychological mechanism at issue. Insofar as the as-if model is capable of accounting for the decision maker’s choice behaviour or indicating ways to effectively control it for economic policy purposes, the model is considered scientifically valid.

The two other classes of models are made to be more acceptable from an epistemological viewpoint for they rest on human behaviour’s assumptions which correspond more to what is otherwise observed in experiments. Moscati defends two main positions: 1/ *all* decision theoretical models, including behavioral and heuristic ones, are as-if models – whatever they claim to be; 2/ there is *no worry* at all with this situation, as long as it is seen within the right antirealist framework. According to him, the latter goes beyond traditional instrumentalism, as explanation is considered a genuine and important aim of science.

In this paper, we would like to support Moscati’s claim of the positive epistemic value of as-if modelling, while keeping a realist interpretation. Our paper is structured as follows. In Section 2, we first clarify the discussion by distinguishing between “realisticness” of basic assumptions, in the sense of being compatible or coherent with what is otherwise known, and “realism” of theories’ contents, in the philosophical sense of scientific realism. We argue that this distinction is relevant precisely because the question of scientific realism takes on its full meaning in the cases we cannot know directly whether these assumptions are true and whether the entities to which they refer actually exist. We then take some examples in physics to argue that requiring realisticness of basic assumptions is too strong and demanding a criterion which is not even reached in physics, contrary to what is sometimes claimed, without a priori impinging on the acceptability or scientificity of physical models. This observation ties in with the more general fiction view of models, presented in section 3. From this viewpoint, scientific models are fictional entities the components of which do not have to be taken literally but only as an embodiment of an epistemic representation of the target systems. Moreover, the fiction view does not a priori commit to any form of antirealism or realism. Nevertheless, two important questions still have to be addressed: 1/ can we build an

epistemological criterion of model selection escaping mere relativism without demanding realisticness of basic assumptions? (epistemological question) 2/ accepting that unrealistic assumptions still play an important cognitive role in scientific inquiry do condemn us to any form of antirealism? (realism-related question). In this paper we address these questions alternatively. In section 4 we develop an epistemological criterion encompassing the observations previously made. In section 5 we then argue that embracing the fiction view of models does not force us to any form of antirealism. On the contrary, we build a realist framework compatible with a fiction view of models and argue that the answers we provide to both the epistemological and the realism-related questions actually strengthen each other. In section 6 we finally provide a veritist account of scientific explanation compatible with the as-if view of models. We argue that even if Moscati defends a position going beyond mere instrumentalism, his mechanism-based account of explanation actually do not succeed in reaching his goal.

2 What does it mean to have “realistic” assumptions?

Basic assumptions of decision theory models are thus often criticized for their lack of realism, which is seen as a fundamental epistemological drawback. Yet, there are two different - although related - philosophical discussions at play here, which are sometimes confused and that we would like to distinguish. Indeed, when scholars talk about the “realisticness” of basic assumptions in such models, they often mean their *consistency* with what is otherwise known – about human behavior, for instance. Consistency of basic assumptions with what is otherwise known is an epistemological demand – that is, a criterion of acceptability or justification of a good scientific model. However, this “realisticness” of hypotheses should not be strictly confused with “realism” in the sense of *scientific realism*. As Moscati points out in his paper, the latter is a philosophical framework of interpretation of the content of scientific theories, in their ontological, semantic or epistemic aspects, and in relation to their empirical success.

Discussions about scientific realism are particularly focused on hypotheses bearing on entities which are *not* observable but still play an important role, e.g. in the explanation of the phenomenon studied. The distinction between realisticness and realism is thus quite relevant in this case. Indeed, it is precisely because we cannot know directly whether these hypotheses are true (and whether the entities to which they refer actually exist) that the question of realism (semantic and epistemic in particular) takes on its full meaning. Moreover, in the case these assumptions can be assessed and are shown to be false, or more precisely inconsistent with some background knowledge, this distinction allows to ask the question of how knowingly false assumptions can still play an important cognitive role in scientific inquiry. This question is actually relevant because in a lot of cases, models which are fully acknowledged as being of high reliability and scientificity, e.g. in physics, do rest on such not-only-idealized-but-false assumptions. We now argue that the comparison with these physics’ examples suggests that realisticness does not seem to be a necessary criterion of acceptability.

Lack of realisticness of basic assumptions in theoretical models in physics has at least

three dimensions that we would like to highlight here. First, some basic assumptions are not testable independently of the very theoretical framework within which they are used. Indeed, on the one hand, fundamental principles as Newtonian law of motion cannot be tested per se, but only through the specification of some auxiliary hypotheses – here, the specification of types of forces or energy which are assumed to be at play in the phenomena studied. On the other hand, these specific hypotheses cannot be independently tested neither. Take the fact that kinetic energy E_k of a particle of mass m and speed v reads:

$$E_k = \frac{1}{2}mv^2. \quad (1)$$

Hypothesis (1) does not have any meaning outside Newtonian mechanics, and more precisely without a general law about how energy of physical systems described by Newtonian mechanics behaves. Then, when hypothesis (1) is used in a physical model, it is not justified by its mere independent truth, but only because the whole system:

$$\{\text{fundamental principles} + \text{specific hypotheses}\} \quad (2)$$

gives rise to a model which produces falsifiable and empirically adequate predictions.

