# Idealized models as inferentially veridical representations: a conceptual framework

Juha Saatsi*

March 12, 2010

**Abstract**

This paper erects a framework for analyzing some idealized models as (what I call) inferentially veridical representations. It adopts a version of the semantic view of theories that focuses on properties, and mobilizes conceptual resources associated with properties and the way that properties are related in various ways. The outcome is an elaboration of some aspects of the analysis of Jones (2005).

## 1 Introduction

Idealized models misrepresent the world, but typically only partially. That is, idealized models typically latch onto reality in significant respects. How should we analyze the way in which idealized models can provide partially veridical representations of their target systems? Or take some idealization—a respect in which an idealized model misrepresents the world—and ask: what is its function, and how is it related to the truth-latching aspects of the model? Here I provide some answers to these questions, applicable to some idealizations.

My analysis is incited by my scientific realist inclinations. Scientific realists are in the business of defending the claim that (strictly speaking) incorrect theories and models which are predictively successful (in appropriate measure) typically latch onto reality in a way that is explanatory of their success. The challenge most often raised against the realist concerns theories and models of past science, inconsistent with today's best justified beliefs. But idealized models present a well-known challenge to realism as well. (Cartwright, 1983; Morrison, 2000) The conceptual framework developed here is the first step in an argument to the conclusion that predictively successful idealized models often latch onto reality in those respects

*I wish to thank two anonymous referees for their valuable comments.

that are explanatory of their success, even if they are heavily idealized in other respects. This first step focuses on analysing a way in which idealized models can provide partially veridical representations of unobservable aspects of the world.[1]

Much work has gone into analysing 'partial' or 'approximate' truth in connection with linguistic representations. There are highly sophisticated accounts of the logic and semantics of verisimilitude, for example, and one can approach idealization from that perspective. An alternative approach, with a venerable tradition in philosophy of science, is to construe scientific modelling as incorporating an element of non-linguistic representation. (e.g. Giere, 1988) In the spirit of this tradition there is incentive to analyze approximate or partially veridical nature of scientific models directly in terms of non-linguistic representation relation. This is the approach that I adopt, following a variant of the semantic view of theories that is representative of this essentially non-linguistic tradition. 'Semantic view' labels a broadly model-based view of scientific theorizing. By virtue of focusing on model systems as essentially non-linguistic entities, the semantic view (arguably) liberates philosophy of science from unnecessarily tight involvement with logic and propositions, providing us with new conceptual and technical resources to analyze the nature of models, approximations, and idealizations (amongst other things).

There are various versions of the semantic view that come in various degrees of formalization. My perspective is most closely aligned with the version espoused by Giere (1988) and Teller (2001). At the heart of this view, I take it, is the idea that models function by allowing modellers to attribute properties to their target systems. A *model description* specifies all the properties of a *model system*, and also the representation relationship between the model system and the target. Although language is involved in specifying a model system and also the relation between a model system and its target, the latter relation—the relation that is critical for analyzing how models can partially latch onto reality—is itself non-linguistic. For Giere and Teller this relationship is a symmetric one of *similarity*. I do not subscribe to this specific assumption, nor do I wish to make any universal assumptions about the nature of the model system. For me it only matters that a model system is used to *attribute properties to its target* in such a way that we can start thinking

[1]The second step in the realist argument would be to argue that the predictive success of some idealized models can indeed be *explained*, in a realist sense, in terms of those models' partial veridicality. The third and final step would be to argue that the second step *justifies* realism regarding these kinds of idealized models. These further steps are beyond the scope of this paper. Here I just focus on elucidating a kind of partial veridicality, which is of interest independent of the realism debate. Nevertheless, I will be open about my ultimate realist agenda throughout.

about a model partially latching onto reality by virtue of correctly attributing to the target some properties, but not others, if taken as a fully faithful representation of reality.[2] This focus on properties is profitable, I will argue, because in analyzing the way that models approximate, idealize, or abstract, we can mobilize conceptual resources associated with properties and especially the way that properties are related in various ways.

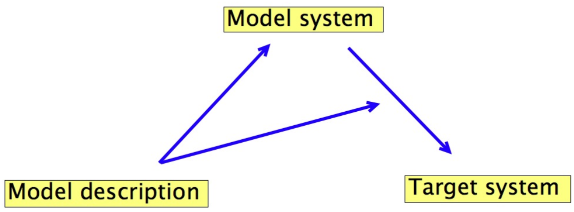


Figure 1: A schematic representation of a model, comprising a model description and a model system. (Modified from Giere, 1988.) The two arrows from the model description are (typically) linguistic specifications of a model system, on the one hand, and of the non-linguistic representation relationship between the model system and the target, on the other.

