Fiction, Models and the Problem of the Gap

Frederick Kroon

The University of Auckland

Abstract

An increasingly popular view holds that scientific modeling involves something akin to the imaginative construction of a fictional story along with its cast of fictional characters, not just the positing of entities (models) that yield a false but useful representation of their targets. The present paper focuses on the following problem for this view of models. If a model is a fiction how can it possibly be said to represent some aspect of the real world? How can the unreal represent the real, and in a way that allows modelers to make predictions about the real, and even explain some of its features? Call this the problem of the gap. The paper begins by motivating the fiction view of models, describing and contrasting the two most popular types of view (both based on Walton's pretense theory of fiction), together with the way they deal with the problem of the gap and some other, related problems. I then sketch a modified version of the fiction view, one that takes on board aspects of both of these approaches by utilizing an important but under-appreciated feature of fiction, and I argue that the view provides natural solutions to this suite of problems.

Keywords: Models, Fiction, Nonexistence, Pretense, Fictional surrogate objects.

1. Introduction

In the early 20th century there was an interesting form of anti-realism to match the anti-realism of logical positivism: Hans Vaihinger and the philosophy of "as if" (Vaihinger 1911). For Vaihinger, the posits of science were to be seen, by and large, as fictions, where fictions were construed as falsehoods: false assumptions that "contradict reality", including false assumption that there exist things of a certain kind (call this the error-theoretic sense of 'fiction'). Such fictions were nonetheless to be retained because they were instrumentally useful. This form of fictionalism is most closely mirrored, perhaps, in the work of van Fraassen, although van Fraassen doesn't regard the theoretical unobservable posits of science as fictions but only as posits whose existence is irrelevant to the development and usefulness—in the sense of empirical adequacy—of science (van Fraassen 1980).

Argumenta 6, 1 (2020): 47-63

ISSN 2465-2334
DOI 10.14275/2465-2334/202011.kro

Not surprisingly, it is hard to find fictionalists in Vaihinger's strong sense. There is widespread agreement among scientists and philosophers of science that we should take current formalisms of this or that theory with a grain of salt, but relatively few would agree that even when we get to the final theory about some domain we should reject its posits as non-existent and the theory itself as only instrumentally useful. In his seminal retrospective assessment of Vaihinger's ideas (Fine 1993), Arthur Fine offers a different view of the applicability of Vaihinger's ideas. He thinks that where they have particular resonance is in the area of *scientific modeling* rather than theoretical science:

> Preeminently, the industry devoted to modeling natural phenomena, in every area of science, involves fictions in Vaihinger's sense. If you want to see what treating something "as if" it were something else amounts to, just look at most of what any scientist does in any hour of any working day (Fine 1993: 16).

Many philosophers of science and scientists have come to accept that models should indeed be classed as fictions of a certain kind, but there is an escalating debate about what this means. As suggested earlier, Vaihinger understood 'fiction' in the error-theoretic sense, and it clear that at least in 'Fictionalism' Fine uses the term in the same way. But an increasingly popular view holds that models are fictions in a somewhat different and arguably richer sense—that modeling involves something akin to the imaginative construction of a fictional story along with its cast of fictional characters, not just the positing of entities (models) that, by dint of the involvement of processes like idealization and abstraction, yield a false but useful representation of their targets. When contemporary theorists talk of *the fiction view of models*, it is this work-of-fiction understanding that they tend to have in mind.

The present paper is mainly about work-of-fiction fictionalism about models but its focus is a problem that also arises for the error-theoretic kind. The problem is this. If a model is a fiction, whether because its posits are akin to fictional characters in a fictional story or because it posits nonexistent items, how can it possibly be said to represent some aspect of the real world? How can the unreal represent the real, and in a way that allows modelers to make predictions about the real, and even explain some of its features? Call this the problem of the gap.

The paper's format is as follows. Section 2 sketches and motivates work-of-fiction fictionalism about models (from here on: *the fiction view of models*), while sections 3 and 4 describe and contrast the two most popular types of view (both based on Walton's pretense theory of fiction), together with the way they deal with the problem of the gap and some other, related problems they face. Section 5 sketches a modified version of the fiction view that takes on board aspects of both of these approaches by utilizing an important but under-appreciated feature of fiction, and describes its own solution to this suite of problems.

2. Models and Fiction

Roughly speaking, in modeling scientists apply prepared descriptions and theoretical laws that they know to be false in order to understand and predict features of target structures in the world. How to characterize the activity of modeling and its products in more precise terms is of course a difficult and contentious matter, something that is underscored by the sheer variety of "things" (the scare quotes are there to remind us of the problem being tackled) that are called models. On the surface, not only do we have concrete models like wind tunnels or string-and-ball models of the solar system, but there are also models that involve idealization and abstraction with respect to properties of a target (frictionless planes, point masses, etc.), as well as models of a more mathematical kind that are focused on hypothetical structures, such as the Lotka-Volterra model of predation.

What is at any rate clear is that, while models all involve falsehood (broadly speaking, the error-theoretic sense of fiction described earlier), for some theorists there is more to the role of fiction as a way of understanding modeling than falsehood, even imaginatively constructed falsehood. To understand models, so they think, we need to appeal to *literary fiction*. A relatively early proponent of this idea was the philosopher of biology Peter Godfrey-Smith, who noted that theorists tended to talk about their models in concrete terms, and that this was even true in the case of mathematical models (Godfrey-Smith 2007, 2009). As he saw it, these are never *purely* mathematical, since we tend to distinguish models that use the same mathematics (e.g., the harmonic oscillator model whose mathematics can be used to describe both an idealised spring and a chemical bond). A closer look at such models shows that they purport to represent the real causal structure of target phenomena, with the mathematics serving as an essential tool to doing so. A description of the Lotka-Volterra model of predation, for example, doesn't begin with the mathematics, but might be introduced with talk of two imaginary populations that have properties like birth rates, capture rates, and so on, that can then be described mathematically.