Second, some basic assumptions can be tested independently of the model at play – however, they can still be considered as good and justified hypotheses even when they turn out to be wrong, i.e. refuted by independent investigations. Take as an example the kinetic theory of gas. Ideal gas law: $PV = nRT$ can be derived from microscopic considerations, modelling the gas as a set of $N \approx 10^{23}$ identical particles (material points of mass m) confined in a volume V . The particles are assumed to interact only with the walls of the box they are confined in via elastic shocks, and not between them. These basic assumptions are false at two levels here, and still the model they take part allows for a computation of the ideal gas law which is considered as a genuine explanation. On the one hand, if we consider that the gas is constituted of material particles, then at temperature under consideration they probably experience *millions* of shocks per second. Thus, the absence of interaction between particles is not a mere approximation but a knowingly false assumption. On the other hand, we know from more fundamental theoretical framework as quantum mechanics that the gas molecules are *not* mere material points or particles but more complicated objects as wave functions only giving information about probabilities of different states in which the system can be.

Finally, as it is presented e.g. in [Suppe, 2000] as an argument against some logical empiricists' positions, a given physical theory can usually be formulated in several distinct ways such that these formulations, even if theoretically and empirically equivalent, postulate *different* entities or mechanisms. A well-known example is Newtonian mechanics, which can be formulated postulating either a set of forces modifying the trajectory on material points, a Lagrangian associated to the material point such that the latter follows the trajectory which minimizes the integral of the Lagrangian along the trajectory or a Hamiltonian driving the dynamics of the physical system seen as

a point in a phase space, according to Hamilton equations. More specifically, classical *gravity* can be formulated in the Newtonian framework by postulating well-known $1/r^2$ -gravitational force, but can also be formulated in the Newton-Cartan geometrical framework without postulating forces but rather describing it as the manifestation of the curvature of an underlying (classical) spacetime [Ehlers, 1973, Chapter 1]. This is not a specificity of Newtonian mechanics, but a feature shared by all physical theories. For instance, quantum mechanics acknowledges at least nine different formulations [Styer et al., 2002]: Heisenberg’s (matrix), Schrödinger’s (wavefunction), Feynman’s (path integral), Wigner’s (phase space), density matrix, second quantization, variational, de Broglie-Bohm’s (pilot wave) and Hamilton-Jacobi’s. As for General Relativity, it can also be formulated different ways [Göckeler and Schücker, 2011, Krasnov, 2020, Arnowitt et al., 2008]: using tensor calculus on differential manifolds, Cartan geometry on principal fiber bundles, encoding gravitation in torsion instead of in curvature, or deriving fields equations from Lagrangian or even Hamiltonian principles. All these formulations are equivalent in the sense that it is perfectly known how to mathematically pass from one to the other, and have the same empirical success. However, entities that they postulate (forces or Lagrangians, quantum states as vectors or density matrix, spacetime torsion or curvature, and so on) together with fundamental mechanisms and laws which lie at the foundation of their explanatory power have sometimes nothing to do with each other. Picture of reality given by these formulations are thus deeply distinct, even though the latter are epistemologically equivalent and all highly acceptable as explanatory models.

This observation then highly *suggests* that realisticness of basic hypotheses is finally neither necessary nor relevant a criterion as for the acceptability of theoretical models. This very statement actually ties in with the more general *fiction view* of models, presented in section 3. As already written in the introduction, this view nevertheless raises two important questions: an epistemological one (how to justify the use of a model rather than another one if the realisticness of their basic assumptions is not necessary?) and a realism-related one (do such a view condemn us to antirealism?). We tackle these issues in section 4 and in section 5 respectively.

3 As-if modelling and reference

3.1 Models are fictions, and that’s no worry at all

The central tension at play may be called the *modelling conundrum* and can be stated as follows:

Definition 1 (Modelling Conundrum (MC)). Scientific models play a central role in representing and explaining, and yet they feature idealisations and deliberate falsities.

MC is often taken as a basis of a critique of truth-oriented positions in various sectors of epistemology and philosophy of science. An obvious example is *naïve scientific realism*: if an empirically adequate and predictively powerful model is interpreted as a faithful description of a phenomenon, then how is it possible to account for its idealised characteristics?

There is therefore a close relationship between the emphasis placed on the “as-if” nature of models and antirealist or instrumentalist conceptions of science. Idealisations, simplifications, abstractions or even the introduction of purely fictional entities or mechanisms, possess obvious epistemic virtues (see e.g. [Sullivan and Khalifa, 2019, Lawler, 2019, Nawar, 2019, Pincock, 2021] for a further discussion of idealisation’s epistemic virtues): simplicity, computability, enabling understanding, etc.

As we showed in section 2, the conundrum has less to do with the realisticness of the model’s basic assumptions than with the interpretation of its empirical and predictive power. We gave several examples in physics where models feature this kind of false, or incompatible with background knowledge, assumptions. Even the concept of physical law considered as *ceteris paribus* generalisation [Cartwright, 1983] has to do with the idealised status of models: laws are true of the models, not of the messy reality they, at most approximately, represent.