Some properties are related by virtue of laws of nature, whilst some other properties are related by some stronger modal connection. For example, we have *determinable* properties which logically necessitate the instantiation of a corresponding *determinate* property, and vice versa. We have properties that are *multiply realisable* by virtue of some law-like connection. And so on. I will argue below that we can make headway in analysing at least some idealized models by focusing on non-causal relations that hold between various properties that models attribute to their target systems. A little more precisely, I will be arguing that at least some idealized models can be understood in terms of those *models being 'inferentially veridical' regarding some relevant properties, despite purposefully misrepresenting some unnecessarily specific properties of the target system, for some pragmatic gain*. I will spell this out by explaining what 'inferential veridicality' means; what

[2]I do not take the semantic view to be wedded to structuralism of any sort, either. A structuralist rendering of the semantic view allows one to analyze approximation and partial truth in terms of formal notions such as partial structures and partial isomorphisms, introducing a degree of welcome precision. (E.g. da Costa & French, 2003) At the same time, focusing on such formal notions doesn't itself bring out certain conceptual resources that allow us to understand how certain idealizations naturally arise and function in science.

‘relevant properties’ are; what I mean by ‘unnecessarily specific properties’; and what sorts of ‘pragmatic gains’ can be had by thus misrepresenting reality.[3]

The rest of the paper will proceed as follows. After setting out my initial assumptions in §2, I will further set the context by adopting a particular conception of idealization in §3. Then I will spend some time spelling out the key notion of inferentially veridical representation, with examples, in §4, before finally applying it to idealization in §5.

## 2 Initial assumptions

In this section I present four initial assumptions regarding properties and representation underlying my analysis. But first some prefatory remarks on properties.

Properties are features that particular systems can exemplify. The kinds of properties that science is concerned with are *general*: more than one thing can exemplify (or ‘share’) them. I use the term ‘property’ to cover both monadic properties and relations. There are many well-known philosophical questions concerning the nature and ontology of properties. I’m not interested in these questions here. Rather, I just want to help myself to the various well-established *conceptual* distinctions and resources that property-talk offers. For example, consider the common idea that psychological facts may be ‘fixed’ by the physical facts (together with the laws of nature), but not vice versa. The key concepts here, *supervenience*, and *multiple realizability*, are typically construed as relationships between pairs of families of properties. Or consider the fact that the electron and the positron are similar in a certain respect by virtue of being charged, unlike neutrons, say. This similarity is construed as sharing a *determinable* property, being charged, that relates to the two more specific *determinate* properties, being positively charged and being negatively charged, respectively. These are examples of typical property talk in analytic philosophy. According to (the Giere-Teller version of) the semantic view models represent by specifying properties exemplified by their target systems, and I want to investigate what use can be made of such conceptual resources in analysing models that are less-than-fully-veridical representations.

Now, my four basic assumptions:

[3]The notion of ‘inferentially veridical representation’ developed here is not a philosophical theory of representation per se, but rather an account of how false scientific representations can latch onto reality. Hence, it is quite distinct from the inferential conception of scientific representation developed by Suárez (2004).

**1)** Given the way that properties can be (non-causally) related, *when one explicitly represents a system as exemplifying some particular property, one ipso facto implies (implicitly, perhaps) that the system exemplifies some other properties.* For example, by representing a system comprising 10 electrons and 10 positrons, one ipso facto implies that the system comprises 20 charged particles. One also implies that the system exemplifies the property <u>having a composite number of particles</u>, which is a property also shared by a collection of 21 neutrons, but not by a set of 19 neutrons (given the primality of the number 19). And so on, for countless other properties. The implication regarding such properties is usually only implicit, of course.

