

Credibility, Idealisation, and Model Building: An Inferential Approach

Xavier de Donato Rodríguez · Jesús Zamora Bonilla

Received: 15 April 2008 / Accepted: 25 September 2008 / Published online: 8 January 2009
© Springer Science+Business Media B.V. 2009

Abstract In this article we defend the inferential view of scientific models and idealisation. Models are seen as “inferential prostheses” (instruments for surrogative reasoning) construed by means of an idealisation-concretisation process, which we essentially understand as a kind of counterfactual deformation procedure (also analysed in inferential terms). The value of scientific representation is understood in terms not only of the success of the inferential outcomes arrived at with its help, but also of the heuristic power of representation and their capacity to correct and improve our models. This provides us with an argument against Sugden’s account of credible models: the likelihood or realisticness (their “credibility”) is not always a good measure of their acceptability. As opposed to “credibility” we propose the notion of “enlightening”, which is the capacity of giving us understanding in the sense of an inferential ability.

1 Models as Inferential Tools

One of the most significant features of the philosophy of science in the last two decades is the relevance it has attached to *models* as the central building block of scientific research and scientific knowledge, a role that has, to some extent, obscured the prevalence of *theories* in the major approaches of the 20th century. This change of focus was mainly due to the stronger attention given by philosophers to how science is actually practised, for in many disciplines scientists spent most of

X. de Donato Rodríguez
UAM–I, Mexico City, Mexico
e-mail: xavier_donato@yahoo.com

J. Zamora Bonilla (✉)
Dpto. de Lógica, Historia y F. de la ciencia, UNED, Humanidades, Paseo de Senda del rey 7,
28040 Madrid, Spain
e-mail: jpzb@fsf.uned.es

their working time developing, checking, testing and comparing specific models, which is particularly true in the case of economics. One pressing philosophical question about scientific models has been that of *how can the scientist learn about the world with the help of such models*, or as it is sometimes put, how can we “learn from” them (cf. e.g. Morgan 1999; Knuutila 2005). This question has been related to the discussion on the nature of scientific *representations*, given that models are, among other things, intended representations of some aspects of the world. The current debate about the role of economic models, initiated in Robert Sugden’s paper on “Credible worlds” (Sugden 2002), could be interpreted in the light of this discussion. His main thesis was that models serve as the basis of a kind of inductive inference, on an equal (or at least similar) footing as the “real”, empirical cases from which the traditional view of induction assumed that this kind of inference had necessarily to start. Opposed to this view was another conception of economic models, according to which they are rather kinds of “thought experiments” that help us to isolate and identify the causal factors that really intervene in the world (e.g. Cartwright 1983; Mäki 1994). The main aim of our article is to show how an inferentialist conception of representation and idealisation can do justice to both approaches to economic (or, more generally, scientific) models. As another consequence of our approach we will offer a unified account of three aspects of the use of scientific models that may be difficult to harmonise: their role as representations, as explanations, and as instruments for learning and manipulation.

Models can be interpreted *as many and very different things* (maps, worlds, experiments, artefacts, social constructions, and so on), and most of these interpretations capture an important aspect of them without necessarily being in mutual contradiction. Obviously, since models do not constitute anything like a “natural class”, the choice of one of these aspects as the cornerstone of an approach to scientific models is, to some extent, subjective and can only be justified in terms of the light it helps to shed on the rest of the relevant aspects. If we take this for granted, our preferred interpretation is that *scientific models are basically a type of artefact* (cf. Knuutila 2005). In particular, they are *instruments for surrogate reasoning* (cf. Swoyer 1991; Hughes 1997). Ibarra and Mormann (2006), Contessa (2007), Bueno and Colyvan ([forthcoming](#)), most of them referring back to the ideas of Heinrich Hertz in *Die Prinzipien der Mechanik* in the late nineteenth century). Suárez (2004) offers what is perhaps the clearest and most synthetic statement of an inferential view of scientific representations. The basic intuition behind the concept of surrogate reasoning is slightly but significantly different from the already mentioned idea of “learning from models”: the latter assumes that models are in some sense a *source* of knowledge about the world (as we will show, this may sometimes be the case, but not necessarily every time), whereas according to the former, the main use of models is in helping us to draw inferences *from* the system they represent. Surrogate reasoning consists of the following three basic inferential steps (see also Fig. 1, adapted from Ibarra and Mormann (2006, p. 22) and Bueno and Colyvan ([forthcoming](#), p. 12):

- (1) An empirical system is *interpreted* in terms of the structure of the model (this is the step that Hughes calls “denotation”, and Bueno and Colyvan

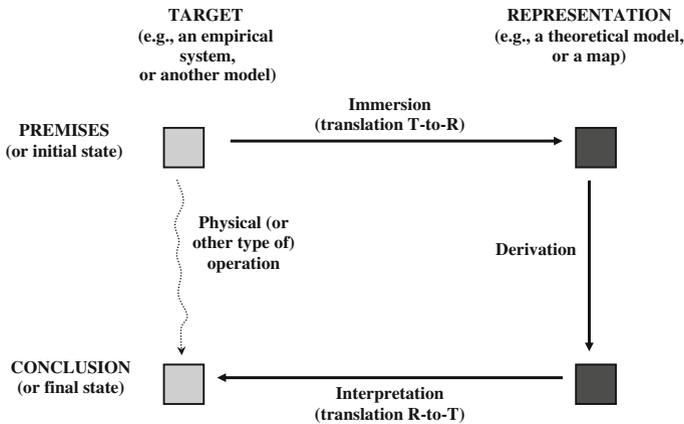


Fig. 1 Three steps in the making of inferences with the help of models

“immersion”); we make an *inference* from some propositions about the empirical system to propositions about the model. The structure of the latter is usually mathematical, but it may be materialised in any “format” that allows us to *reason* about it, such as in a physical model the movement of the components of which can be materially performed and described.