If Moscati’s first assumption about “as-ifness” of models is inaccurate, we can still make sense of the core of his argument: the as-if nature of modelling is problematic for truth-oriented conceptions of scientific success.

The problem can now be framed in more specific terms: is it possible to accept Moscati’s second claim, namely that as-ifness is not problematic, while resisting antirealist or instrumentalist tendencies? We think the response is positive and the remainder of this paper is intended to defend this claim.

MC is particularly noticeable when adopting the so-called “fiction-view of models” [Frigg and Nguyen, 2016, 2020, Toon, 2012]. In this perspective, models are conceived as props in a game of make-believe: an agent using a model of some target system, be it physical, social or economical, is engaging in a game of make-believe by accepting a set of hypotheses for the sake of manipulating the model, i.e. deriving new fictional propositions (propositions true inside the fiction) from known ones. This framework is interesting because it makes clear the distinction between what is true inside the model and what is possibly true outside, i.e. in what the model portrays. Engaging in a game of make-believe involves accepting some knowingly false assumptions only for the sake of the game: no ontological weight is attached to them.

That explains the coexistence of incompatible hypotheses or models, as in the case of semi-classical models in physics. A physicist manipulating the billiard ball model in order to work out the temperature or pressure of a perfect ideal gas using the well-known equation $PV = nRT$ is not committed to the claim that molecules bounce off each other *contre-sens ici non? pour dériver $PV=nRT$ on suppose pas que les molécules sont des sphères qui tapent dedans, mais juste des points matériels qui ne rebondissent que sur les parois de la boîte où est enfermé le gas. Donc on pourrait plutôt dire: “is not committed to the claim that molecules do not bounce off each other”*. The important aspect of the fiction view is then to highlight the epistemic role of idealisations while keeping a distinction between what the model describes literally and what are the conclusions the user might want to infer to the target.

Therefore, we see that the critique of economical as-if models as faithful representations of target systems and mechanisms rely on an inaccurate understanding of the status

of idealisations and knowingly false assumptions in modelling. The latter have epistemic virtues and explanatory power (see Section 6 hereunder for a further discussion of the explanatory power of fictional models), but they do not reduce the models to predictive tools. Of course, that does not mean that antirealism and the fiction view cannot go hand in hand. Indeed, it is easy to see what a position like that would look like: the comparison between scientific models and fictions leads to the conclusion that models are nothing but fictions, that is things that can be more or less useful for certain goals, but certainly not true or adequate in any sense. For example, Vaihinger introduces in his seminal work *Philosophy of 'as-if'* a distinction between hypotheses and fictions, where the latter are evaluated on their expediency:

That is why in the case of a number of hypotheses, all equally possible, the most probable one is always selected. On the other hand, in the case of a number of equally possible fictions, the most expedient is chosen. [Vaihinger, 2009, p. 85, note 1]

Regarding as-if modelling, and keeping in mind the content of MC, we can now better understand Moscatti's proposal of being agnostic about "whether this mechanism actually operates in the mind of the decision maker". [Moscatti, 2023, p. 2]

This attitude is exactly what the make-believe view of modelling puts forward: an agent might want to use the assumption that such and such mechanism operates in the mind of the decision maker while not believing it does, just as a reader might engage in a novel without believing a school of magic does exist. All the reasons we have to believe that these assumptions are not realistic (because they are not compatible with what is known of human psychology, for example) do not change anything to the kind of game the modeller is playing when using the model. Trying to save the realisticness of these hypotheses by assuming they are not conscious in the mind of the decision maker simply shifts the problem. As Moscatti puts himself about these model-mechanisms: "it is more appropriate to acknowledge that they take place in the conscious mind of the decision theorist." [Moscatti, 2023, p. 12]

The tension between realist and antirealist positions regarding (MC) is then to be found in the process of justification of the models. If a model is (in, for example, Vaihinger's perspective) only assessable in terms of (epistemic) utility, if no external justification can be found, then it provides an argument in favour of antirealism.

Our goal in the remainder of this paper is to show that the fictional context may as well be used to defend a realist position by providing a truth-oriented justification of the models.

3.2 Realism and as-ifness: good friends after all?

As we have seen, the realist challenge is not about interpreting psychologically the functions that appear in the economical models. With this regard, they are perfectly acceptable as they are: fictional statements true only in the model and not applicable to the target systems, i.e. human agents in a specified economical system.

The realist challenge rather concerns the possibility of establishing independent means of justification of the model. To understand the very nature of the problem, let us take a closer look at Moscatti’s flavour of antirealism. As he defines it [Moscatti, 2023, pp. 18-20], following the three-fold definition of [Psillos, 1999], he:

- Accepts the metaphysical thesis according to which there is a mind-independent reality which is the object of scientific inquiry;
- Accepts the semantic thesis about observable, but remains sceptical about unobservable;
- Accepts the epistemic thesis about observables, models are true descriptions of the observable, while being sceptical about unobservable entities and mechanism that produce the observed behaviour.