In general, by explicitly representing the system as exemplifying a *more specific* property one ipso facto implies that the system exemplifies some *less specific* properties. By '*more specific*' property I mean just this: exemplifying a more specific property is *one way (out of many possible ways) of exemplifying* a less specific property. For instance, one way of comprising 20 charged particles is to comprise 10 positive and 10 negative particles. Another way is to comprise 15 positive and 5 negative. And so on. Familiar examples are afforded by other determinate–determinable pairs: being red is one way of being colored; weighing 10kg is one way of weighing between 9 and 11kg; being an equilateral triangle is one way of being a triangle; being neuro-toxic is one way of being poisonous; etc.[4]

**2)** *Misrepresentation regarding a more specific property is compatible with veridicality regarding some implied less specific property.* This is obvious, as it just follows from the fact that for any less specific property there are more than one possible ways of exemplifying it. It is trivial to come up with examples with gerrymandered, disjunctive properties. To give a more interesting example, consider a detailed model that represents a gaseous system of particles in a box, specifying the momentum of each particle. In as far as temperature of this system is determined by the mean kinetic energy of its constituent particles, our model may be fully veridical with respect to the property <u>temperature</u>, whilst widely misrepresenting the system regarding some (or even all) particular momenta.

**3)** *There is a sense in which a model's predictive success may be due to the model being veridical with respect to some appropriate less specific properties that it im-*

[4] See, Funkhouser (2005), for example, on the determination relation and how it differs from other kinds of the specification relations, such as multiple realization. My 'more specific' and 'less specific' is meant to cover all of these relations.

*plies.* This happens if *any* model, such that it explicitly attributes to the target *some* way of exemplifying these less specific properties, entails commensurate predictive accuracy (when conjoined with the same laws of nature). For example, consider the model of gas above, assuming that we are predicting the system's temperature: any specific distribution of particular momenta will do, as long as the average kinetic energy is right.

Here's a less obvious example. It has been recently argued that some properties of soft condensed matter (polymers, e.g. DNA in solution) are independent of the specific interactions between the individual molecules and of the exact geometrical properties of the macromolecule, but rather depend only on their (less specific) *topological* properties. A model of the inter-molecular interactions and the molecule geometry that gets the topological property right can make right predictions by virtue of this fact. After all, different more specific modelling assumptions regarding the macromolecule's geometry would entail the same predictions, since these are just different ways for the macromolecule to have the appropriate topological feature on which the relevant behaviour exclusively depends.

Viewing this from a realist perspective—in term of how predictive success and truth are arguably typically connected—we can put it as follows: *the more specific properties and the less specific properties need not be on a par in explaining a model's predictive success.* The less specific properties can be solely responsible for predictive success—or *'success-inducing'*—in the following sense: if the prediction hangs on getting a less specific property right, then *any* more specific property that is just a way of exemplifying this success-inducing less specific property, is equally good in the model (as far as predictive accuracy is concerned). A more specific property, on the other hand, can be play various less essential roles in bringing about success, or be *'success-catalyzing'*. For example, we might be able to explain how scientists were able to cognitively access some success-inducing properties by dealing with more concrete, more specific success-catalyzing properties that turned out to misrepresent reality.

**4)** *A model that is veridical vis-à-vis some success-inducing properties, but misrepresents some more specific properties, can be an idealized model.* This follows from the intuition that (many) idealized models are models that make 'simplifying' assumptions that are not too detrimental for their (predictive, say) purpose. Consider a model that is idealized by virtue of providing a simplified representation of some aspects of its target. If those simplifying assumptions concern purely some more specific properties that are suitably related to the success-inducing less specific properties, then the misrepresentation that issues from the idealization should not be detrimental for predictive success since the model is veridical regarding success-inducing properties themselves.

How can misrepresentation at the level of more specific properties yield an *idealization*? That is, in what sense can a misrepresentation vis-à-vis more specific properties *simplify* a model? The idea here is that an idealizing assumption that misrepresents the target plays a purely pragmatic role in giving us cognitive or mathematical handle, say, on the success-inducing properties. In other words, idealizing assumptions are success-catalyzing by virtue of allowing us—with our limited cognitive and mathematical powers—to build models which are veridical in the respects that are required for bringing about an accurate prediction. We could try modelling directly in terms of the success-inducing properties, but this can be cumbersome or impossible both cognitively and mathematically, because success-inducing properties can be highly abstract and unspecific. This is why there is often *pragmatic need* to introduce more concrete and more specific assumptions into a model, even if those assumptions per se are not really success-inducing.

In principle we could try modelling a system with any of the more specific properties that are ways of exemplifying the success-inducing less specific properties, since these less specific properties are really responsible for the model's success, and these properties underdetermine the more specific properties. But it may happen that some of the more specific properties are privileged in terms of yielding a model that is the simplest (for us) to operate with, and those privileged more specific properties need not be the ones that the target actually exemplifies. When it comes to pragmatic considerations it is better to build an idealized model that lies about the more specific properties, but is easier to work with, than a model that is veridical with respect to the more specific properties, but leads to mathematical complications, say. I will look at an example of this later (§5).