- (2) Starting from the statements derived from the interpretation of the empirical system, *formal inferences* are made “within” the model, taking advantage of its structure (Hughes calls this step “demonstration”, and Bueno and Colyvan “derivation”); depending of the case, these inferences may be deductive, inductive, statistical, abductive, counterfactual, analogical, and so on.
- (3) Some of the conclusions derived in this way are re-interpreted or *re-translated* into the language of the empirical system (both papers call this step “interpretation”); of course, this new interpretation is in itself a kind of *inference*, like those in the first step.

Thus, what the model has allowed us to do in the end is to derive some conclusions about the empirical system, starting from information extracted *from this same system*. Hertz’ original idea (particularly well represented in the theory developed by Ibarra and Mormann in their distinction between conceptual and physical operations) was that the *inferences* we draw within the model have to reproduce the *causal* connections between the real events occurring in the empirical system, but the inferential framework is general enough to include systems in which *it is not the causal structure* the model aims to reproduce. For example, maps function exactly in this way: one starts by translating one’s “real-world” starting point and destination into the map, then uses the map to look for the route one can take to reach one’s destination, and, lastly, transfers that information again to the “real world”, in the form of the direction in which one has to move (of course, as the trip goes on, many such inferential circles must be drawn). Hence, the inferential approach allows us to vindicate the old intuition that “models are maps” (or that maps are models), without committing us to

anything like a “pictorial” view of scientific representations (think of a GPS device transmitting only *verbal* information: it contains a *representation* of the terrain, but not a *pictorial* one).

Suárez (2004) dubbed the inferentialist view of scientific representations a “minimalist” one in the sense that it is not committed to any specific thesis about the *nature* of the relation of “being a representation of”. From the point of view defended here, since models are primarily *artefacts* (cf. Knuutila 2005), such a feature is hardly a surprise: artificial objects are usually identified not by their specific material constitution, but for their *function*. Consequently, *any entity (either real or abstract) that can sustain the working of conceptual inferences on the right-hand side of Fig. 1 can serve as a representation*. In contrast to what Suárez suggests, however, this does not mean that we cannot have a “general theory of scientific representations”. Even if we admit that perhaps we cannot draw up a finite list of necessary and sufficient conditions for something to instantiate the concept of “scientific model”, what follows from an inferentialist view such as the one depicted by Suárez and others is that the theory of scientific representations needs to be a *pragmatic* theory, i.e. a theory about what scientific models are *for*, how they are *used* and, most importantly, what the features are that allow scientists to distinguish *successful* models from the not so successful (cf. Contessa 2007). Suárez is thus right if we interpret him as claiming that a “merely logical” general theory of representations is not possible. In other words, what we need is not merely a theory about what scientific models *are*, but an explication of what it is that makes good scientific models *good*, i.e. an explication of how they are *evaluated* (which we offer in Sect. 2). This does not entail, however, that the *formal* study of the connection between models and their *representanda* is not important in the philosophy of science, for even if there are no necessary and sufficient *formal* conditions for *being* a model, or for being a *right* model, most scientific models fulfil their inferential role *thanks precisely to their having some specific mathematical properties*. We will argue, in particular, that many scientific models are construed by means of an idealisation-concretisation process, and that this process can be analysed from an inferentialist perspective (this will be done in Sect. 3). Idealisation allows us to create more simplified representations and, conversely, concretisation allows us to correct our original idealised models in order to improve them and make them more reliable. Idealisations are also on many levels, from those of a hypothetical character (because we are totally ignorant about their capacity to be realised) to those that are known to be not only empirically false, but also to contradict some theoretical or background assumptions, or even internally inconsistent (and successful in spite of this). On the other hand, the value of a scientific representation is understood in terms not only of the success of the inferential *outcomes* arrived at with its help, but also of the *heuristic* power of representations and their capacity to correct and improve our models, i.e. in terms of how the inferential steps, and their connections with other inferential artefacts, can be *improved* thanks to the properties of a model. Finally, in Sect. 4 we relate our inferential conception of scientific models to the discussion on the role of models as credible worlds or isolating tools.

2 Scientific Models and Inferentialism

An appropriate way to progress from an inferential *definition* of scientific models and representations to a pragmatic *theory* about the use and evaluation of these cognitive tools is by connecting the view sketched in the previous section to a more general inferentialist account of knowledge and action, such as the one proposed by the philosopher of language Robert Brandom (1994). This theory has been applied in some previous papers (Zamora Bonilla 2006a, b) to the process by which a scientific claim becomes *accepted* among the relevant research community; we will use it in the following rather to illuminate the process of model building and testing. Brandom describes his theory as “inferential semantics grounded on a normative pragmatics”. The “inferential semantics” part means that instead of understanding the *meanings* of linguistic expressions as founded on the *representation* of aspects of the world by pieces of speech or text, we primarily explain them through the *inferential links* that each expression has with the others, and it is this inferential entanglement of expressions that allows them to serve as representations. The “normative pragmatics” part refers to the idea that human *actions* (not only verbal ones) are understood as structured and motivated by the *deontic statuses* of the agents in terms of what they are committed to or entitled to do. Lastly, the fact that the semantic part is “grounded” on the pragmatic part means that it is not the semantic properties of language that determine what people can do with it, but the other way around. The essential elements of Brandom’s inferentialist model that may be helpful now are the following (see Fig. 2):

- (1) At any given moment, each agent is *committed* (or entitled) to certain claims (“doxastic commitments”) and to certain actions (“practical commitments”).
- (2) These sets of commitments evolve according to the *inferential norms* that are (tacitly or explicitly) accepted within the relevant community of speakers. The