As he himself notes, this position is a classical form of instrumentalism, the only difference is that Moscatti also defends a variant of mechanistic explanation that goes beyond traditional instrumentalism. For that reason, we have refrained from using the qualifier “instrumentalist” for his position. We will specifically address the explanation problem in Section 6.

There is no particular worry concerning the metaphysical thesis, whereas both the semantic and epistemic need to be further discussed.

Concerning semantic realism, Moscatti defines it in terms of *literal representations*:

According to semantic realism, scientific terms should provide literal representations of the observable and unobservable entities they refer to, although also approximate representations, such as abstractions or idealizations, are admitted. [Moscatti, 2023, p. 18]

Stated in these terms, it becomes clear that the as-if interpretation of modelling carries an antirealist understanding of the situation: models, as supports of games of *make-believe*, are non-committal by nature about the entities they portrays. However, it could be argued that this use of the semantic thesis is quite inadequate for its purpose. The core of the semantic thesis is not the approximate validity of the representations provided by the model, but the fact that models provide representations. The term “literal” means that models or theories are descriptions of matters of facts, representations of phenomena. One of the most discussed antirealist position is the constructive empiricism proposed by van Fraassen [Van Fraassen, 1980], and he is quite clear about his acceptance of the semantic thesis: theories (or models) are interpreted literally in the sense of being considered as representations, while remaining agnostic about the existence of the referents. That is, when a model describes such and such entity to such and such mechanism, the semantic realist considers that the meaning of the model, what the model tells us about reality, is that such and such entities behave like it is portrayed, with no regard for the representation’s adequacy. Accepting the semantical thesis does not imply any commitment about the existence of the depicted entities. As van Fraassen himself puts it:

Not every philosophical position concerning science which insists on a literal construal of the language of science is a realist position. For this insistence relates not at all to our epistemic attitudes toward theories, nor to the aim we pursue in constructing theories, but only to the correct understanding of *what a theory says*. (The fundamentalist theist, the agnostic, and the atheist presumably agree with each other (though not with liberal theologians) in their understanding of the statement that God, or gods, or angels exist.) After deciding that the language of science must be literally understood, we can still say that there is no need to believe good theories to be true, nor to believe *ipso facto* that the entities they postulate are real. [Van Fraassen, 1980, pp. 11-12]

In his definition, Moscati includes approximations and idealisations. Why so? Interpreted literally, a model does not portray anything as an idealisation: it is the user of the model that plays the game of make-believe. It is true in the perfect gas model that molecules bounce off each other (**mm problème de cohérence ici**). It is only when used as a representation of real gases that this assumption acquires the status of an idealisation. In fact, Moscati's semantical thesis already incorporates an instrumentalist flavour, as it makes impossible any justification that goes beyond the fictional status of the content of the model. The meaning of the model is already defined in terms of approximation, and then there is no hope for verification. But how can we know that something is an approximation if no prior verification has been made (or, at least, if the modeller does not have an a priori idea of what he is modelling)?

Being a literal description is a necessary condition of being an adequate description, the latter question being the core of the epistemic thesis that states empirical adequacy and predictive power as signs of truth of the model qua description of a state of facts. In other words, empirical success and (at least partial) truth are related in some way. The critique of realism performed by Moscati attacks the possibility of justifying such a relation.

Let us push the fictional analogy a step forward, and let us try to rephrase Moscati's argument in this perspective. Models tell us stories about entities behaving in a variety of ways. Engaging in the modelling activity is akin to playing a game of make-believe: accepting idealised assumptions, deriving new fictional truths, exploring the world of the model. The semantic thesis states that the story told by the model is actually a story about real physical systems, that is, models are not instruments judged on their predictive power and empirical adequacy only, they tell stories about facts, and they are therefore capable of being true or false. The epistemic thesis states that empirical adequacy is a good guide to truth.

Moscati's definitions are unable to keep the two aspects of realism separate, and that leads him to jump too quickly to an antirealist conclusion. We hope that the fiction view has helped clarify those matters.

In the next section, we build an epistemological criterion of models selection able to take into account the central feature of the fiction view of models as summarized in the MC (see def. 1).

4 “Unrealistic” assumptions as substrates of modelling and classification principles

From the fiction view of modelling developed in section 3, it turns out that some basic assumptions does not have to be taken too literally, as really describing what is going on beyond direct observations, but rather as mere *substrates of modelling*. That is to say, as hypotheses which are necessary to take but which are not genuine empirical statements. Take a metro map as an illustrative example. A metro map can be seen as a model of the real metro network, a lot of features of which are highly idealized or even literally wrong, as the form of the lines between metro stations. However, if the purpose is to be able to move from a station to another one, it would be silly not to trust such a map for the reason that it contains idealisations or “false assumptions” about the form of the lines between stations. Our point is thus not that empirical false statements do not matter in general, but rather that these assumptions are simply not empirical statements. The form of the lines between stations has to be represented in some ways for the map to exist, but they are not claiming anything about the lines between stations as they really are out there. Indeed, this is simply not the purpose of the map and it is assumed that the form of the lines does not impinge on this purpose. Likewise metro lines have to be represented having a certain form (for instance, as straight lines), agents in rational choice, behavioral or heuristic models have to be represented computing information in some way in order to reproduce the observed behaviors. However, there is no reason for this representation to be realistic, i.e. consistent with what is otherwise known about human behavior, as long as the model reaches the purpose for which it is built.