* * *

These are—in very general terms—the initial assumptions underlying the conceptual framework that I wish to erect. No doubt these will have to be refined and clarified in due course. I will spend the rest of the paper developing these initial ideas in some detail and situating my position in wider context, beginning with a review of the clearest extant analysis of idealization. This paper can be viewed as an attempt to clarify and extend some aspects of that analysis.

## 3 Jones on Idealization

We should follow Jones (2005) in partially defining 'idealization' and 'abstractions' as follows. *It is a necessary condition for a model to contain an **idealization** (viz. to be an idealized model) that the model misrepresents at least some properties of its target system. Models incorporate an **abstraction** in those respects in which they completely omit certain properties, i.e. say absolutely nothing about them.* There are further conditions that we will want to impose on idealizations, but stipulating this much is already clarifying given the various conflicting uses of these terms in the literature. I maintain that this is the best starting point for thinking about idealization.

A couple of remarks about Jones's framework before proceeding further. (i) It should be seen as a regulative framework for philosophers of science to define their terms. The literature is replete with contradicting uses of these terms, so this is much desired. (ii) These definitions are not meant to capture the various ways in which scientists use terms 'idealization' and 'abstraction'. (iii) A model that only contains abstractions is simply veridical regarding the target system, but limited in its content. There is no good sense in which abstractions are non-veridical. (iv) When we say that abstractions omit properties by saying nothing about them, we really mean that. For example, if we have a model of a particle that is an abstraction with respect to the property of of being electrically charged, it doesn't mean that the charge is zero, so that if the particle was placed in an electromagnetic field, the field would have no effect on it. For this would be an idealization, falsely representing the charge as vanishing.

Saying this much only partially defines idealization as opposed to abstraction, and Jones (2005) recognizes that there must be some further factors that are often present when we speak of an idealization, simply because not all misrepresentations are idealizations. Jones takes the following further factors to be 'typical':

- TRUTH APPROXIMATION: what an idealized model says about its target is often 'approximately true'.
- SIMPLIFICATION: the function of an idealization is to simplify by misrepresenting some features of the target.
- GOAL RELEVANCE: somehow what gets misrepresented are the properties relevant for our modelling goals.

There's broad agreement in the literature that all of these are typical aspects of idealization. But so far this only amounts to little more than sign-posting, and it would be good if we could sharpen the key notions 'approximate truth', 'simplicity', and 'relevant features'. Below I will try to elaborate on Jones's framework by focusing on the kinds of conceptual resources already alluded to.

## 4 Inferentially veridical representations

Let's begin with approximate truth. First of all, I propose that we save the term 'truth' (approximate or otherwise) for the logico-semantic notion that relates to linguistic representation and propositions. Instead, let's use 'veridical representation' in connection with models that can be construed as incorporating non-linguistic representations: a model is a (fully) veridical representation if all the properties the model attributes to the target are actually exemplified by the target. For most models full veridicality is a chimera, of course, and we need a notion qualified somehow. I propose two qualifications below, to define *'partially veridical'* and *'inferentially veridical'* representation. After outlining these concepts, I will illustrate them with some simple examples. (My ultimate realist agenda is worth keeping in mind.) In the next section I will return to idealization.

Having a *partially veridical representation* can be initially construed simply as a matter of having a representation that is correct about some of the properties it attributes to the target. This notion as such, taken at face value, is rather trivial and uninteresting, for just about any model in science is a partially veridical representation in *some* respects. Teller (2001) suggests a good way to proceed in order to avoid triviality. What really interests us is not just the partial veridicality of a representation, as defined above, but partial veridicality *with respect to those properties that are appropriate for the modelling purpose*. From here on I will use 'partially veridical representation' in this *contextualized* sense that takes into account the use to which a model is put. Hence, from now on, a model is a partially veridical representation (relative to some context of use) if and only if it correctly attributes to the target some properties that are appropriate for the purpose of our modelling. This is a wholly contextual notion, since the partial veridicality of a model depends on what we are trying to do with it.