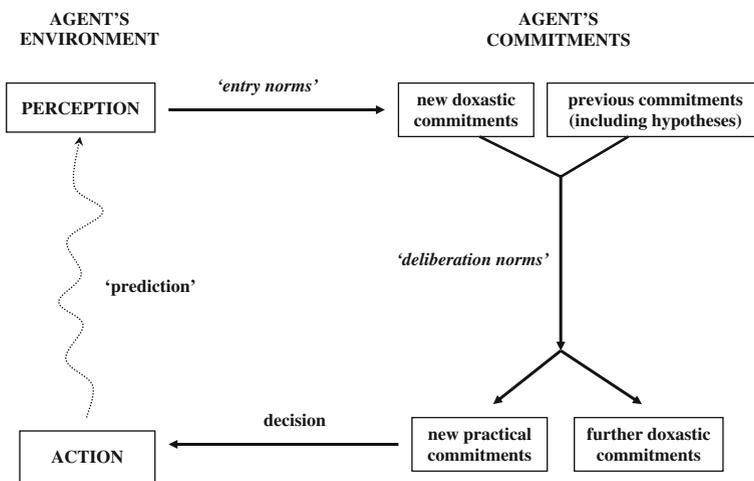


Fig. 2 A normative inferentialist view of rational action

norms indicate to what claims and actions one *becomes* committed or entitled according to what commitments one previously had, or according to the realisation of certain public events.¹

Although in our analysis of scientific models and representations we will concentrate on doxastic commitments (the pragmatic commitments we take into account mostly relate to intervention and prediction), this is not to obscure the fact that the *value* of a given set of commitments, and of a system of inferential norms, depends essentially on the efficiency with which the actions such commitments and norms lead to help to satisfy the agent's *goals*. Since the flux depicted in Fig. 2 is dynamic (new actions lead to new events and to interactions with other agents, which cause new perceptions, and so on), the commitments and norms of an agent or a community will *evolve*: commitments change according to what new inferences are made, and inferential norms may also change for there may be a "natural selection" process that allows successful norms to "survive and reproduce" (i.e. passed more frequently to other agents and situations) at a higher rate than less successful ones (cf. Zamora Bonilla 2006a, b). Thus, "good" scientific models will be those that, in the loop described in Fig. 1, lead to *satisfactory* results, or to better results than other models give.

What can this show us about scientific models? Within the scheme of Fig. 2, a model primarily consists of a set of interconnected doxastic commitments and inferential norms, some of the latter made explicit by *formal rules* and others taking the form of *practices* (such as simulation techniques; cf. Winsberg 2006). The main question to put in order to elaborate on a pragmatic theory of scientific models would thus concern what the *most general functions* of such "inferential modules" within the economy of practices depicted in Fig. 2 could be. We suggest that these functions could be reduced to the following three:

- (a) First, the addition of a model to the corpus of our commitments should increase the ratio of *successful* inferences (i.e. "right predictions") to not so successful ones: a glance at Fig. 2 shows that, apart from the relation that success has with the practical goals of our actions, all possible *epistemic* criteria of success depend in some way or other *on the coherence between commitments coming from different inferential routes*. This is not an easy goal. Due in part to previous commitments coming from independent sources, and in part to the use of fallible inferential norms (e.g. induction, abduction, inference to the best explanation), there is no guarantee that the working of an agent's inferential machinery always leads to conclusions that are logically consistent between them.
- (b) Secondly, the model should also *increase the number and variety of inferences* we were able to draw from the rest of our commitments. This does not only mean that the addition of the model should have new *logical* consequences: in some cases it is also important that it allows us to *actually* reach the logical

¹ It is also important to note that inferential norms include not only such licensing logical, formally valid inferences, but also *material* inferences (e.g., from "it's raining" to "the floor will be wet"). The inferential role Grüne-Yanoff (2009, Sect. 2) attached to the *interpretational* component of models would refer precisely to such material inferences.

consequences our previous commitments *had* but that we were not capable of drawing *in practice*. Mathematical and computer models are usually of this kind, but so are wood-and-wire models, for it is by manipulating their parts that we can often *see* what follows from some objects being in a particular causal or geometrical disposition.²

- (c) Thirdly, it should help us to reduce the cognitive (or computational) *costs* of the activity of drawing consequences, either doxastic or practical.

We describe some of the virtues models may have according to these functions in more detail in Sect. 4, whereas in Sect. 3 we consider some of the ways in which models may be constructed in order to fulfil these goals. What we would like to discuss now is the connection between these general functions of models and the general topic of this monograph: what is it that makes a model “credible”? Our thesis is that it is mainly to do with the first function indicated above: do *any* other criteria exist by which to judge whether a model is *right* (or “probably true”, or “approximately true”, or “probably approximately true”) besides the fact that it leads us to adopt conclusions that are corroborated by different means? *We think they do not*. If we do not know in advance whether the hypotheses of a model are true or not, then the coherence of its predictions with other commitments we arrive at in an independent way is a powerful argument to conclude that such assumptions are probably right (this is anything but the good, old-fashioned hypothetico-deductive method). What happens, then, when a model contains assumptions (say, “virtual commitments”) we know to be false, or very far indeed from the truth? Traditionally, it has been asserted that if the model makes many right predictions we will probably be led to the conclusion that, in spite of being false in some aspects, it nevertheless depicts more or less well *the aspects of the world that are responsible for the truth of its predictions*, and hence the false assumptions will be taken as “isolating” assumptions. We find Michael Strevens’ (forthcoming) way of putting it particularly clear: the role of these assumptions is often that of showing which aspects of the system *do not actually* play a causal role. We think this could also be related to what Nancy Cartwright (2009) expresses as “probing models as a means to understand how structure affects the outcomes”. All this has something to do with function *a* in the list above. The credibility of a model would consist, then, simply in a “measure” or estimation of how probably it is right, which may mean “right in counterfactual situations”. From an inferentialist point of view, however, false assumptions such as simplifications, idealisations and the like also serve some other functions: in some cases they just allow us to *draw* conclusions where we could not do so before (function *b*), i.e. they show us *how to reason* in certain circumstances (for example, how to apply economic reasoning in cases in which agents have different information). This is a kind of *heuristic* role, often referred to as “finger exercises”, and “thought experiments” could also be included within this category. Lastly, some assumptions may also have the role of making calculations easier, or even possible (e.g. approximation methods, statistical assumptions), which concerns functions *b* and *c*. Our thesis is that when a scientific model satisfies these goals

² In this sense, mathematical models also allow us to learn about (mathematical) theories, in the sense of helping to *actually* derive consequences from the axioms of the latter (cf. Morgan and Morrison 1999).