We thus agree with Moscati on the point that any model (in general, but in particular in decision theory) is as-if, and that it is not a fundamental epistemological issue. As for the epistemological criteria of acceptability of a given model, here is the account that Moscati gives from his antirealist perspective [Moscati, 2023, p. 22]:

Insofar as a decision mechanism produces, i.e. explains, the targeted choice-behaviour phenomena in a simple and parcimonious way, and insofar as the decision theory incorporating the mechanism is successful [...] in describing, predicting or controlling the targeted choice-behaviour phenomena, the explanation provided by the mechanism is a valid one.

We obviously agree that classical epistemological criteria as simplicity, parsimony or predictive capacity are good criteria in general and in the case of mechanisms-based explanations in particular. However, we would like to go a bit further and build up a an epistemological criterion dealing explicitly with the issue of false assumptions in the modelling process and possibly realist-compatible. We now introduce this “structural invariance” criterion, as developed by one author in more details in [Attard]. Later on, it will be argued that this criterion is able to take into account all features of as-if modelling issues as described above while possibly being justified from a realist perspective.

The construction of the criterion starts with the following observation: the non triviality of an explanation of the form {fundamental principles + specific hypotheses}

as described above in equation (2) rests on the constraints which are imposed to the set of specific hypotheses used in a given empirical context. More precisely, if acting rationally only means maximizing a utility function, then given any behavior, it is always possible to find a utility function to maximize, under the form of *ad hoc* hypotheses. Besides, this is a classical criticism of the (careless) use of rationality in the social science, as expressed e.g by P. Ylikoski and P. Hedström in [Manzo, 2014, Chapter 2, pp. 59-60]:

From the point of view of the core assumptions of [rational choice theory (RCT)], broadening the range of individuals' concerns to non-monetary goods and to the welfare of others is straightforward, although there is concern about whether this can be done in a non-arbitrary manner. [...] Finding a RCT model that fits a particular phenomenon becomes almost trivially easy as there are no real constraints on preferences and beliefs that can be attributed to the individuals in question.

Thus, relevant epistemological criteria do not apply directly to the fundamental principles or the specific hypotheses *per se*, e.g. demanding their accuracy, realism or external consistency, but rather to the set of constraints which are imposed to the specific hypotheses in order to turn (2) into a non trivial explanation within a given set of fundamental principles. In physics, what makes a specific hypothesis (for example, $1/r^2$ -gravitational term in Newtonian physics) epistemologically acceptable is not that it is realistic or merely true (for reasons cited above), but rather that it is *systematically* associated to a certain set of empirical situations (here, some gravitational phenomena) with great empirical success (non trivial adequacy), and that in other formulations of classical mechanics it has a corresponding term which is also systematically associated to the same set of empirical situations with the same empirical success. Thus, the constraints imposed to the set of acceptable theoretical hypotheses ends up exhibiting a certain *structure*: a systematic association between classes of phenomena and classes of models which explain them. The final point is that even if general principles and theoretical hypotheses can be formulated in different ways, they provide isomorphic classifications of models and thus a classification of phenomena which does not depend on any particular formulation.

There is no obvious reason for this observation to be restricted to physics: on the contrary, we claim that it is a quite fundamental epistemological feature. As long as the logical pattern of explanation looks like (2), the question has to be addressed of how to render the explanation non trivial – more precisely, what constraints apply on the specific hypotheses used to instantiate the given set of fundamental principles. Meanwhile, any explanatory model often enjoys several equivalent formulations which are either incompatible to each other or not necessarily coherent with what is otherwise known. Our point is that relevant epistemological criteria apply to the set of constraints bearing on the specific hypotheses used in the explanatory model in the following manner: a set of fundamental principles derives its epistemological value from the extent to which it allows a salient (i.e. not trivial) classification of empirical facts within classes of models (corresponding to specific finite sets of theoretical hypotheses) which explain them,

such that this classification does not depend, at the end, on any particular formulation. Again, from our viewpoint we should not take fundamental principles or hypotheses too literally, as really describing what is going on beyond direct observations, but as a (necessary) representation of an underlying classification principle. “Realisticness” of these assumptions is thus not the relevant epistemological criterion at work: their role is only to concretely embody this classification, not to represent reality.

That is why basic assumptions in decision theory models do not need to be realistic, or consistent with what is known about human behavior. Taking rational choice models as an example, what really counts is not whether agents are really computing complicated utility functions in their mind, because this, from our viewpoint, is not an empirical question. Indeed, such basic assumptions in theoretical models are not, despite their form, genuine empirical statements but only a way of embodying a certain classification principle which is the actual empirical proposal to be tested. From our structural invariance perspective, what really matters is rather what types of utilities are taken into account in the optimization process, and whether these types of utilities are already used in other similar contexts – in order for an invariant structure to be eventually highlighted.