Defining partial veridicality in this way as a representation that is correct about some relevant properties attributed to the target is ambiguous: do we mean the properties that are *explicitly* attributed to the target, or do we consider the properties that are implied *implicitly*? I propose that we disambiguate by reserving 'partial veridicality' for explicitly attributed properties, and introduce the term *'inferentially veridical representation'* to cover both explicitly attributed and implicitly implied properties that the target exemplifies. For example, a detailed particle model of a gaseous system (cf. §2) may not explicitly represent its target as having a temperature, even though it implies it (assuming that temperature is identified with mean kinetic energy). The model may misrepresent all individual particle momenta, yet be partially veridical with respect to particle number, say, and only inferentially veridical with respect to the property $\underline{\text{temperature}}$. The notion of inferential veridicality is still used in the contextualized sense that takes into account the use of a model. Otherwise the notion threatens to become trivial by virtue of every model being inferentially veridical in countless ways, due to the fathomless gerrymandered properties implied by a model.

Inferential veridicality, thus sketched, is serviceable for analysing how non-veridical models can latch onto reality in ways that make those models successful. But it raises issues about property individuation. How do we identify a model's *success-inducing* properties (cf. §2) to see whether or not (or to what extent) the model is an inferentially veridical representation? At this point it is paramount to recognize the importance of various (non-causal) relations between properties, as we need to be able to consider and contrast the role of less specific versus more specific properties in producing a given success. We need to be able to consider the possibility that a model is an inferentially veridical representation due to some relevant less specific properties being success-inducing, whilst more specific properties are misrepresented. This only requires that the properties critical for our modelling interest are actually less specific than the properties the model falsely attributes to the system. We can then go on to consider the role played by the more specific properties in the model. If not success-inducing, they may nevertheless be 'success-catalyzing' in some sense, e.g. by making the model simpler, or by making a heuristic contribution, say.

I will give illustrative examples of inferentially veridical models before connecting inferential veridicality with Jones's understanding of idealization in the next section.

**Example 1.** Here's an intuitive toy-example adapted from Strevens (2004) to begin with. Assume that you are interested in a system comprising a white billiard ball, with momentum $\mathbf{p}$, hitting a glass window pane. In your model a black billiard ball travels at the same velocity, but has momentum $\mathbf{p'}$ (with $p' > p$) (due to larger mass), hitting and breaking the glass. The model is inferentially veridical with respect to predicting the glass breaking. We can ignore the properties of the system it misrepresents: the color of the ball (obviously), and also the exact mass and momentum of the ball. The model's predictive success is due to it correctly attributing the following key property to the target: the ball's momentum is higher than some critical $\mathbf{p_c}$ (with $p_c < p$) required for breaking the class. This property is less specific than the exact momentum the model attributes to the ball. That is, the model correctly attributes (implicitly) to the system a property it exemplifies—*the ball having momentum bigger than* $\mathbf{p_c}$—and latching onto this property is appropriate vis-à-vis your predictive use of the model. (The model is also veridical with respect to the exact speed of the ball, but this in itself is not success-inducing property.) This is an apt model for its purpose, and its aptness is down to it being a inferentially veridical representation in this way.

**Example 2.** Assume that you have a model that purportedly represents a particular type of knot puzzle. By using the model you are able to predict that most people in the test group cannot systematically figure out how to untie this type of knot when first presented with it. Your prediction is based on a topological property of the model-knot, combined with cognitive science knowledge of humans' limited ability to mentally manipulate spatial figures of certain similar topological types. It might happen that the actual knot used in testing human subjects is different from the model-knot, but not too different: it exemplifies the same topological property. (The model knot is right-handed, whilst the actual knot is left-handed, say.) How do we explain your predictive success in the face of this representational discrepancy? We should attribute this predictive success to the model being veridical with respect to the topological property of the knot. After all, the regularity discovered by cognitive science (we assume, for the sake of the argument) concerns specifically *topological* properties of a spatial figure—not this or that more specific geometrical property that is a way of realizing the topological property. Our model, though strictly speaking non-veridical, is an inferentially veridical representation by virtue of correctly attributing the more abstract topological property that is appropriate for this purpose of modelling.

**Example 3.** The toy-examples above are meant to give an intuitive handle on the kind of inferential veridicality of models that I have in mind. Here's a more realistic example from the realist literature.