(allowing us to make more claims and at a lower cognitive cost), it is not appropriate to call it “credible”, for it may be very far removed from any actual or even possible circumstance. We prefer to say that a model like this is *enlightening*, i.e. its main virtue is to give us “understanding”, not in the sense of a kind of “mental feeling”, but as an *inferential ability* (cf. Ylikoski 2008). We will come back to this question, and to the related topic of the explanatory power of models, in Sect. 4.

We could also express our view as follows: the complex set of factual claims and inferential dispositions constituting the body of commitments of a scientific community has the *potential* of generating a lot of consequences, but the particular connections between many of those claims and dispositions may make it very difficult, or even impossible, to *actually* derive these consequences by using the calculating and reasoning techniques the community has; in particular, our factual claims cannot be “appropriate” as premises to extract from them (and with the help of our inferential norms) too many relevant conclusions. In these cases, the introduction of a definite set of “virtual” commitments, depicting a counterfactual situation (that perhaps is even incompatible with some accepted principles), may act as a “canal system”, or a “pruning” that channels the inferential flow from our inferential dispositions towards more interesting and numerous logical consequences. This justifies our thesis that *idealisation by counterfactual reasoning* (i.e. searching for what *would* happen under certain assumptions) is the main tool in model construction, as we will show in the next section. Let us first summarise the properties that constitute a model’s *virtues*, according to the inferential approach we have just sketched; recall that a model is identified with a particular set of “commitments” (some given as premises, and others as conclusions):

- Its “size” (the number of questions it answers)
- Its coherence (especially between the commitments derived from others and those derived from perception)
- Its manageability (models as “inferential prostheses” allowing the drawing of many consequences at a low cognitive cost)
- Its heuristic capacity to produce new and more reliable models.

Further, what makes a *good* model? Here there is a list of possible virtues:

- *adequacy*: the consequences we obtain with its help must correspond to the consequences that follow from the real system (or, more exactly, from other inferential channels we are committed to);
- *versatility*: the model allows us to extract inferences from different kinds of claims about the *representandum*;
- *ergonomics*: the model must be easy to use (manageable), at least easier than extracting inferences directly from the real system, or from the bare theoretical principles;
- *compatibility*: the model must be easy to connect, if necessary, with other models (the same *representandum* may have more than one model for different cases, and these different models may be connected in various ways; a model could be taken as the *representandum* of other models; it could be based on

different or even incompatible sets of theoretical principles and counterfactual idealizations, and must make them compatible enough to allow inference making).

Analogously, what makes idealisations good are the manageability they induce, our measurements and calculations, and the heuristic power of guiding us in the formulation of abstract models with the help of which we are able to explain a great variety of phenomena. As we have shown in Sect. 3, if an idealisation were too easily made concrete, its heuristic power would be small. If idealisations impose highly ideal, unfeasible conditions, but have the ability to suggest appropriate concretisations, then their heuristic power would be great, and therefore they would have the chance of being successful in the production of more adequate theoretical and empirical models. It is also important to note that different disciplines may attach a higher value to some of these reasons than to others, reflecting the relative difficulty or benefits associated with each in the corresponding field (cf. Zamora Bonilla (1999) for such a comparison between the natural and the social sciences).

What picture of scientific knowledge derives from this view? In a sense, it could be considered a rather “relativist” one. After all, there is no recipe for the best combination of virtues models must have. Each scientist or scientific community will have its own *preferences* (and it is thus, in part, how communities and schools are identified), preferences that will be materialised in the *inferential norms* each group has, i.e. in the patterns establishing what inferences are taken as (more or less) *appropriate*. Furthermore, in our view, models are kinds of artefacts, or *inferential prostheses*. Our whole system of knowledge is mediated through an inferential network and there is, in fact, no way of distinguishing between inferential knowledge obtained “with the help of prostheses” and that obtained without them. Nevertheless, there are some prostheses that seem more “natural” than others, for they provide us with the *feeling* that we are “nearer to reality” with them than without them. In this sense the “degree of realism” of a model or theory must be taken as the *feeling* it gives us of “being closer to the truth” (basically, if its predictions correspond more closely to our observations and its assumptions are ones we are strongly committed to, for different reasons)³; the more, and in the more contexts, the *representandum* is replaceable by the *representans* in practice, the more “realistic” will the latter seem—this is particularly evident in the case of simulations. All this shows that our view is not as relativistic as it might seem: it is one thing to assert that the criteria for judging the goodness of models are “subjective”, and a very different thing to claim that the degree to which a model *actually satisfies* those criteria is “non-objective”. Even if the “realisticness” (or “credibility”) of a model depends on what an agent wants it for, it will be a matter of fact whether it fulfils these goals or not. Furthermore, models may help us to discover new aspects of reality, for they may be useful for testing the assumptions

³ This is in line with the “methodological” approach to the concept of verisimilitude that has been defended elsewhere by one of us, in which “truthlikeness” is taken as the “epistemic utility function” of scientists and, being a kind of *utility*, it is assumed to be something that can be *subjectively experienced* by the relevant agents. Epistemically speaking, truthlikeness (or the *appearance* of being closer to the truth), is hence, a more primitive concept than truth (Cf. Zamora Bonilla 1992, 2000).

they contain as hypotheses: scientific discovery could be seen as a way of introducing hypotheses that are so successful that they lead to particularly “realistic” new models.

3 The Role of Idealisation and Concretisation in Model Construction

In this section, we give a brief account of the role of idealisations and concretisations in the process of model construction from an inferentialist perspective. One of the main functions of idealising in model building is not to represent reality such as it is, but to allow inferences and interventions (i.e. practical inferences) about certain relevant aspects of the world. According to our account, idealisations could be understood as counterfactual inferential moves allowing more manageability in measurements and calculations, as well as in arriving at causal hypotheses in relation to the isolability of systems (these hypotheses will, in turn, allow us to make inferences about the role certain causal factors play in certain phenomena).