Thus, embracing a fiction view of models does not imply arbitrariness in modelling choices – which is of course acknowledged even by antirealists. We nevertheless constructed an epistemological criterion of model selection which explicitly takes into account the fact that models contain falsities or idealisations. In the next section, as announced, we argue that fictionalism does not necessarily rime with antirealism neither, and bring the discussion at a realism-related level. We also argue that the realist view developed then is fully compatible with the structural invariance criterion presented in the present section, and that both developments actually strengthen each other.

5 As-if modelling, justification and realism

The second part of the realist challenge stated in preceding sections consists in the construction of independent empirically based justification of the model.

More specifically, we consider a response to be realist if it accepts the semantical and epistemological theses. As already mentioned, Moscati accepts both for observables, but also denies both regarding unobservables, in particular for the mechanisms postulated by the models.

Of course, in a fictional perspective, being realist about these mechanisms and some (if not all) the entities in concerns and involves is a no-go: they are knowingly unrealistic. The effectiveness of the $PV = nRT$ equation is certainly not a proof that molecules are bouncing spheres **encore ici**, neither the utility of RCT is a proof that agents are non-stopping-utility-maximisers.

What, then, can be taken as the basis of justification of the model? We argue that the fiction view pulls towards a counterfactual view of justification and explanation. We address both in the remainder of this section and in the next one, respectively.

The fiction view portrays scientific models as set of propositions deriving from some

hypotheses and principles. Playing the game of *make-believe* is accepting idealised hypotheses in order to derive new fictional truths, with the hope the latter are applicable to the target system (i.e. the model supports inferences to the target). Let us call these propositions the agent using the model is willing to apply to the target system *model-world propositions*.

What is the nature of model-world propositions? From a formal point of view, they are conclusions of a deductive process starting from fictional hypotheses. In other words, model-world propositions are counterfactual propositions. The example from the ideal gas law speaks for itself. Any proposition about a real gas, which is supposed to be sufficiently similar to a perfect gas for the inference to be reliable, derives from a set of hypotheses in which we find at least one idealised assumption. When making a prediction about the pressure or the temperature of the modelled gas, the complete form of the proposition is something like “if the perfect gas law is true, then this gas will behave as such and such” In fact, the very application of the perfect gas law to a phenomenon carries with it the idealised assumption that lies behind. Model-world propositions then take the form of an implication where at least one of the antecedents is false, and that is one of the reasons why many epistemologists defend a modal view of modelling (see e.g. [Verreault-Julien, 2019]): the very counterfactual nature of model-world propositions can only produce possibility claims.

So far so good, but an important question stays unsolved: are model-world propositions immune to the fictional process of modelling? Indeed, the modelling conundrum we are interested in is still problematic. Even if we accept not to reject the possibility of justifying the model in a way that is compatible with the traditional theses of realism from the observation that its basic assumptions are knowingly false, the question is still begging for model-world propositions. In other words, if the only claims a scientific model is able to provide are possibility claims, then how can we hope to empirically justify them?

Coming back to the characteristics of the epistemological criterion developed in section 4, it appears that the focal point of justification must remain as far as possible from entities and mechanisms, as many different such entities and mechanisms may be instantiated in different and incompatible models of the same phenomena. On the other hand, the main feature of models is that they provide a way to manipulate hypotheses, to examine their empirical consequences, in a non trivial manner. Hypotheses and principles are arranged in a model in the form of a counterfactual structure: all fictional hypotheses are a priori acceptable, but they are subject to a constraint that is both epistemological and empirical. A model is adequate in the sense that it reproduces sufficiently well the counterfactual behaviour of the target system. The counterfactual structure the model embodies can then be taken as the ground of a truth-oriented justification: what makes the model adequate is that it properly captures the counterfactual behaviour of the target. In this context, fictions, idealisations and “as-if” modelling are no worry and have epistemic value because they highlight the counterfactual structure and make it possible to be manipulated.

Let us illustrate our point with a rational choice model. If such a model is empirically

successful, we agree with Moscati in that it does not commit us to the belief that agents are really computing and maximising complicated utility functions, and that it is not even necessary for the model to be acceptable. However, from the realist point of view developed above, the reason why this model is successful is that it captures the right counterfactual patterns responsible for the observed behaviors. The maximisation of a utility function is just a way of concretely formulating these counterfactual relationships, that stay the same independently of any formulation, and not something to take literally. This realist viewpoint is fully compatible with the epistemological criterion of structural invariance developed in section 4. According to the latter, what really matters is not the hypotheses themselves but rather the constraints which apply to them in order to render the explanatory pattern (2) non-trivial. More precisely, the epistemological value of fundamental principles is not given by their mere truth but by the extent to which they allow a non-trivial classification of a given set of phenomena within classes of models. From our realist viewpoint, this non-triviality is precisely a direct consequence of the fact that somehow we succeeded in grasping the right counterfactual patterns. Thus, both our epistemological criterion and our realist account of the empirical success of models strengthen each other, while being fully compatible with a fiction view of scientific modelling.

