Moving Targets and Models of Nothing

A New Sense of Abstraction for Philosophy of Science

Michael T. Stuart and Anatolii Kozlov

To appear in Julia Sánchez-Dorado and Chiara Ambrosio (eds.). *Understanding through Abstraction in Science and Art*. London: Routledge.

Abstract

As Nelson Goodman highlighted, there are two main senses of “abstract” that can be found in discussions about abstract art. On the one hand, a representation is abstract if it leaves out certain features of its target. On the other hand, something can be abstract to the extent that it does not depict a concrete subject. The first sense of “abstract” is well-known in philosophy of science. For example, philosophers discuss mathematical models of physical, biological, and economic systems as being abstract in this (“subtractive”) sense. However, it is the second sense that dominates discussions of abstract art in aesthetics. For example, abstract art was (and is) considered revolutionary precisely for being non-figurative. Through an analysis of artists including Kandinsky, Malevich, and Mondrian, we develop a reading of this second sense, which we call “generative abstraction”. Generative abstraction is a

process in which a new artefact is created which represents something other than the initial concrete target system that inspired it (if there was one), where the artefact's features are explored for their own sake, and where the "language" of the new artefact is in some way more "universal". Focusing on this sense of abstraction is helpful in revealing the complexity of the process of crafting an abstract artefact, in problematizing the notion that abstraction can always be un-done (or concretized), as well as revealing new ways for abstractions to be epistemically (un)successful.

6.1 Introduction

Abstraction is a process that is required for building scientific models. Typically, it is thought to involve subtracting irrelevant details. In this sense, it is sometimes portrayed as a necessary but acceptable epistemic evil, since it doesn't introduce anything false, and it is reversible.

However, "that there are different kinds of abstractive processes is not often addressed in philosophy of science or cognitive science" (Nersessian 2008, 191). We want to focus on another, very different sense of abstraction, one that is found in discussions of abstract art in aesthetics. This sense is non-representational in some ways, but not in others (Goodman 2003). We think this concept of abstraction better describes certain processes of model-building in science.

In this chapter, we follow the approach of Nancy Nersessian (2008, section 6.2.2.) and Sabina Leonelli (2008) in foregrounding the cognitive-epistemic *process* of abstraction. We begin by looking at the history of work on abstraction in the philosophy of science, to get clear on the “standard” notion of abstraction, which we label “subtractive” abstraction. Then we canvas the history and philosophy of abstract art to present a different notion of abstraction, which we label “generative” abstraction. Then, we employ case studies to show that some scientific processes of abstraction are correctly labelled as generative, not subtractive. Finally, we consider some philosophical implications.

6.2 “Subtractive” Abstraction in Philosophy of Science

In philosophy of science, abstraction is usually discussed with reference to scientific representation, especially scientific models. Demetris Portides sums things up: “models are primary devices of scientific representation” and “idealizations and abstractions are manifest in most (if not all) kinds of scientific representation”. Thus, “it has become commonplace that scientific models, scientific representation and idealization/abstraction are entangled concepts” (Portides 2021). While the terms “abstraction” and “idealization” are sometimes applied to objects other than models, for example, explanations, objects, or “paths” to representations, these discussions are usually closely related to models (Carrillo and Martínez 2023; Jansson and Saatsi 2019; Verreault-Julien 2022).

Why are these notions so closely entangled? Representation is important (at least) because of its central role in epistemological questions about surrogative reasoning in science. Ideally, it is thought, a representation would capture all the aspects of a target system, and so whatever we learn about the representation will also be true of the target system that was represented. In practice, however, scientists cannot represent all the aspects of any target system. So, they employ abstraction and idealization. This complicates the idea that models give us straightforward epistemic access to the world. The main epistemological question, then, asks how a model can provide new epistemic desiderata (knowledge, truth, approximate truth, epistemic access, understanding, pursuit-worthy hypotheses, etc.) about a target system despite (or in virtue of) being abstract or idealized.