For Godfrey-Smith, attending to the way theorists *talk* about models is crucial to understanding models:

> I take at face value the fact that modelers often take themselves to be describing imaginary biological populations, imaginary neural networks, or imaginary economies. An imaginary population is something that, if it was real, would be a concrete flesh-and-blood population, not a mathematical object (Godfrey-Smith 2007: 735).

(Reflecting this stance, Thomson-Jones talks of 'the face-value practice of modeling'; cf. Thomson-Jones 2010.) Godfrey-Smith then asks what the best way is to account for this way of speaking and the uses to which models are put. His answer points to a striking similarity to fiction. Typical models, including mathematical ones, involve imaginary systems that would be concrete if they were real.[1] The Lotka-Volterra

[1] In this paper I am setting aside the tricky status of concrete models (e.g., wind tunnels, string-and-ball model of the solar system, etc.), although there is reason to think that the way these represent their targets involves us in somehow imagining them *as* their target (Toon 2012: Ch.5; cf. also the DEKA model of Frigg and his co-authors; Frigg and Salis 2020: §3).

model of predation, for example, consists of a model system that is an imaginary population of predator animals and prey animals.[2] These have the properties explicitly attributed to them in the act of modeling (growth and death rates, say); others are inferred from what has been stipulated, using mathematics and biological "laws"; others are don't-cares. (See especially Godfrey-Smith 2009.) But this, he notes, is *very* similar to the way fictional worlds are constructed. The world of Sherlock Holmes is partly a matter of stipulation, partly a matter of inference from what is stipulated; much of it, say the number of hairs on Holmes's head, is a don't-care.

Godfrey-Smith thinks that this analogy between model systems and fictional objects is non-accidental and important. Model-based science is in the business of specifying imaginary worlds, although its purpose in doing so is not literary but purely cognitive: it aims at understanding, explaining, and predicting features of the world. To this end, Godfrey-Smith points out a nice feature of the analogy between models and literary fiction: the way modelers often talk of the *similarity* between models and their targets when they apply the model. This kind of talk is tricky if you think of models as mathematical, say, while the idea of relevant similarity between models and targets appears natural and "unintimidating" from the fiction point of view. Thus, just as we can compare two physical systems, we can compare two fictional systems (e.g., Tolkien's "Middle Earth" and the world of Malory's King Arthur in *Morte D'Arthur*). And just as we can compare a model system to its target physical system we can compare a fictional system with a physical system (e.g., events in Orwell's *Animal Farm* are similar to those in Russia in the first part of the 20th century).

Godfrey-Smith doesn't have a great deal more to say about how best to understand the analogy to fiction, but he says enough to indicate a potentially serious problem for the fiction view of models. In the comparison between a model system and its target physical system, what we seem to be doing is comparing properties associated with one system with properties associated with another. But if model systems are imaginary objects such as concrete infinite populations, frictionless planes, and so on, they don't exist and so can scarcely be said to have properties that can be compared to properties of things in the real world. There is, to put it differently, an ontological gap between model systems (when seen as analogous to fictional characters) and their real-world targets, a gap that makes it hard to see how we can learn about the real world from models. Because its aims are different, there is no such problem for literary fiction.

How is the problem best solved? Godfrey-Smith (2009) discusses a number of options without coming down on one side or the other. What is clear, however, is that one's preferred solution will depend on which version of the fiction view of models one chooses. Consider, for example, the view that model systems are abstract entities. In this case there is no ontological gap since model systems so construed *exist*. What this view retains from the face-value picture of modeling is the idea that the model system is an extra entity. What is not preserved, at least not straightforwardly,

[2] Like many authors, I tend to use 'model' and 'model system' interchangeably. Strictly speaking, however, a model system is the (purported) entity that serves to portray or represent some target or other. The model describes how the model system does this, using whatever parts of science and mathematics are needed.

is the idea that model systems and their targets are similar to the extent that the properties of the one broadly correspond to properties of the other. Instead there is a more abstract mapping of some kind, with only the formal structure of the relations between objects on each side being preserved. As Godfrey-Smith points out, on a familiar Platonistic interpretation of structures such a view seems to inherit well-known problems encountered in the literature on the semantic view of theories.

These problems may disappear if the idea of models as abstract structures is tied more closely to the idea that model-building involves pretense. Thus Thomasson (2020) and Thomson-Jones (2020) develop an "artifactualist view" of models, following the contours of Thomasson's view of fictional characters, on which the content of a text that introduces a model should be understood as occurring in pretense, while in producing such descriptions authors create abstract cultural artifacts. On such a view, there is a sense in which it is correct to say 'point masses don't exist' (just as it is correct to say 'Holmes doesn't exist'), even though as artifacts they literally do exist (just as Holmes literally exists). Proponents claim there is a clear reason for positing such entities: we routinely assert truths external to the pretense, such as 'Bohr's quantized shell model of the atom gets more of an atom's structure right than the plum pudding model'. Given, as they argue, that statements of this kind have no plausible paraphrases that eliminate reference to models, that suggests that such models really do exist.

But of course, abstract objects cannot really have such properties as being a biological population or composed of protons and electrons, so the view doesn't conform as straightforwardly as one might like to the "face-value practice of modeling". In addition, the idea inherits other problems to do with the nature of such abstract objects (see, for example Brock 2010). I will here set it aside in order to consider pure pretense theories that try to do without such extra objects.