This kind of realist interpretation of the success of the as-if model then goes beyond mere instrumentalism or antirealism. Empirical adequacy of the model finds an explanation via an external justification: empirically adequate fictional models are capable of capturing the correct (or, “correct enough”) counterfactual patterns and that is what explains their efficiency. It is easy to see how this meets the needs of the three realist theses presented above: the model exemplifies and makes manipulable a structure of counterfactual dependences that correspond to real patterns that exist in the physical system. The empirical and predictive success of the model rests on the scientist’s capacity to implement the relevant patterns in his model.

6 Realism and explanation veritism

As already mentioned, Moscati’s antirealism goes beyond traditional instrumentalism because it makes room for the cognitive and epistemic function of explanation:

the version of antirealism I advance in this article emphasizes the role that mechanistic explanations play in decision theory and therefore goes beyond traditional instrumentalism [Moscati, 2023, p. 4]

Acknowledging the importance of explaining actually goes beyond instrumentalism because it sets a further goal to scientific inquiry that merely saving the phenomena:

However, I do not think that instrumentalism provides a satisfactory account of the as-if modelling practices of decision theory. The problem I see with instrumentalism is that it does not account for the important role that the epistemic goal of explanation plays in decision theory. Decision theorists do

not want only to save the phenomena, but also to explain them. [Moscatti, 2023, p. 21]

Moscatti adopts a mechanistic conception of explanation,

I think that the one [theory of scientific explanation] that best captures the epistemic attitude of decision theorists is the mechanistic account of explanation. According to this account, to explain a phenomenon is to identify the mechanism that, via the activities and interactions of its components, produces the phenomenon. The mechanism and its components can be partially, or even entirely, unobservable. [Moscatti, 2023, p. 21]

According to the conception of modelling we have used to frame Moscatti’s position, the mechanisms in question are of course fictional:

She may even deem, and explicitly acknowledge, that the posited mechanism and its components (such as the utility function, the preference relation or the heuristic rules), are only fictional constructs. Nonetheless, the decision theorist explains, describes or predicts the decision maker’s choices as if they were generated by the posited psychological mechanism at issue. [Moscatti, 2023, p. 2]

The main idea at play is that the mechanisms that explain in the models are only representations of the real mechanisms that produce the phenomena. Model-explanations then have a clear cognitive function (such as enabling understanding) dissociated from any requirement for truth.

There are two main categories of theories of explanations: ontological and epistemic (respectively OC and EC). According to the former, explanations are “full-bodied things” [Craver, 2013, p. 40], “objective features of the world” [Craver, 2007, p. 27]. It is the very stone that hits the window that explains why the glass is shattered. On the other hand, epistemic conceptions consider explanations to be arguments used by agents in order to achieve certain epistemic and cognitive goals. There are no explanations if no one is doing the explaining in EC, while explanations exist, according to OC theorists, independently of any epistemic activity performed by scientists.

Moscatti clearly falls under the epistemic umbrella: in his view, mechanisms are fictions, representations or stories that feature in arguments.¹ What remains unclear is the epistemic status of the mechanism. Moscatti seems to conflate the production of observed behaviour and explanation. But, if the mechanism is fictional, then it produces no observed behaviour: only fictional (i.e. inside the model) behaviour that may or may not be similar to what is observed:

Insofar as a decision mechanism produces, i.e. explains, the targeted choice-behaviour phenomena in a simple and parsimonious way, and insofar as the

¹Bokulich [Bokulich, 2018] has recently proposed what she calls the *eikonic view* of explanation, which is a proposition that encounters most of the requirements Moscatti emphasises.

decision theory incorporating the mechanism is successful – possibly for reasons that are different from the ones imagined by decision theorists – in describing, predicting or controlling the targeted choice-behaviour phenomena, the explanation provided by the mechanism is a valid one. [Moscati, 2023, p. 22]

It is hard to see what explanation adds to the traditional instrumentalist understanding of models: there are more epistemic goals than just saving the phenomena, but all these other goals reduce to the reproduction of observable phenomena by the model.

An interesting distinction introduced in [Bokulich, 2018] may be of great help, here. Bokulich distinguishes between accounts of explanation (claims about how explanations work) and conceptions of explanation (claims about what explanations are). While we agree with Moscati on the fiction view as an *account* of explanation, we disagree when it comes to the nature of explanations themselves, and on the antirealist consequences that allegedly follow. Explanations work via mechanisms embodied in idealised models. Does it mean that we must refrain ourselves from looking for an external justification of these explanations? Must we abandon the “truth is a necessary condition of explanation” requirement?

One of us ([Brandelet, in press]) has showed how the fiction view of models can cope with the issue of veritism² of explanation. Briefly put, as an *account* of explanation, the fiction view is compatible with both epistemic and ontic *conceptions*. One is then able to consider the counterfactual structure exemplified by the model to be the ground on which the explanation is constructed. Our proposition has the two following characteristics:

- *How model-explanations work*: by postulating mechanisms in idealised models that correctly reproduce the phenomena;
- *What are explanations*: the counterfactual structure of the phenomena which is reproduced by the model.

Neither of these limit explanations to their antirealist understanding, while keeping the epistemic virtues associated with “as-if” and fictional arguments. This permits us to respond to the worry of the end of section 8.5 of Moscati’s paper:

my objection to the realist would be that, by sticking to his strict notion of explanation as true explanation, he can account for little of what goes on in decision theory. [Moscati, 2023, p. 22]

We have shown that it is perfectly possible to account for the “as-if” use of models in decision theory while providing an understanding of model-based explanations that is truth-oriented.