As has been stated (see in particular Morgan and Morrison 1999), scientific models are often better conceived of as autonomous and independent of theory. This means that there is no algorithm for constructing adequate models just from theoretical principles. Sometimes they are basically inspired or guided by principles, and sometimes they are abducted (at least in part) from empirical data and experimental sources, but the process of model construction usually brings together elements from different sources including calculation techniques, background knowledge, analogy and intuition, so that they could be seen as something autonomous and independent. In Boumans’ words: “model building is like baking a cake without a recipe. The ingredients are theoretical ideas, policy views, mathematisations of the cycle, metaphors and empirical facts” (Boumans 1999, p. 67). Idealisations play a crucial role in this process because the application of models to reality clearly involves taking into account many simplifications, and the abstraction and selection of relevant parameters. It is a commonly accepted view to understand abstraction as the deliberate omission of certain parameters, whereas idealisation could be viewed as the conscious misrepresentation of certain factors (cf. Jones 2005).

Abstraction, then, concerns the selection of parameters and the isolation of systems. In some sense it is a kind of idealisation, too (in the more abstract sense of “counterfactual deformation”), because we do know that there is no real system isolated from the rest of the world. The number of variables exerting some influence on a particular phenomenon and their measurement is too great to allow taking them all into account and thus offering analytical solutions to our problem. This is one of the main reasons for idealising: to simplify our model in order to render it computable. When, on the other hand, the data are too sparse, simulation methods begin to assume importance. The construction of simulation models in order to replace experiments and observations as sources of empirical data also involves many idealisations, or as Winsberg (2006) calls them, “falsifications”, in the sense of contrary-to-fact modelling principles that are involved in the process of model

building. His case study, on artificial viscosity, constitutes a good example of a technique for the construction of simulation models in computational fluid mechanics. It seems that practical advantage and its helpfulness in terms of constructing manageable models are what give “credibility” to this particular technique. According to Winsberg, the same could be said of other simulation models, but our conjecture is that this is not exclusive to simulation techniques, and also applies to other and more standard model-building procedures.

Not only models, but also laws and theories, as well as calculation techniques, measurement methods, graphs, explanations and practically every aspect of scientific method, involve idealisation. Of course, none of these exclusively include idealisations: the curves drawn among the data, for example, combine the results of approximation and optimisation, but in a sense we can tell that what justifies the use of either technique is the *counterfactual* assumption that the more a function approaches the data, the more probable it is that the inferences made by using that function will be right. Another important example is constituted by the ideal and counterfactual assumptions involved in approximations and reductive explanations between different theories and laws. We therefore have idealisations on the following levels:

- The abstraction and selection of parameters (isolation)
- Deformations introduced in the parameters considered in the model
- Idealisations made during the process of calculating and measuring these parameters and in the construction of data models
- Idealisations involved in the simplified form of laws and principles
- Idealisations needed in approximation relations between laws and theories
- Idealisations that are taken into account in the elaboration of computer-model simulations.

This list, of course, is not exhaustive. In real science these different levels of idealisation occur together and are usually found in diverse combined forms. Each of them deserves a particular study, having its own mechanisms of application, although a good question is whether they exhibit a common structure, such as a counterfactual deformation procedure that could be formalised in terms of modal logic or through the use of any other formal tool. Nevertheless, it would seem important to distinguish between the different concepts involved in the above classification: abstraction, idealisation, and approximation.

As we show in Fig. 3, idealisations (together with optimisation and approximation) play a role in constructing data models from the empirical data as well as theoretical models that purport to explain/predict some aspects of the real world, idealised in turn in the form of data models. An important function of idealisations is to allow the construction of more idealised models from the more concretised ones, or vice versa: in other words, idealisation/concretisation could refer to relations between theoretical models as well (as in the Kepler–Newton relation or the case of thermodynamics and statistical mechanics). For some purposes and in certain contexts to be determined by the scientific community, it is convenient to treat a highly idealised system as if it were real, or to treat a non-isolated system as if it were really isolated. In the same vein, we could work with more idealised

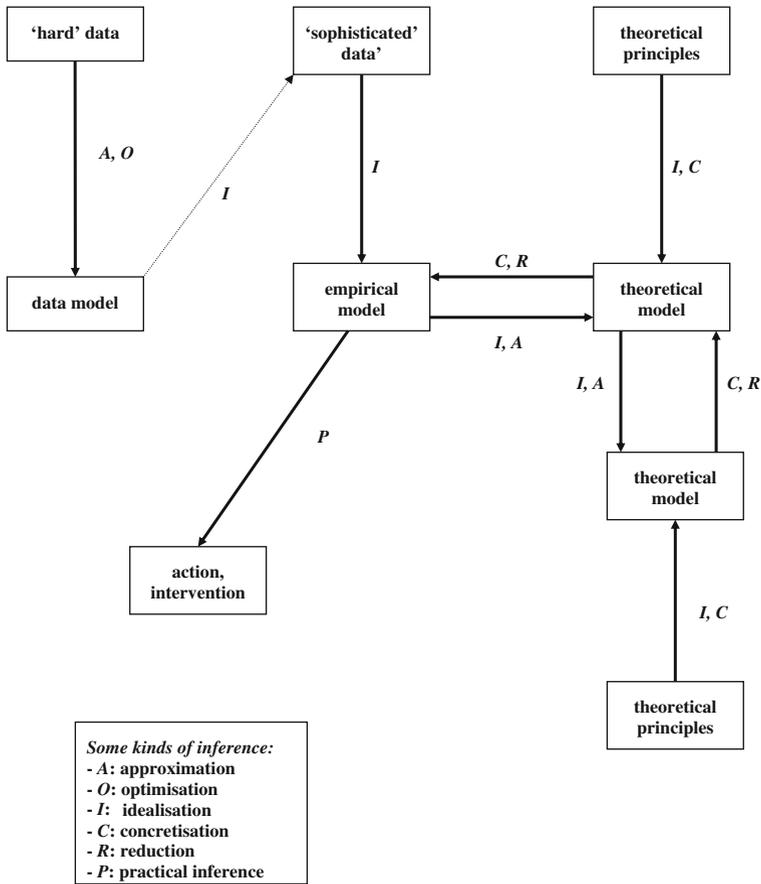


Fig. 3 Different types of inferential moves in the construction and testing of models

models instead of our more concretised ones, which perhaps are not useful for certain purposes or are not so easily workable. Analogously, data models could also be idealised or concretised for convenience. Figure 3 summarises these relations (cf. Arroyo and de Donato, Idealization, data, and theoretical constructs: The case study of phylogenetics, unpublished manuscript).