3. The Fiction View of Models and *De Dicto* Imagining

Consider one such account, due to Roman Frigg (this is the account most fully discussed in Godfrey-Smith 2009). The account adapts Kendall Walton's well-known make-believe account of fiction (Walton 1990) to the case of models. Walton focuses on the way a text can be used as a resource in games of make-believe in which participants pretend, imagine, or make believe that the world is as the text represents it as being. If readers let their imaginings be directed in this way, they are then participating in a game of make-believe that is *authorized* by the work. For Walton, a proposition can be said to be fictional—true in the fiction—just in case participants in such a game of make-believe are supposed to imagine it as true. There are two types of fictional truth: the *primary* fictional truths are evident in the work itself, taking proper account of the linguistic conventions that allow us to understand the work, while the *implied* fictional truths are generated from the primary ones by taking into account what the world is like, or perhaps what the community of origin of the text believed the world was like.[3] It thus turns out that it is true in the Holmes stories that *Sherlock Holmes lived nearer to Paddington Station than to Waterloo Station* (no sentence in the Holmes corpus

[3] For further details, including criticism, see Kroon and Voltolini 2019.

actually says this, so this is an implied truth), while it is false in the stories that Holmes had a wife, for example.

Frigg thinks descriptions of models are structurally much like works of fiction, even if their purpose is very different (they are supposed to provide an understanding of the world, not be a source of entertainment). A model description serves as a prop for a game of make-believe in which participants imagine that the world is as the model description represents it as being. Mimicking the case of fiction, not only are there propositions explicitly authorized by the model description, but in addition there are implied truths:

> What is explicitly stated in a model description (that the model-planets are spherical, for example) are the primary truths of the model, and what follows from them via laws or general principles are the implied truths (Frigg 2010: 260-61).

For another example, take the classical case of Fibonacci's population model, as described in Frigg and Salis 2020. Here the primary truths of the model include such claims as *The rabbits breed in six month intervals*, and the implied truths include claims like *The rabbit population grows monotonically*, which can be derived from the basic assumptions of the model supplemented with some basic facts of arithmetic.

Note the apparent lack of worrisome metaphysical commitments. As it seems, such models are committed only to systems comprising concrete things such as populations of reproducing rabbits, perfectly spherical planets in circular orbits around a massive sun, planes not subject to friction, and so on. The modeler proceeds by imagining a system comprised of such things, and then draws conclusions about their properties using relevant theories and mathematics. There are no further commitments, say to model systems as abstract artifactual entities.

But how do we relate the model system to the target? Unless we are talking of physical models, there is no physical resemblance between model system and target system—the left-hand side of any relation of resemblance is purely in the modeler's head. How can we possibly plug this gap and show how modelers can apply their models to the real world? Here we seem to encounter the problem of the gap in its most pernicious and challenging form.

Frigg's answer is to draw on the way we can use 'transfictional' claims to say what the real world is like. Uttering a sentence like 'Morris Zapp is no more conceited than most academics' allows us to state something about the conceitedness of academics by taking a property that Morris Zapp has in David Lodge's *Changing Places* and affirming that this same property is abundantly instantiated among real world academics. He thinks we can do the same with models. If, for example, I say that some actual rabbit population behaves like a population in the model, certain properties are on the table that can be compared to the properties of a real rabbit population:

> [T]ransfictional statements about models should be read as prefixed with a clause stating what the relevant respects of the comparison are, and this allows us to rephrase comparative sentences as comparisons between properties rather than objects, which makes the original puzzle go away (Frigg 2010: 263).

This is a kind of reductive account: rather than comparing nonexistent things directly with existent things (an impossibility, Frigg thinks, since there are no such things as nonexistents), we compare (existent) properties that imagined things have in a model to (existent) properties of things in the real world. But this account faces a number of objections. Godfrey-Smith points out that many of the properties that are being introduced when dealing with fictional models will be uninstantiated ones and that these may raise special problems of the same kind as those seen with fictional objects.[4] But I suspect that Godfrey-Smith's deeper underlying worry is that models should be seen as *representing* their targets, and a package of allegedly nonexistent entities and properties seems particularly ill-suited to this task. So the wider problem that is not solved by the property-comparison account is this: *how can the model-target gap be closed when there literally are no concrete models to represent the target*?[5] Call this the no-representation problem.

In later work, Frigg is explicit about the need for an account of how models *represent* their target systems. To deal with the problem of representation, Frigg and his co-authors develop the idea, inspired by Goodman and Elgin, of a model M *t-representing* a target system T. Briefly, M *t-represents* T if it denotes T and represents T *as* being a certain kind of thing Z exemplifying Z-style properties, properties that are then related via a key to another set of properties at least some of which M imputes to T (see, e.g., Frigg and Salis 2020: §3). This makes sense if models are existing concrete objects, such as a string-and-ball model of the solar system, but not when they involve such things as immortal rabbits and frictionless planes. Following Salis (see especially Salis 2020), Frigg and Salis respond to this problem by modifying the account of a model. Instead of taken models to include such things as nonexistent immortal rabbits, they associate the model with the content of the fiction together with the text that describes the content, not with the fictional object that is described in the text. So conceived,

> [a] model is a tuple $M = [D, C]$, where D is the description of the model and C is the [full] content of the description … (i.e. the set of propositions that are specified by D together with the principles of generation). ... C now takes the place of what one intuitively would call the 'model system' (such as Fibonacci's immortal rabbits). Because model-descriptions and their contents exist, models thus construed are *bona fide* objects (akin to fictional stories) that can enter into relations (Frigg and Salis 2020: 202).