Augustin Fresnel's ether theory of light has been much discussed as a historical case that demonstrates how a radically false theory can yield novel predictive successes. One of these successes—Fresnel's derivation of the so-called Fresnel equations for the intensity of reflected and refracted polarized light—has been particularly prominent in the literature. Should we accept the anti-realist contention that Fresnel's predictive success relied on false theoretical assumptions? Or should we explain Fresnel's success perhaps in terms of whatever 'structural' assumptions, say, he made regarding light? (Worrall, 1989)

I have argued that neither response is appropriate in this case. (Saatsi, 2005, 2008) A closer look at Fresnel's derivation indicates that underlying the derivation there was a set of modelling assumptions according to which the undulating elastic solid ether realizes some rather abstract spatial continuity properties. The mathematical equations that Fresnel used to derive his end result capture these more abstract properties. With respect to these properties there is a perfect correspondence between Fresnel's theorizing, on the one hand, and the contemporary derivation of Fresnel's equations from Maxwell's equations, on the other. Hence, the realist can put her finger on this kind of continuity in the less specific properties in spelling out her realist commitments. Fresnel's theorizing was latching onto reality and the model underlying his derivation was inferentially veridical.

**Example 4.** Assume that you wish to model, with good predictive accuracy, a tennis ball rolling down an inclined asphalt plane. You use a simple model found in introductory physics textbooks, with good results. The model misrepresents the target in various respects, but very little in each respect: the gravitational field is not *absolutely* homogeneous; the ball is not *absolutely* smooth, hard, and spherical; etc. But intuitively speaking, the model latches onto reality in a way that ensures its predictive success.[5] How should we conceptually precisify this intuition about the model latching onto reality? I prefer to think in terms of the properties that the model explicitly attributes to the system—the ball is perfectly spherical, there is absolutely no slippage, and so on—and see how they relate to the various less specific properties that the model in a sense must get right in order to be accurate.

For the model to be considered successful there has to be a bit of leeway between the actual motion of the ball and what the model says: the model's predic-

[5]I'm not *arguing* for realism about such a model here. Rather, I'm illustrating how the notion of inferentially veridical representation can be employed to articulate the sense in which the model is arguably latching onto reality in a way that is responsible for its predictive accuracy.

tions are good, but not perfect. Once we (contextually) decide how much leeway is accepted, we can start thinking about the various ways in which the model could be modified without too big an effect on predictive accuracy. For example, we can replace the homogeneous gravitational field by an inhomogeneous one without too big an effect, as long as the inhomogeneities are very small. The perfect sphericality of the ball and the perfect smoothness of the surface can be replaced by properties that are, intuitive speaking, "close enough". And so on.[6] There is an enormous number of ways in which a model's assumptions can be thus modified without affecting the model's predictive accuracy too much. Most of such modifications lead to mathematical complexities in practice, of course, and one has to resort to various approximation schemes or computer simulations in order to get predictions out of most of the alternative models.

But ignoring such usability issues altogether, and focusing purely on the connection between a model's predictive accuracy and the properties it attributes to the target, we see that there is clearly a sense in which the original model is *unnecessarily specific* in representing the system as having the exact properties. For there is an enormous range of alternative models—each more or less on a par in "approximating" the actual system—that are equally good in terms of predictive success (relative to the amount of leeway accepted). What really matters for predictive success, then, is that each exact property that a model attributes to the system belongs to some appropriate *range* of specific properties. Or, in other words, that each exact property attributed by a model is a way of exemplifying some appropriate less specific property. Any predictively successful model has to imply that the ball exemplifies the less specific property <u>being *approximately* spherical</u>, where 'approximately' is a shorthand for a disjunction over the set of those more specific properties that entail an outcome within the limits of agreed accuracy.[7] (This set includes the exact shape actually exemplified by the ball.) In the same sense, the surface has to be *approximately smooth*, and the gravitational field has to be *approximately homogeneous*, and so on. These are the properties that any predictively successful model will correctly attribute to the system, and any model that is not successful in this way will fail to attribute at least one of these properties to the system. The intuition that the original model latches onto reality gets thus analysed

[6] Models represent by specifying properties exemplified by their targets, and we can contemplate various changes to the model *directly* in terms of these properties, regardless of whether or not these changes can be nicely captured through the equations that partly define the original model.

[7] Hence, 'approximately' is fixed by a contextual error bound, together with the other properties attributed to the system.

in terms of the model being inferentially veridical by virtue of correctly implying, implicitly, that the target exemplifies these critical less specific properties.