Ernan McMullin differentiated between several concepts that would later form the basis of the distinction between abstraction and idealization (McMullin 1985). These two concepts were more recently redefined by Martin Thomson-Jones such that “idealization” should refer to misrepresentation and “abstraction” should refer to mere omission (Thomson-Jones 2005). Idealization “requires the assertion of a falsehood”, and abstraction “involves the omission of a truth” without misrepresentation (175). Thomson-Jones doesn’t claim to capture all the useful ways of talking about model-building, but proposes this distinction as a useful framework for analysing the epistemology of scientific representations. Peter Godfrey-Smith presents a view that differs “only in points of emphasis” (Godfrey-Smith 2009, 48). Specifically, an abstract description “leaves out a lot”, while an idealized description “fictionalizes” in the sense that it does not present a literally true description of the target, and at the same time, it describes an imaginary system that would be concrete if real. This way of thinking about abstraction and idealization still dominates the literature in philosophy of science. Here is a recent statement: “An abstraction is the wholesale omission of a property…An idealisation is the distortion of a property…For this reason, abstractions offer a literally true (albeit incomplete) representation of the target, while idealisations assert, if understood literally, falsehoods” (Frigg 2023, 317).

The literature on abstraction and idealization has since exploded, and many detailed epistemological accounts of both now exist. On abstraction, Michael Strevens argues that one model is more abstract than another if the causal influences described in the latter are also described by the former, and every proposition in the latter model is entailed by the former (2008, 97). Leonelli distinguishes between abstract models understood as (a) non-concrete models, (b) models requiring more information to make empirical statements about the real world, and (c) models applying to more phenomena (2008, 520). Arnon Levy argues that one representation is more abstract than another if it is relatively less informative about the same target (2021). Thus, “mammal” is more abstract than “Red-tailed Chipmunk”. Idealization has tended to take up more of the spotlight because idealizations are (or include) misrepresentations, which present a greater challenge to those trying to account for epistemic uses of scientific representations. Many strategies now exist to deal with this challenge. For example, we can claim that only the true parts of idealized models refer, or that idealizations are merely practical shortcuts, or that idealizations do some epistemic heavy lifting without figuring into the content of the scientific understanding they produce, or that idealizations are not misrepresentations (see, e.g., Strevens 2008, 2017; Khalifa 2017; Lawler 2019; Yablo 2020; Levy 2021; Nguyen 2020).

There is an important assumption shared by almost everyone who participates in the above debates, and it can be traced back to McMullin. Idealization and abstraction (and their subtypes) have *one general aim*: "a deliberate simplifying of something complicated (a situation, a concept, etc.) with a view to achieving at least a partial understanding of *that* thing" (1985, 248; emphasis added). It is important to point out that this aim focuses on simplifying a single, unchanging target system. The goal "is not simply to escape from the intractable irregularity of the real world" but "to grasp the real world from which the idealization takes its origin" (1985, 248). In other words, the target system for abstraction and idealization is deliberately set from the start, and learning about *that* target is the aim. This is a natural assumption to make given the prevalence and importance of surrogative reasoning in science. And indeed, a great deal of scientific modelling does take place in this way.

This idea, that abstraction is a process directed at a single (concrete) target that doesn't change, also plays a role for philosophers who focus on abstraction as a process rather than as a product. For example, Leonelli characterizes abstraction as "the activity of selecting some features of a phenomenon P, as performed by an individual scientist within a specific context, in order to produce a model of (an aspect of) P" (2008, 521). For Leonelli, abstraction is a process that transforms features of the target system into parameters used to model *that very target system*. In Leonelli's case study, the target might be a concrete organism, such as *Arabidopsis thaliana*, or something less-concrete, such as a signalling pathway. But it remains constant from the beginning to the end of the process.

Due to this assumption, scientific model-building is imagined as beginning with the choice of a particular target system (e.g., the pendulum, the atom, an economy, a population of rabbits, a signalling pathway, etc.). Idealizations and abstractions are introduced which shape the model, that is then manipulated, and conclusions are finally drawn about *that* target system. This raises the epistemological question of how those conclusions are justified. In the case of abstract models, the answer might go something like this: the scientist had originally removed some details without misrepresenting the system, so the model will deliver (at least approximately) correct information about the aspects of that target that were not removed. Better still, the scientist can add back in the details that they had earlier removed, to produce conclusions that are even more justified.

In this chapter, we focus on a different way of thinking about abstraction, one that does not hold the target fixed. That is, rather than considering abstraction as a process that is always epistemically tied to a single target system which is given from the start, we consider processes of abstraction that create new systems, new targets, and leave the old targets behind. This is inspired by a notion of abstraction that we find in discussions of abstract art.