[4] It is not entirely clear what Godfrey-Smith had in mind. Levy takes him to mean that "the model, being merely imaginary, cannot instantiate properties" (Levy 2015: 790). More plausibly, he has in mind such "properties" as *being a non-extended physical object*. Thomasson (2020) responds that the existence of uninstantiated properties can be argued for 'by making pleonastic inferences' such as moving from "The wand is not magical" to "the property of magicalness is not possessed by the wand" to infer that there is a property of magicalness that the wand (indeed everything) lacks. But this move looks question-begging. If someone says, pretending that a gu-gu is a new kind of primitive primate, that you are a "gu-gu", your reasonable protest that you are not a gu-gu (on the grounds that you are a human) doesn't show that there is such a property as being a gu-gu. That there is a genuine property of being a gu-gu requires at the very least some intelligible account of what it is to be a gu-gu.
[5] See also Toon 2012: 58 and Levy 2015: 789-90.

Assuming this notion of a model is even coherent,[6] how do models so construed *denote* and *represent* target systems? Frigg and Salis give only the briefest of indications:

> The model thus defined exists and therefore can stand in the denotation relation with real world systems (Salis 2020: 20),

and

> a look at scientific practice suggests that in many cases the denotation of a model piggy-backs on the denotation of denoting linguistic symbols. In our example, Fibonacci's model denotes what it does because we use the denoting expression "the rabbit population in the London Zoo" (Frigg and Salis 2020: 203).

Salis (2020) argues in some detail that we can exploit our knowledge of models so construed to learn about target systems. Perhaps. But the question remains in what sense this is like *denotation*, the relation that plays such a straightforward, pivotal role in their account of t-representation.

4. The Fiction View of Models and *De Re* Imagining

Salis herself sees her account as fixing the failures of a rather different way of understanding the way models denote their targets: what she calls the *direct fiction* view defended in somewhat different ways by Toon (2012) and Levy (2015) and one which is immune to the no-representation objection. One way to motivate their view is to look at the work of a prominent opponent of the fiction view of models like Paul Teller. For Teller a point particle or continuous fluid represents *real* objects such as extended objects and bodies of water:

> A real extended object is fictionally described as having no extension. A real body of water is fictionally described as being a continuous fluid. Such cases constitute fictional descriptions of real objects. So such cases should be thought of, not as object fictions, but as state of affairs fictions, as fictional characterization of states of affairs of real objects (Teller 2009: 244).

(Here 'fictions' and 'fictional' applies to anything that is non-veridical, while *object fictions* are non-veridical, i.e., nonexistent, entities.) Teller, and, following him, Ronald Giere, argue that while scientists might *call* entire models fictional, this may

[6] Note the following prima facie semantic problem: according to the kind of Millian orthodoxy Walton accepts, there may be no propositions expressed by the relevant text. Take model descriptions that contain names of, as it seems, nonexistent entities, for example the silogens referred to in models of fractures in micron-sized pieces of silicon (Giere 2009) or the ether, as modeled by Maxwell's mechanical model. Frigg's original account can maintain that the model descriptions only describe entities from the point of view of the pretense, and so may only express what Kripke (2013) calls *pretend* propositions. The new account cannot afford this luxury, or may need to adopt a more descriptivist, e.g., Ramseyan, view of the content of model descriptions.

be for no other reason than that they contain component object fictions (cf. Giere 2009).

Although Teller does not explicitly put it this way, his account suggests that a model just *is* its target, but a target that is misdescribed through the use of idealization, abstraction, and approximation. That is precisely how Toon (2012) and Levy (2015) see models. But where Teller sees a role for "fictional" (i.e., "non-veridical") characterizations of states of affairs involving real target objects, Toon and Levy see something much closer to the deployment of the literary notion of fiction. Modeling, on their view, aims to provide an imaginative description of real things, with a description of the model prescribing, effectively through the use of Walton's machinery of rules of generation, what we are to imagine about the real system. In the case of the ideal pendulum, for instance, model-users are required to imagine real springs as perfectly elastic and the bob as a point mass, with laws and mathematics needed to supply a stock of inferred truths about the movement of the bob under these conditions. These inferred truths can then be used to make predictions about, and explain features of, real pendula. Levy's presentation of this idea appeals to Walton's notion of prop oriented games of make-believe, for example games in which we imaginatively speak of Italy as a boot or of thunderclouds as faces as a means of thinking and reasoning about them (Walton 1993). Levy's suggestion is that

> we treat models as games of prop oriented make-believe—where the props, as it were, are the real-world target phenomena. To put the idea more plainly: models are special descriptions, which portray a target as simpler (or just different) than it actually is. The goal of this special mode of description is to facilitate reasoning about the target. In this picture, modeling doesn't involve an appeal to an imaginary concrete entity, over and above the target. All we have are targets, imaginatively described (Levy 2015: 791).

This is not the only seemingly significant difference between the way Toon and Levy describe their accounts. For Toon (2012), model descriptions of models with targets (e.g., the simple pendulum) prescribe imaginings about real concrete targets, while model descriptions of models without targets simply prescribe imaginings or, at best, imaginings about purely fictional systems (Toon 2012: §3.3). By contrast, Levy (2015: §4.4) argues that there are no targetless models, appearances notwithstanding. Models that appear targetless may do so because, for example, "the specific range and features of the intended target are not known for sure', or because they are "generalized models [that] work as hubs anchoring specific models" (2015: 796-7). Other apparent models like the Game of Life are genuinely targetless, but Levy thinks that these are little more than "bits of mathematics" rather than models in a full-blooded sense (2015: 797).