* * *

All the above examples indicate that inferential veridicality of a model can be due to a representation latching onto reality via a false property attribution that is *unnecessarily specific*.[8] In each case one could try to present a more abstract model that only attributes a less specific property, correctly, to the target system. For example, one could have represented the white billiard ball as an object of some appropriate size hitting the glass with *some* momentum larger than $\mathbf{p_c}$ (without specifying any particular momentum). Similarly, one could have represented the knot mathematically as having a particular topological property, without saying more precisely whether it is this or that specific knot that realizes this particular topological character. Likewise, Fresnel could have (in principle, even if not in practice) abstracted away from the unnecessarily specific properties that he mistakenly attributed to light, focusing purely on the less specific spatial continuity properties required for his derivation. Also, one could try presenting a less specific model for the rolling tennis ball, only attributing to the system those less specific properties that the original model correctly implies, together with all the different models that are predictively successful (where predictive success is determined by some accuracy requirement).

But even if there exists a corresponding abstraction that is fully veridical, in many cases it is quite natural not to present these less specific properties directly, but instead give a model that is more specific in such a way that the appropriate less specific properties are ipso facto implied. There are various reasons why our models often involve such unnecessarily specific properties, related to various features of our modelling practices. I will next finally argue that some idealized models can be analyzed in terms of such unnecessarily specific properties playing a pragmatic, simplifying role.

[8]Unnecessarily specific in the sense that some of the specific properties the model attributes to the system are not really responsible for predictive success, but at best play some lesser, ‘success catalyzing’ role.

## 5 Elaborating on Jones's framework

The ubiquity of idealization in science presents the scientific realist with some difficult questions. What good can it be for a model to *lie* about the system represented by the model? This general question can be narrowed down by asking it with respect to different uses of models; we can ask the question, for example, with respect to explanatory or predictive worth of idealizations. In this paper I have focused on the predictive aspect, which I consider to be more pressing for the realist.[9]

The realist intuition is to claim that we misrepresent in order to simplify our model, because misrepresentation in some respects does no (significant) harm for the predictive ability of the model, whereas not misrepresenting in those respects would lead to enormous complexities or to the lack of usable model altogether. Conceptually precisifying that intuition is a great challenge: the realist needs to account for the indispensability of idealizations, on the one hand, and at the same time defend the realist claim that the predictive success of idealized models is in a substantial sense down to those models latching onto reality, and not down to their lying about the world. My contention that some idealized models may be construed as inferentially veridical representations goes some way toward answering that challenge, by providing a conceptual framework for analyzing a model's 'latching onto' its target in a way that can be explanatory of its predictive success.[10] Finally, I will now explain how misrepresenting in some respects can *simplify* a model and thus be indispensable to our modelling practices.

Recall the initial assumption (4) in section §2: an idealized model can be a model that (a) is inferentially veridical by virtue of correctly implying some less specific success-inducing properties exemplified by its target, but (b) purposefully lies about some more specific properties, because doing so does no harm, but rather provides some pragmatic benefits. This is a way of elaborating on Jones's idea that idealizations do not misrepresent any old property, but typically properties that are *somehow* relevant for the purpose of the model. The properties being misrepresented are those properties that are very intimately connected to the properties that are success-inducing. Namely, the misrepresented properties are possible ways of

[9]For the explanatory dimension of this question, see Strevens (2008) ch. 8.

[10]The next step in the realist argument would be to defend the idea that the explanations furnished by my conceptual framework really are *realist* explanations. I believe (but will not argue) that this step can be taken, on a case-by-case basis, for many idealized models.

the target system having the less specific, success-inducing properties.[11] All that is required for such misrepresentation to count as an idealization is that the misrepresented more specific property is somehow pragmatically useful.

Some of the examples in the section above indicate how it may be natural not to present the success-inducing less specific properties directly, but instead give a model that is more specific in such a way that the relevant less specific properties are ipso facto implied. For example, any *material* model of the billiard ball hitting the glass is necessarily going to attribute this or that specific momentum to the target, even though the success-inducing property is $\underline{\text{having a momentum greater than } \mathbf{p_c}}$. Regarding the knot example, we can just note that abstract topological properties or equivalence classes cannot be simply pictured, unlike the particular knots realizing those properties. These are not clear cases of idealization, however, since there is nothing to make it pragmatically useful to adopt any particular way of implying the success-inducing property via some non-veridical more specific property.[12] Also, the model that Fresnel employed is not idealized, for it is not the case that due to some pragmatic constraints Fresnel could only latch onto the more abstract success-inducing properties by way of making knowingly false, more specific assumptions regarding the ether. Rather, it was natural for Fresnel to make such more specific assumptions, and they were heuristically indispensable, given the metaphysical and explanatory presuppositions of the ether paradigm.