Data are always constructs, not only because of the well-known distinction between hard data and phenomena (cf. Suppe 1989; Woodward 1989; see also Laymon 1982; Ibarra and Mormann 1998), but also because the selection of all kinds of data or relevant parameters is always made as a consequence of some idealisation. The relevant parameters change depending on the level of idealisation at which we are. What we have, then, is more a hierarchy of phenomena interspersed by the process of idealisation-concretisation induced by theoretical means. As Suppe correctly put it, “Theories are not concerned primarily with applying laws to phenomena, but rather with using laws to predict and explain the behaviour of physical systems abstracted from phenomena” (Suppe 1989, p. 68).

Suppe considers “hard data” idealised descriptions of phenomena, which are more or less accurate depending on their degree of idealisation. What we do, in fact, is make more precise our descriptions of the facts and increase the predictive and explanatory power of our data models. We select some relevant parameters (abstracting from the rest) and idealise them. The data determination consequently becomes a complex process of elaboration from the phenomena, involving a great number of theoretical principles and assumptions related to parameter selection, their measurement, and the choice of boundary conditions. At the same time, we could even re-describe the data for theoretical purposes and revise them if something goes wrong during the process.

All these deformation procedures have a common structure. The idea is that by imagining counterfactual conditions we are departing more or less from a fixed world. As we have already stated, this idea resembles Lewis’ notion of distance between possible worlds (Lewis 1973), in which the actual world is fixed by the context (and may, in fact, be another idealised model). In a broader philosophical discussion, Nozick (2001, pp. 148–155) highlighted the importance of distinguishing between different “degrees of contingency”, meaning that statements that are contingently false may be true in possible worlds that differ from our actual world in very different ways.⁴ We will use the same term, “degree of contingency”, to mean that idealisations may differ in many different ways from the “actual world”. In this sense, our approach supports Mäki’s “isolationist” understanding of models.

Let us now briefly sketch our approach.⁵ Idealisations could be analysed as statements (S_i) that are the consequent of certain counterfactual (or subjunctive) conditionals, in which the antecedent expresses the ideal (or virtual) conditions (C_i) under which the idealisation holds. We could use the term “idealised law” for the whole conditional, and for the antecedent we could just use the term “ideal conditions”. The structure would then be: $C_1, \dots, C_n \Rightarrow S_1, \dots, S_k$, where C_1, \dots, C_n are the ideal conditions, S_1, \dots, S_k are the idealisations, and the connective “ \Rightarrow ” could be understood in terms of Lewis’ semantics for counterfactual conditionals.

Our basic intuition is that ideal conditions may have different degrees of contingency, or may be contingent in relation to different aspects:

- (i) At the highest degree of contingency, C_1, \dots, C_n are completely idealised in the sense that they contradict some accepted theoretical *principles*.
- (ii) C_1, \dots, C_n are contingently false but conflict with well-established empirical *regularities*.
- (iii) C_1, \dots, C_n are also contingently false but do not conflict explicitly with a well-established regularity. In any case, we have strong reasons to believe that they are false in the actual world and can only be approximately met under experimental control.

⁴ The intuitive idea is that a statement holding in $\frac{1}{4}$ of all possible worlds will have a degree of contingency of $\frac{3}{4}$. Later he interprets this in terms similar to those of David Lewis. The degree of contingency of a statement S in the actual world is the maximum degree of closeness of the worlds in which S does not hold with respect to the actual world.

⁵ Adapted from Arroyo and de Donato, The structure of idealization in biological theories, unpublished manuscript.

- (iv) C_1, \dots, C_n are purely contingent assumptions that, despite not seeming plausible, we do not even know if they are true or false in the actual world.

Again, the present classification is not exhaustive, and only purports to illustrate what we understand of idealisation.⁶ It should also be remembered that these different types of idealisation normally occur in different combinations, so that the structure of scientific theories may exhibit a very complex network of “multi-level” idealisations. The main idea is that a theory (or a law or a model) is formed of idealisations that hold under conditions ranging from (i) to (iv). This is needed if we would like a theory (or law or model) with heuristic, epistemic and cognitive virtues: if it were too idealised in the sense that it only contained ideal conditions of type (i), then it would be practically impossible to have concretisations because it would have no realistic connection to the actual world. However, theories have their great explanatory power (and other epistemic virtues) precisely due to these highly idealised conditions, for these assumptions are what make the working of their inferential machinery possible. On the other hand, if the ideal conditions involved were very likely to occur in the actual world, i.e. if they were only ideal cases of type (iv), then the theory would not be fruitful enough because once it had been concretised it would not have anything else “to say”. From an inferential point of view, it would not be fruitful as an inferential tool.

Weisberg (2007) referred to the different (and incompatible, in the sense of being not achievable together) representational goals involved in idealisation. Different forms of idealising are frequently based on different representational ideals. Galilean idealisation, which is the kind involved in thought experiments, the isolation of systems, and approximation and optimisation techniques, is used in order to achieve simplicity and manageability in the process of calculation and measurement, whereas idealisation as a method for seeking the main causal factors that have an influence on a certain phenomenon is more focused on discovering the real causal structure behind a class of phenomena. In this case, idealisations allow us to infer general causal hypotheses, which will in turn allow us to infer causal explanations of concrete phenomena. Concretisations are thought to achieve completeness and accuracy in the description of real systems. They allow us to make more accurate and precise inferences, which if successful will give to the theory a particular degree of enlightenment.