Following Salis, I want to highlight two problems for Levy's account in particular: the *no-target* and *indirectness* problems. First, the account seems to have nothing to say about certain familiar kinds of models that, unlike the targetless models Levy mentions, seem *aimed* at a target. Examples include Maxwell's mechanical model of the ether and models of synthetic molecules that will never be created in a lab, perhaps because the models reveal that any such entities would cause great harm. Secondly, the relation between model and target seems typically far less direct than Levy and Toon make it out to be. As Salis (2020) puts it:

> Stating that model descriptions are about real objects does not dispense with fictional entities (and the controversies they generate) because model descriptions always involve apparent reference to some fictional objects. … [Take the simple pendulum]. The model description of the simple pendulum is not about any particular pendulum. It does not start with 'Imagine of this particular pendulum in front of you that it is a point mass suspended by a massless, unstretchable string'. Rather, it apparently refers to an imaginary system consisting of a point mass and a massless string and hence prescribes imagining about a fictional system (Salis 2020: 12).

I think we should, with some qualifications described below, accept Toon's answer to the first objection, which is that some imagining is directed at fictional target systems (and hence that this kind of imagining is effectively embedded in imagining, rather than knowing, that there is a real target system). The second objection strikes me as in some ways more important, although it is not hard to see the beginnings of an answer. As Salis points out, modelers typically do not begin with an instruction to imagine, of some particular thing or of any of a class of particular things, that it has certain properties specified in the model. The relation between model and target is often, even typically, more indirect. But *contra* Salis, we should at the same time note that there are also cases where the relation is described as being much more direct, in much the way emphasized by Toon and Levy. Science texts, for example, often talk about the different idealized ways in which atoms are *described* by Rutherford and Bohr, say, not just about the different ways in which atoms are *modeled*, or about the way a model of the solar system might *describe* the Earth as being a point-like object that doesn't rotate.[7] A rather nice example is given by Levy, who quotes the two ways in which Turing characterizes his mathematical model of the growing embryo: in one version "the cells are idealized into geometrical points" while in the other "the matter of the organism is imagined as continuously distributed" (Levy 2015: 782).

Not only do both the direct and indirect perspectives occur in the literature. It is also clear that the difference in perspective would not strike modelers themselves as particularly significant. They would be unlikely, for example, to reject a presentation of the Fibonacci population model that went as follows:[8] "My kids have two rabbits, one male and one female. Their names or identities don't matter. What matters is that they are ready to mate. Let's describe how their number will grow by making some simplifying assumptions. Assume that rabbits always mate six months after birth, that the female of each pair gives birth to exactly one male-female pair another six months after mating, that they never die, etc. [Now comes the Fibonacci calculation of rabbit pair numbers at all future moments.] That is the Fibonacci population model!" None of us, modelers included, would be nonplussed by such a presentation of the model.

[7] See Matthews *et al.* 2005 for the various ways in which the simple model of the pendulum has been described.

[8] If the presentation was intended for a journal or a text meant for researchers, they would, perhaps, be chided by a referee; but there the reason has to do with the culture of scientific academic writing (which is also the reason why there might not be criticism if the presenter was known to be famous).

What is more important to scientists who use the model is that it can be applied to target populations other than the one featured in the introduction.

5. Learning from Fiction

To see how the fiction view of models is able to cast light on this seemingly odd duality of perspectives, it is time to apply some lessons from the case of literary fiction. Consider real objects that feature in fictional works, for example Napoleon in *War and Peace*. While real individuals can appear in fictional works, in the works they are often very different from the way they actually are. These differences, whether large or small, have given rise to the view that real individuals as they appear in works of fiction should be regarded as distinct fictional characters: fictional *surrogate* objects for short. Meinongians in particular should see the attraction of the view. Just as they think it is true that Andrei Bolkonsky (a purely fictional character in Tolstoy's *War and Peace*) was wounded at the Battle of Austerlitz, they should also think it true that Napoleon rescued Andrei at this battle: for Meinongians, fictional truth suffices for truth. Since in reality Napoleon did not rescue Andrei, it must be a surrogate object that did so—the Napoleon of *War and Peace*. Artifactualists, who think that fictional objects are abstract objects created by authors, have provided other reasons for thinking that there must be a surrogate fictional Napoleon.[9]

But even pretense theorists should admit a sense in which real objects have fictional surrogates. After all, in our pretenses the Napoleon of the story must be distinguished from the real Napoleon. We are pretend-referring to someone who rescued Andrei at Austerlitz. Napoleon wasn't like that! Similarly, when I read *War and Peace* and admire Napoleon for his kindness in rescuing Andrei it is the Napoleon of *War and Peace* I admire; I might detest the real Napoleon on the grounds that he, on the other hand, would never have done such a thing. Note again how we use certain familiar qualifiers to draw the contrast; we talk of *the Napoleon of* War and Peace or of *Napoleon as he was in* War and Peace, and contrast that person with other versions of Napoleon: the *real* Napoleon, say, or *Napoleon as he was in Vincent Benét's* 'The Curfew Tolls'. Our ability to make sense of these distinctions doesn't depend on whether one is a pretense theorist or a fictional realist of some kind (Kroon 1994).[10]

But how do these fictional surrogates of Napoleon relate to (the real) Napoleon? Fictional realist proponents of fictional surrogacy tend to agree that the surrogates in some sense represent their real-world counterparts, even if there is disagreement about the nature of this relation.[11] If one is a pretense theorist, however, there is much simpler account one can give of such relationships: there *is* in fact no relationship between a Napoleon surrogate like the Napoleon of *War and Peace* and Napoleon himself since

[9] See Motoarca 2014 and Voltolini 2013, 2020.
[10] In fact, artifactualist believers in fictional surrogate objects could probably adapt the kind of surrogacy-friendly view of modeling defended in this paper to yield an alternative to the artifactualist account of modeling found in Thomasson 2020 and Thompson 2020. (They may, of course, have independent reasons to resist such an extension of the notion of surrogacy.)
[11] It needn't be representational in any strong sense; Voltolini, for example, considers it a many-many relation of similarity (Voltolini 2020: 815).