The final example in the section above is different, however. The inclined plane model of the tennis ball is clearly idealized by virtue of the fact that the various false assumptions it incorporates yield the mathematically simplest—it is safe to assume—model amongst all the altenative models of acceptable predictive accuracy. So when it comes to demonstrating or predicting the speed and position of the tennis ball over time (within given error-bounds), we can do it with the least amount work with this model. There are other models which are equally good if judged purely with respect to our goal of predictive accuracy, each specifying a particular property to the ball, to the inclined surface, to the gravitational field,

[11] This dovetails with the intuition that idealized models represent *counterfactual* systems which demonstrate, explain, or predict the relevant phenomenon in fiction, or in other possible world. Much more work is needed in order to make precise the sense of modality involved, however. It's true that some idealizations seem to represent systems that are nomologically possible in the actual world, but most idealizations are not nomologically possible. So we need to relax at least some laws of nature to view such idealized system as a possible system. But at the same time idealized models are faithful to some actual laws of nature.

[12] A mathematical model of the window-breaking billiard ball might be idealized assuming that, say, some relevant calculations are simplified somehow by attributing a round number 10 kgms$^{-1}$ ($> p_c$) as the momentum. Thanks to an anonymous referee for this suggestion.

etc. Not all these models are on a par regarding their mathematical simplicity, however, and the mathematical description of the actual system may also be overly complicated and unworkable. We might try and operate with the less specific, success-inducing properties directly, without specifying this or that particular way of being *approximately* spherical, *approximately* homogeneous, and so on. But doing that would be extremely cumbersome or impossible mathematically. It is so much easier to write down and deal with an equation for a model that attributes a precise force acting on the ball at each moment of time, than it is to deal with the much more abstract model that *only* says that there is *some* force, within some bounds, acting on the ball at each moment.

Idealization like this is rife in science, but admittedly there are different kinds of cases that may push the conceptual framework that I have developed here over its limits. For example, it is not obvious how to characterise the success-inducing properties specified by infinite population models in biology, say, or by a model of waves in an infinitely deep pond of water, and I don't claim that the framework presented here is applicable across the board.[13] But the realist's conceptual framework needs to cover these cases, too, for how can she take these models to 'latch onto' any real system, unless we can somehow discount those infinite numerical differences between the model and the actual system as kinds of differences that are ultimately immaterial to explaining the success of the model in terms of its truth-latching aspects?

## References

Batterman, R. W. (2001). *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford University Press.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Oxford University Press.

DaCosta, N. and S. French (2003). *Science and Partial Truth*. Oxford: Oxford University Press.

Funkhouser, E. (2006). The Determinable-Determinate Relation. *Noûs 40*, 548–569.

[13]Cf. Batterman 2001 for examples of models with infinities.

Giere, R. (1988). *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

Jones, M. R. (2005). Idelization and Abstraction: A Framework. In M. R. Jones and N. Cartwright (Eds.), *Idealization XII: Correcting the Model. Idealization and Abstraction in the Sciences*, Volume 86 of *Poznan Studies in the Philosophy of the Sciences and the Humanities*, pp. 173–217. Amsterdam: Rodopi.

Morrison, M. (2000). *Unifying Scientific Theories*. Cambridge: Cambridge University Press.

Saatsi, J. (2008). Eclectic realism—the proof of the pudding: a reply to Busch. *Studies in History and Philosophy of Science 39*, 273–276.

Saatsi, J. T. (2005). Reconsidering the Fresnel-Maxwell case study. *Studies in History and Philosophy of Science 36*, 509–538.

Strevens, M. (2004). The causal and unification approaches to explanation unified – causally. *Noûs 38*, 154–176.

Strevens, M. (2008). *Depth: An Account of Scientific Explanation*. Harvard: Harvard University Press.

Suárez, M. (2004). An Inferential Conception of Scientific Representation. *Philosophy of Science 71*, 767–779.

Teller, P. (2001). Twilight of the Perfect Model Model. *Erkenntnis 55*(3), 393–415.

Worrall, J. (1989). Structural Realism: The Best Of Both Worlds? *Dialectica 43*, 99–124. Reprinted in Papineau, D. (ed.) *The Philosophy of Science* (Oxford: Oxford University Press), pp. 139 ? 165.