4 Are Good Scientific Models Representations of Credible Worlds?

As noted above, according to Sugden (2002), the main role of idealised models is to allow a particular kind of *inductive* inference about real systems (i.e. at least some *conclusions* of those inferences are about the real world): in a similar way to how we make inferences from some observed *real* cases to other unobserved real cases, or from some individuals to others (if the latter are analogous or similar “enough”

⁶ This may be related to Mäki’s idea (this issue) that isolation could be seen as a linear process, at one end of which is a real system, and at the other an abstract system; what we are adding to this view is that the different steps of the isolation process involve different kinds of *modalities*.

to the former; e.g. inferences about molecular genetic mechanisms in all animals, starting from the study of a few species), we might also infer something about a real system from what we learn from an *imaginary* system if this is *analogous or similar* in a relevant way to the real case. This similarity is surely not only to do with the relation between the *formal* structures of the real and the imaginary systems, but arises mainly from the *causal* factors present in both cases (although, after all, what we compare is the mathematical or logical structure of those causal factors). This could be connected to other approaches to explanation and idealisation, such as those of Uskali Mäki, Michael Strevens and Frank Hindriks: according to these views, idealised models are explanatory because they make (approximately) *true* statements about isolated causal factors. An idealised and in this sense “unrealistic” model can explain some real phenomena if it just describes the way in which the real causal factors operate. If the suggestions put forward by Mäki, Strevens and Hindriks are right, then so is Sugden’s, for an idealised model would be a system in which we “observe” the operation of some causal mechanisms, and we can draw empirically valid conclusions from this “observation”. As in the case of a controlled experiment, in the idealised model we “observe” (usually, more clearly than in the real world) the working of the mechanisms we wish to study. Thus, Sugden’s view of models can be resurrected in the thesis that economic models are *thought experiments*, in spite of his own claims against equating the two (cf. Sugden’s paper in 2009). We thus tend to support Mäki’s claim that the “isolationist” and the “inductivist” views of models are not so different. Our reason is that what allows the making of an *inductive* inference from the model to the real system is *only* the belief that the model rightly includes aspects that, in the real world, produce the phenomenon our inferences are about.

Sugden’s approach indicates a very important aspect of the use of models. However, we think that his use of the concept of “credibility” is confusing, and mixes the two different elements of the value of a hypothesis we have identified so far: idealised models are not simply “isolated”, but often describe situations that we know perfectly well *cannot* exist. Sugden’s favourite examples (Akerlof’s “market for lemons” and Schelling’s “checkerboard segregation model”) are cases that might perhaps be approximated in a real situation, but many other successful models have different “degrees of contingency” (as we called it in Sect. 3). This may be due to the existence of “infinity assumptions”, such as in economic models assuming that information is perfect and quantities infinitely divisible, or in physical models assuming infinite planes. On the other hand, models may describe impossible situations because they describe them by means of mutually contradictory theories (as with many solid-state models). In all these cases, we know for sure that the modelled system *could not* have been real, and so we cannot “believe” that it is. The model is completely *incredible* (cf. Grüne-Yanoff 2009). However, even in this case it may be extremely *enlightening* (with the type of enlightenment that one could also derive from literary *fiction*s, for example; cf. again Grüne-Yanoff’s classification of types of credibility).

The difference between credibility and enlightenment relates to the distinction between the two goals we may try to fulfil when adding a new claim to our set of commitments. As we have shown, on the one hand the new claim *H* (i.e. the set of

claims of which the description of the model consists, together with the propositions claiming that it represents some aspect of the world, or is to some extent applicable to it) will be more or less valuable for the same reasons why we have to think H (or the relevant part of it) is true or approximately true, which depends, in turn, mainly on two factors: the number and quality of right “predictions” H allows, and whether the accepted inferential norms allow H to be derived from other commitments. Models that are valued mainly according to this criterion are intended as a form of *discovery*: what we want from them is to know *how the world is*, what structures, forces, mechanisms and entities exist that are not manifest in what we know up to now. Of course, these models attempt to explain some facts, but *that* they successfully explain them is just one of the *reasons* why we are promoting our claim that we have *discovered* something real in the world.

On the other hand, the new claim H may be valuable because of the easiness and fluency it induces in our capacity to *navigate* the network of inferential links according to which our box of commitments is structured. Basically, the value of a model may derive from the new reasoning strategies it allows to be implemented, i.e. in the way it combines old inferential norms in order to create new ones.⁷ It can do this by different means: for example, it could amount to the discovery of a new algorithm, it may provide a new heuristic suggesting how to connect several collections of independent data, or it could simply allow the derivation of a known anomalous fact from previously accepted claims. For example, Akerlof’s “market for lemons” story succeeded particularly in showing *how to apply rational-choice modelling* to many market situations that had not been taken into account in neoclassical theories (i.e. those with information asymmetry); the point of his theory was not to claim that there were (or could be) real cases in which the very simplified conditions defining his toy examples, or something very similar, were met, but to show (in his own words, as a “finger exercise”) the direction of a research programme capable of delivering much more detailed and predictively successful models.

Therefore, we argue that the likelihood or realisticness of models (their “credibility”) is not in itself a good measure of their acceptability. According to Sugden, idealised systems are those that could be real, defined only by low-level contrary-to-fact conditionals, a kind of *thought experiment*, which in principle could be realised, but in practice cannot for practical or fundamental reasons. In this sense, Sugden seems to follow Atkinson’s approach to thought experiments (see Atkinson 2003). According to the latter, “thought experiments [...] are of value only when they are related to or inspire real scientific experiments” (2003, p. 209). Atkinson claims that this is at least the case in physics. Even this is highly debatable, however: there are frequently idealisations (idealised models) in physics that are defined by ideal conditions and are *impossible* to realise (this is one of the reasons why certain kinds of thought experiments are criticised as effective forms of

⁷ If inferential moves are viewed as a kind of cognitive operation of the mind, it is possible to distinguish two different systems of reasoning (cf. Carruthers 2006, p. 254 ff): one operates in parallel, usually in a quick, automatic and unconscious way, and is constituted of innate mechanisms, and the other supervenes on the operation of the first, usually through conscious inner speech, and depends on culturally transmitted combinations of simpler inferential norms.

reasoning in science). The assumption of counterfactual conditions, in other words the counterfactual deformation of real situations, seems to be essential to the structure of thought experiments, as shown in Sorensen (1992), having the purpose of excluding spurious possibilities and also of showing that others are perfectly genuine. If this is the usual case in physics, it seems to be true for stronger reasons in economics. “Credibility” in Sugden’s sense cannot then be the measure of the acceptability of models as “good models”. Unrealistic models are valuable, on the other hand, because (and when) they show us how to fruitfully apply to new cases the theoretical principles and inferential norms we knew from before, but were unable to use in those cases. Good unrealistic models are, hence, those that are *enlightening*.