the Napoleon of *War and Peace*, an occasionally kind person who rescued Andrei Bolkonsky, doesn't exist. Instead, we should say the following. In writing *War and Peace* Tolstoy wanted his readers to imagine, of Napoleon, that Napoleon did certain things that he did not in fact do. Much else that he wanted his readers to imagine about Napoleon is based on facts about the latter's actual life and deeds. But when we, in response, engage imaginatively with *War and Peace* we do so from the inside: in the scope of our pretending that the world is as reported in *War and Peace* we represent to ourselves someone who is rather different from the real Napoleon (of course in so doing we must import facts about the real Napoleon and his exploits in so far as these don't conflict with the prescription to imagine what the novel tells us—there could be no fiction without such an anchoring in reality). So, while the story is partly *about* Napoleon and his exploits (here 'Napoleon' refers in the standard way to Napoleon), when we engage with the story we are not referring to him. This is because we are not referring at all: we are only *pretend*-referring, referring from inside the scope of the pretense (or, as some prefer, referring at a pretend or fictional context instead of at a real context).[12] In the scope of the pretense, we are referring to someone who has the properties he is ascribed in the novel—and that person, aptly characterized as *the Napoleon of* War and Peace, doesn't exist.

(Quick proof that we are not really referring. If we were, our utterances would be up for evaluation for truth or falsity in the usual way, and so it would be entirely appropriate for listeners to accuse us, over and over again, of uttering falsehoods. But this would not be appropriate—our utterances are not truth-normed in this way. Despite this, we can learn a lot about the world of Napoleon by reading *War and Peace*. Doing so requires some sensitivity, but, roughly speaking, if Napoleon is described in *War and Peace* as having done X and there are no artistic ends that would be served by Tolstoy asking us to imagine this even though he believed that Napoleon did not in fact do X, then, given that Tolstoy is reliable where Napoleon is concerned, it is probably safe to infer that Napoleon *did* do X. In short, it is appropriate to *export* fictional truths under certain circumstances, that is, to interpret them as genuine, non-pretend truths, even though it is admittedly difficult to frame rules about how to do this.)[13]

Return now to the case of models, and consider again the kind of Waltonian pretense accounts discussed in previous sections. Levy and Toon think that in modeling we are imaginatively re-describing real-world systems (but sometimes fictional systems, if Toon is right). By contrast, Frigg and co-authors like Salis argue that model systems are the product of *de dicto* imagining the existence of model systems like point masses, infinite populations of immortal rabbits, and the like. The view I prefer borrows from the lessons we have just learned about Napoleon and his surrogates: to the extent that in modeling we are indeed imaginatively re-describing real-world systems, that idea does not in any way get rid of the idea of nonexistent model systems. It just requires us to rethink their role and nature.

Here, in brief, is the idea. Let X be a real-world system or object—a real pair of rabbits, say, or the solar system or a pendulum, or…—and suppose we imaginatively

[12] For discussion of this use of the notion of context, see Kroon and Voltolini 2019: §2.1.
[13] For useful discussion, see Friend 2014.

represent it as satisfying various assumptions *F*, both idealizing and auxiliary (such as the existential assumption that there exist silogen atoms; Giere 2009), while also abstracting away features of no concern. (For the sake of brevity, I'll simply talk of *assumptions of idealization and abstraction*.) Then the model system is object X conceived of as conforming to these assumptions of idealization and abstraction *F*, say a physical pendulum idealized as a point mass bob suspended from a string of zero or negligible mass, with the only forces acting on the bob being the force of gravity and tension from the string. (Given the principles of generation for the pretense, it is also part of the pretense that this object has properties *P* whose possession by X follows from X satisfying *F*, given mathematics and relevant scientific laws.) That is the surrogate object we encounter from inside our pretense, just as the Napoleon of *War and Peace* is the surrogate figure we encounter as we engage with *War and Peace*. And this object does not in fact exist just as the Napoleon of *War and Peace* doesn't exist.

How do we apply and learn from models once we understand them this way? Here is one suggestion: in the same way as we learn about Napoleon by reading and engaging with *War and Peace*. From in the scope of the pretense we can engage with the model system and work out how it behaves under various conditions. At this point there is only pretend-reference and a pretend-ascription of what properties the system would have under these conditions, or, if you prefer, reference and property ascription at a pretend-context. But as in the case of *War and Peace* there is also a non-pretend way of reading the sentences that record these findings: reference and property ascription at a real-world context. We have been assuming that the pretense was based on (*de re*) imagining that a certain target system conformed to a degree of idealization and abstraction. If so, when we refer *apart* from the pretense (that is, at a real context), we are referring to the target system simpliciter. What it is reasonable to export from the pretend-truths of the model and how to qualify or amend these in light of facts about the target system depends on features and the intended scope of the model. Perhaps the process can be understood using the notion of partial truth that Levy favors (Levy 2015: §4.2).[14] We needn't take a stand on this. All I here want to emphasize is that this is a version of a pretense account that takes due note of the role of surrogate objects, and thereby suggests a prima facie attractive way of closing the gap between model systems and the real-world systems they represent.