²Veritism is the conception that considers truth as a necessary condition of explanation.

7 Conclusion

References

- Richard Arnowitt, Stanley Deser, and Charles W. Misner. Republication of: The dynamics of general relativity. *General Relativity and Gravitation*, 40(9):1997–2027, aug 2008. doi: 10.1007/s10714-008-0661-1.
- Jeremy Attard. Epistemological statuts of rationality principles in the social sciences: a structural invariance principle. URL <https://philpapers.org/rec/ATTES0>.
- Alisa Bokulich. Representing and explaining: The eikonic conception of scientific explanation. *Philosophy of Science*, 85(5):793–805, December 2018. doi: 10.1086/699693.
- Antoine Brandelet. Can fiction and veritism go hand in hand? *Zagadnienia Filozoficzne w Nauce (Philosophical Problems in Science)*, in press.
- Nancy Cartwright. *How the Laws of Physics Lie*. Oxford University Press, 1983. doi: 10.1093/0198247044.001.0001.
- Carl F. Craver. *Explaining the Brain*. Oxford University Press, USA, 2007.
- Carl F. Craver. The ontic account of scientific explanation. In *Explanation in the Special Sciences*, pages 27–52. Springer Netherlands, October 2013. doi: 10.1007/978-94-007-7563-3_2.
- Jürgen Ehlers. Survey of general relativity theory. In *Astrophysics and Space Science Library*, pages 1–125. Springer Netherlands, 1973. doi: 10.1007/978-94-010-2639-0_1.
- Milton Friedman. The methodology of positive economics. In Milton Friedman, editor, *Essays in Positive Economics*, pages 3–43. University of Chicago Press, 1953.
- Roman Frigg and James Nguyen. The fiction view of models reloaded. *The Monist*, 99(3):225–242, jul 2016. doi: 10.1093/monist/onw002.
- Roman Frigg and James Nguyen. *Modelling Nature: An Opinionated Introduction to Scientific Representation*. Springer International Publishing, September 2020. ISBN 3030451526.
- M. Göckeler and T. Schücker. *Differential Geometry, Gauge Theories, and Gravity*. Cambridge University Press, 2011. ISBN 9780511628818.
- Kirill Krasnov. *Formulations of General Relativity Gravity, Spinors and Differential Forms*. Cambridge University Press, 2020. ISBN 9781108674652.
- Insa Lawler. Scientific understanding and felicitous legitimate falsehoods. *Synthese*, 198(7):6859–6887, December 2019. doi: 10.1007/s11229-019-02495-0.

- Gianluca Manzo. *Analytical Sociology. Actions and Networks*. Wiley series in computational and quantitative social science., 2014.
- Ivan Moscatti. Behavioural and heuristic models are as-if models too – and that’s ok. *Economics and Philosophy*, pages 1–31, apr 2023. doi: 10.1017/s0266267123000093.
- Tamer Nawar. Veritism refuted? understanding, idealization, and the facts. *Synthese*, 198(5):4295–4313, October 2019. doi: 10.1007/s11229-019-02342-2.
- Christopher Pincock. A defense of truth as a necessary condition on scientific explanation. *Erkenntnis*, January 2021. doi: 10.1007/s10670-020-00371-9.
- Stathis Psillos. *Scientific realism : how science tracks truth*. Routledge, London New York, 1999.
- Daniel F. Styer, Miranda S. Balkin, Kathryn M. Becker, Matthew R. Burns, Christopher E. Dudley, Scott T. Forth, Jeremy S. Gaumer, Mark A. Kramer, David C. Oertel, Leonard H. Park, Marie T. Rinkoski, Clait T. Smith, and Timothy D. Wotherspoon. Nine formulations of quantum mechanics. *American Journal of Physics*, 70(3):288–297, mar 2002. doi: 10.1119/1.1445404.
- Emily Sullivan and Kareem Khalifa. Idealizations and understanding: Much ado about nothing? *Australasian Journal of Philosophy*, 97(4):673–689, January 2019. doi: 10.1080/00048402.2018.1564337.
- Frederick Suppe. Theory identity. In William H. Newton-Smith, editor, *A Companion to the Philosophy of Science*, pages 525–527. Wiley-Blackwell, 2000.
- Adam Toon. *Models as Make-Believe*. Palgrave Macmillan, January 2012.
- Hans Vaihinger. *The philosophy of ‘as if’*. International library of psychology, philosophy, and scientific method. Martino Pub, Mansfield Center, CT, nachdr. der ausg. new york : harcourt, 1925 edition, 2009. ISBN 157898825X. Includes bibliographical references and index.
- Bas C. Van Fraassen. *The Scientific Image*. Oxford University Press, October 1980. ISBN 0198244274.
- Philippe Verreault-Julien. How could models possibly provide how-possibly explanations? *Studies in History and Philosophy of Science Part A*, 73:22–33, February 2019. doi: 10.1016/j.shpsa.2018.06.008.