References

- Atkinson, D. (2003). Experiments and thought experiments in natural science. In M. C. Galavotti (Ed.), *Observation and experiment in the natural and social sciences* (pp. 209–225). Dordrecht: Kluwer Academic Publishers.
- Boumans, M. (1999). Built-in justification. In M. S. Morgan & M. Morrison (Eds.), *Models as mediators* (pp. 66–96). Cambridge: Cambridge University Press.
- Brandom, R. B. (1994). *Making it explicit. Reasoning, representing, and discursive commitment*. Cambridge, MA: Harvard University Press.
- Bueno, O., & Colyvan, M. An inferential conception of the application of mathematics (forthcoming).
- Carruthers, P. (2006). *The architecture of the mind: massive modularity and the flexibility of thought*. Oxford: Clarendon Press
- Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Clarendon Press.
- Cartwright, N. (2009). If no capacities then no credible worlds. But can models reveal capacities? *Erkenntnis*, this issue. doi:[10.1007/s10670-008-9136-8](https://doi.org/10.1007/s10670-008-9136-8).
- Contessa, G. (2007). Scientific representation, interpretation, and surrogate reasoning. *Philosophy of Science*, 74, 48–68.
- Grüne-Yanoff, T. (2009). Learning from minimal economic models. *Erkenntnis*, this issue. doi:[10.1007/s10670-008-9138-6](https://doi.org/10.1007/s10670-008-9138-6).
- Hughes, R. I. G. (1997). Models and representation. *Philosophy of Science*, 64, 325–336.
- Ibarra, A., & Mormann, Th. (1998). Datos, fenómenos y constructos teóricos—Un enfoque representacional. *Theoria*, 31, 61–87.
- Ibarra, A., & Mormann, Th. (2006). Scientific theories as intervening representations. *Theoria*, 55, 21–38.
- Jones, M. R. (2005). Idealization and abstraction: A framework. In M. R. Jones & N. Cartwright (Eds.), *Idealization XII: Correcting the model. Poznań studies in the philosophy of the sciences and the humanities* (Vol. 86, pp. 173–217). New York: Rodopi.
- Knuutila, T. (2005). Models, representation, and mediation. *Philosophy of Science*, 72, 1260–1271.
- Laymon, R. (1982). Scientific realism and the hierarchical counterfactual path from data to theory. In P. Asquith & T. Nickles (Eds.), *PSA 1982* (Vol. 1, pp. 107–121). East Lansing: Philosophy of Science Association.
- Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.
- Mäki, U. (1994). Isolation, idealization and truth in economics. In B. Hamminga & N. B. De Marchi (Eds.), *Idealization VI: Idealization in economics* (pp. 147–168). Amsterdam: Rodopi.
- Morgan, M. (1999). Learning from models. In M. Morgan & M. S. Morrison (Eds.), *Models as mediators* (pp. 347–388). Cambridge: Cambridge University Press.
- Morgan, M., & Morrison, M. S. (Eds.). (1999). *Models as mediators*. Cambridge: Cambridge University Press.
- Nozick, R. (2001). *Invariances: The structure of the objective world*. Cambridge: Harvard University Press.
- Sorensen, R. (1992). *Thought experiments*. Oxford: Oxford University Press.

- Strevens, M. Why explanations lie: Idealization in explanation (forthcoming).
- Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science*, 71, 767–779.
- Sugden, R. (2002). Credible worlds: The status of theoretical models in economics. In U. Mäki (Ed.), *Fact and fiction in economics. Models, realism and social construction*. Cambridge: Cambridge University Press.
- Sugden, R. (2009). Credible worlds, capacities and mechanisms. *Erkenntnis*, this issue. doi:[10.1007/s10670-008-9134-x](https://doi.org/10.1007/s10670-008-9134-x).
- Suppe, F. (1989). *The semantic conception of theories and scientific realism*. Urbana and Chicago: University of Illinois Press.
- Swoyer, C. (1991). Structural representation and surrogate reasoning. *Synthese*, 87, 449–508.
- Weisberg, M. (2007). Three kinds of idealization. *Journal of Philosophy*, 104, 12.
- Winsberg, E. (2006). Models of success versus the success of models: Reliability without truth. *Synthese*, 152, 1–19.
- Woodward, J. (1989). Data and phenomena. *Synthese*, 79, 393–472.
- Ylikoski, P. (2008). The illusion of depth of understanding in science. In De Regt, Sabinelly, & Eigner (Eds.), *Scientific understanding: Philosophical perspectives*. Pittsburg: Pittsburg University Press.
- Zamora Bonilla, J. (1992). Truthlikeness without truth. A methodological approach. *Synthese*, 93, 343–372.
- Zamora Bonilla, J. (1999). Verisimilitude and the scientific strategy of economic theory. *Journal of Economic Methodology*, 6, 331–350.
- Zamora Bonilla, J. (2000). Truthlikeness, rationality and scientific method. *Synthese*, 122, 321–335.
- Zamora Bonilla, J. (2006a). Science as a persuasion game. *Episteme*, 2, 189–201.
- Zamora Bonilla, J. (2006b). Science studies and the theory of games. *Perspectives on Science*, 14, 639–671.