Note that this way of describing how we learn from models is misleading in so far as it treats the pretense as focused on the content of models—as content oriented rather than prop oriented make-believe, to use Walton's terms (Walton 1993). What is really going on, of course, is a little different, since models are unlike stories in their orientation: stories are meant to be engaged with from the inside (to treat them as learning tools is not to do them justice), while the purpose of models is to facilitate reasoning about external target-systems. So, it would be more accurate to say that we begin with target systems and an interest in explaining and predicting features of such

[14] Levy describes his use of the notion of partial truth as follows: "The idea, to put it tersely, is that while model descriptions are typically idealized, hence not true of their targets simpliciter, they are nevertheless partly true, at least when successful" (Levy 2015: 792). Although Levy doesn't say this, note that this assumes that applying models involves what I earlier called exportation: to apply a model we must export what is in the first instance merely imagined.

systems and that pretend-reference and pretend-ascription of properties only come in once we try to meet this interest by talking about these systems through assumptions of idealization and abstraction. In short, the pretense involved in modeling should be seen as externally oriented, not content oriented.[15]

In fact, it is tempting to say that the purpose of model systems is to *represent* external target systems. But that way of putting the point hides another problem (the *no-representation* problem) which affects both Frigg's early account of models and the present account—strictly speaking, there are no model systems that can do the representing. Before turning to this problem, let me first deal with a problem that may initially loom even larger: the *indirectness* problem. As presented, the account only works (if at all!) for models based directly on target systems; but as Frigg and Salis emphasize, models are often, perhaps typically, not based on real-world targets this way (set aside models of specific objects like the solar system). But here the very nature of models shows us the way out. I have already suggested that there is something almost incidental about the fact that modelers don't appeal *de re* to real-world targets when they describe their models. They could have, without this affecting their models. That suggests the following solution to the indirectness problem. When devising models, modelers imagine that some *arbitrary* system of the relevant kind is subject to certain assumptions of idealization and abstraction, not (*pace* Salis) some *specific* system. In effect, the modeler in the pendulum case means something like "Consider a pendulum—*any* pendulum—made subject to idealization and abstraction as follows". At the point where the modeler comes to apply her model to a real-world system X, she effectively instantiates to the specific system X. At this point, and without any loss of generality, she simply takes X to be the system that is imaginatively reconfigured to conform to certain assumptions of idealization and abstraction. For purposes of application, the simple pendulum model can thus be thought of as *this* particular physical pendulum imaginatively reconfigured as being *F* as easily as *that* particular physical pendulum. So the fact that models are (typically) introduced without reference to a specific target system is not of any significance on this picture, and is fully consistent with the idea that pretense in the case of modeling has an external orientation.

What about the no-representation problem? Here the pretense account on offer has limited leeway. Absent existent model systems, it is only in the scope of a pretense that model systems (now construed as pretended surrogate objects) denote target systems. Frigg and Salis point out that some "may think that [this] is too feeble a notion to account for how science represents its objects and nothing short of 'real' denotation between models and their targets is good enough" (Frigg and Salis 2015: 201-2). They say they want to keep an open mind about this issue, but as we saw earlier they also propose a new version of the fiction view that, in their view, makes room for real denotation. My suspicion is that our intuitions are simply not clear enough to determine if more than pretense is needed. What certainly is clear is that in saying that models denote real-world targets, we are asserting something about the world and the

[15] I don't use the term 'prop oriented' since, as the indirectness problem reminds us, there need not be any particular prop that is invoked in the imaginative construction of the model.

way modeling activity is directed at the world—we are not merely *pretending* that models and modeling are directed at the world. But there is a familiar way in which this worldly orientation of our utterance can be captured by the kind of pretense view on offer: we often make substantive claims about the world by talking *through* our pretenses, and statements to the effect that a certain model represents a particular target system may be no exception.[16] So I remain unconvinced that more than pretense is needed to make sense of the way we describe models as denoting their targets. If more than pretense is needed, and if, in particular, we want a robust, "non-feeble" account of the relationship between model and target system, then—and only then—do we have to change the way we think of models.

That leaves us with the final problem: the no-target problem, which affects the present view no less than Levy's. In the case of some attempts at modeling there turns out to be no target system, and hence no system that is imagined as subject to certain idealizing assumptions. Such is the fate of Maxwell's mechanical model of the ether and models of synthetic molecules that will never be created in a lab. The solution to this quandary is essentially that proposed by Toon (2012): the target systems are themselves (at best) fictional systems.[17] But more should be said. To the extent that there is no ether there can (of course!) be no model of the ether either. But in Maxwell's case there was the *belief* that the ether exists, and that he had a model of it. To be able to talk of Maxwell's work, commentators simulate this belief: they do *as if*, or pretend that, the ether exists. They do this effortlessly, in much the same way as, picking up a book like *The Hobbit*, we effortlessly engage in the pretense that there are such things as hobbits. And if you are a chemist in the business of hypothetically working out what a certain molecule would be like when no such molecule yet exists, you do *as if* there is such a molecule and (in the scope of your pretense) you describe your attempts to model it. Commentators, including philosophers writing about modeling, will effortlessly join in the pretense, much as friends whom you regale with the exploits of Bilbo in *The Hobbit* join you in your pretense.

[16] We can, for example, extend the original pretense to allow talk of a relationships of denotation by letting the extended pretense include something like the following rule of generation: it is correct to pretend that model system X denotes target system Y just if those who instituted the original pretense mandated that we are to imagine X as a system that conforms to certain assumptions of idealization and abstraction, and that we do so by pretending that Y is subject to these assumptions of idealization and abstraction. (As a consequence, the relevant model description denotes Y at a real context while denoting X at a pretend-context in which it is pretended that Y conforms to certain assumptions of idealization and abstraction.) Assuming that in uttering a statement of the form 'Model system X denotes its target system Y' our interest is in communicating the circumstance that makes this statement true in the extended pretense, the *real* content of our utterance is that this circumstance obtains. Hence in uttering a statement 'Model system X denotes its target system Y' we are making a substantive claim about the world. (See Everett 2013: Ch. 3 on 'Talking Through the Pretense', esp. §3.3.2 which is concerned with modeling fictional characters; see also Armour-Garb and Woodbridge 2015 for a general discussion of how an involvement in pretense allow us to make claims about the world.)

[17] For a very different non-pretense way of dealing with models, including targetless models, see Suárez's influential account of an inferential conception of representation in Suárez (2004).

So are there targetless models? Not on the view proposed. In any such case of an apparent targetless model there is a fiction in which it really is a model with a target—a fiction that we, as commentators, effortlessly engage with. Since on the fiction view of models the usual kind of model systems are to be understood as fictional systems, we should consider these wannabe models *fictional* fictional systems, much as Gonzago, a character within a play enacted in *Hamlet*, should be considered a *fictional* fictional character rather than a fictional character (Kripke 2013: 72-73).[18]

References

Armour-Garb, B. and Woodbridge, J.A. 2015, *Pretense and Pathology: Philosophical Fictionalism and Its Applications*, Cambridge: Cambridge University Press.

Brock, S. 2010, "The Creationist Fiction: The Case against Creationism about Fictional Characters", *The Philosophical Review*, 119, 3, 337-64.

Everett, A. 2013, *The Nonexistent*, Oxford: Oxford University Press.

Fine, A. 1993, "Fictionalism", *Midwest Studies in Philosophy*, XVIII, 1-18.

Friend, S. 2014, "Believing in Stories", in Currie, G., Kieran, M., Meskin, A. and Robson, J. (eds.), *Aesthetics and the Sciences of Mind*, Oxford: Oxford University Press, 228-48.

Frigg, R. 2010, "Models and Fiction", *Synthese*, 172, 2, 251-68.

Frigg, R. and Salis, F. 2020, "Of Rabbits and Men: Fiction and Scientific Modeling", in Armour-Garb, B. and Kroon, F. (eds.), *Fictionalism in Philosophy,* Oxford: Oxford University Press, 187-206.

Giere, R. 2009, "Why Scientific Models Should not be Regarded as Works of Fiction", in Suárez 2009, 248-58.

Godfrey-Smith, P. 2007, "The Strategy of Model-Based Science", *Biology & Philosophy*, 21, 5, 725-40.

Godfrey-Smith, P. 2009, "Models and Fictions in Science", *Philosophical Studies*, 143, 101-16.

Kripke, S. 2013, *Reference and Existence*, Oxford: Oxford University Press.

Kroon, F. 1994, "A Problem about Make-Believe", *Philosophical Studies*, 75, 201-29.

Kroon, F. and Voltolini, A. 2019, "Fiction", *The Stanford Encyclopedia of Philosophy*, Zalta, E.N. (ed.), https://plato.stanford.edu/archives/win2019/entries/fiction/.

Levy, A. 2012, "Models, Fictions, and Realism: Two Packages", *Philosophy of Science*, 79, 5, 738-48.

Levy, A. 2015, "Modeling without Models", *Philosophical Studies*, 172: 781-98.

Levy, A. and Godfrey-Smith, P. (eds.), 2020, *The Scientific Imagination: Philosophical and Psychological Perspectives*, Oxford: Oxford University Press.

[18] Many thanks to Carola Barbero, Matteo Plebani, and Alberto Voltolini for organizing the FINO GC / SIFA MidTerm Conference (June 2019) at the University of Turin where an early version of this paper was presented, and to members of the audience for their stimulating questions and criticisms. Special thanks to Alberto Voltolini for suggesting a number of useful revisions.

Matthews, M., Gauld, C. and Stinner, A. (eds.) 2005, *The Pendulum: Scientific, Historical, Philosophical and Educational Perspectives,* Dordrecht: Springer.

Motoarca, R. 2014, "Fictional Surrogates", *Philosophia*, 42, 1033-1053.

Salis, F. 2020, "The New Fiction View of Models", *British Journal for the Philosophy of Science*, 71, 1-28.

Suárez, M. 2004, "An Inferential Conception of Scientific Representation", *Philosophy of Science*, 71, 5, 767-79.

Suárez, M. (ed.) 2009, *Fictions in Science. Philosophical Essays on Modeling and Idealization*, London & New York: Routledge.

Teller, P. 2009, "Fictions, Fictionalization, and Truth", in Suárez 2009, 235-47.

Thomasson, A.L. 1999, *Fiction and Metaphysics*, New York: Cambridge University Press.

Thomasson, A.L. 2020, "If Models Were Fictions, Then What Would They Be?", in Levy & Godfrey-Smith 2020, 51-74.

Thomson-Jones, M. 2010, "Missing Systems and the Face-Value Practice", *Synthese*, 172, 2, 283-99.

Thomson-Jones, M. 2020, "Realism about Missing Systems", in Levy & Godfrey-Smith 2020, 75-101.

Toon, A. 2012, *Models as Make-Believe: Imagination, Fiction, and Scientific Representation*, London: Palgrave Macmillan.

Vaihinger, H. 1911, *Die Philosophie des Als Ob*, Leipzig: Meiner. Translated by C.K. Ogden as *The Philosophy of 'As If'*, London, 1924: Routledge and Kegan Paul.

van Fraassen, B.C. 1980, *The Scientific Image*, Oxford: Oxford University Press.

Voltolini, A. 2013, "Probably the Charterhouse of Parma Does Not Exist, Possibly Not Even That Parma", *Humana-Mente Journal of Philosophical Studies*, 25, 235-61.

Voltolini, A. 2020, "Real Individuals in Fictions, Fictional Surrogates in Stories", *Philosophia*, 48, 2, 803-20.

Walton, K.L. 1990, *Mimesis as Make-Believe*, Cambridge, MA: Harvard University Press.

Walton, K.L. 1993, "Metaphor and Prop Oriented Make-Believe", *European Journal of Philosophy*, 1, 1, 39-57.