6.3 “Generative” Abstraction in Abstract Art

Abstract art is said to originate somewhere between 1910 and 1920. Often regarded as “the most important development of early 20th-century [Western] art”, it is connected with artists like Hilma af Klint, Wassily Kandinsky, Kazimir Malevich, Piet Mondrian, Paul Klee, Mark Rothko, and Jackson Pollock, who were reacting to movements like impressionism and cubism, especially the work of Paul Cézanne, Henri Matisse, and Pablo Picasso (Chilvers and Glaves-Smith 2009). The impressionists, cubists, and abstract artists were united in demanding a new aesthetic that would break away from the kind of mimesis characteristic of artistic realism. However, while cubists and impressionists departed from realism in important ways, it was characteristic of their work that they never gave up on representationalism.

> When Braque and Picasso found their work approaching the non-representational or non-figurative or non-objective (all these terms are used), both artists ‘recoiled.’ They chose, like Cézanne and Matisse and the great majority of post-impressionist

> and modernist painters, not to lose sight of the object. For this reason among others it is often said that the aim of Cubism was essentially to represent reality more accurately and completely.
>
> (Vargish and Mook 1999, 129)

A cubist might present a person, a landscape, or a piece of fruit in a very different way, but they were still presenting a person, a landscape, or a piece of fruit.

What made abstract art different from other kinds of modern art? Calling it "non-representational" is misleading: all its key figures insisted that their work did, in fact, represent something. The difference is that what they chose to represent wasn't a typical visual object, like a person, landscape, or piece of fruit. Abstract artists might start with an object like that, but through a series of changes, they would remove all traces of the object, in order to present a series of lines, shapes, and colours.[1] So far, this looks like subtractive abstraction. But the key is that the result would come to represent something else, something non-visual (Sánchez-Dorado, this volume). This is the sense in which abstract art is non-representational: "modernist abstraction is best understood not in terms of a loss of realistic detail but in terms of shifting the frame of reference away from the object" (Vargish and Mook 1999, 131).

Let's illustrate with some examples. To repeat, omitting details in a painting was something the cubists and other modern artists were already doing. For example, each of Matisse's four nude female backs comprising his *The Back* series (1908–1931) progressively "lose realist detail without losing representational force" (Vargish and Mook 1999, 132).[2] Around the same time that Matisse was working on *The Back I* (1908–1909), Kandinsky was beginning to use the same subtractive abstraction for a different purpose. His early work employs strong blotches of colour and retains a clear link to representational impressionist art, for example, his *Treppe Zum Schloss* (1909).[3] The work that comes even one year later, however, has already moved away from any concrete objects as its focus.[4] For another example, consider Mondrian's increasingly abstract paintings of trees.[5] The point we want to emphasize is that while subtractive abstraction is often involved, even centrally, in creating abstract art, that is not what makes abstract art abstract.

Of course, this wasn't the first time non-representational art was produced (Gertsman 2021), and there were (and still are) disagreements among scholars and practitioners about what abstract art is. Alfred Barr identified two broad approaches, corresponding to the work and motivations of Kandinsky on the one hand, which was intuitional and emotional, and Malevich on the other, which was intellectual and geometrical (Barr 1936). Barr's distinction has been as controversial as influential. More recently, Diarmuid Costello has identified seven kinds of abstraction (2018). But there is always a core idea: abstract art leaves behind figurative visual representations in order to draw attention to a new object that represents a non-figurative target.

By moving away from initial concrete objects, artists were able to break free of the constraints of mimetic representationalism, the "prison" of limited form

(Mondrian 2007). If an abstract artist wanted to investigate something like ambition, for example, they would not need to paint Napoleon on a horse, or anyone, on any horse. Thanks to the cubists, space on a canvas was no longer modelled on a single viewpoint or constrained by the rules of perspective. Further, if you want to express something “divine” or “universal”, as Kandinsky, Malevich, and Mondrian all did, then you will likely need to adopt some kind of common language that will enable you to get the point across to audiences despite the difficult subject matter. In response, artists employed colour, line, shape, contrast, and so on, to present visual melodies and compositions that (they hoped) would convey the right message to different audiences. As Kandinsky wrote,

> Colour is a means of exercising direct influence upon the soul. Colour is the keyboard. The eye is the hammer, while the soul is a piano of many strings. The artist is the hand through which the medium of different keys causes the human soul to vibrate.
>
> (Kandinsky 1977, 43)

Only by moving to more “universal” forms of expression like harmony, line, and colour, which these artists (controversially!) took to be less culturally specific than other means of expression, did these artists believe they could convey more universal messages, or messages about more universal things, like inner harmony, psychic effect (Kandinsky), feeling (Malevich), and pure aesthetic relationships (Mondrian).

One might be tempted to conclude that abstract art is abstract just in the sense that it focuses on abstract targets instead of concrete ones. While this might be, we will remain focused on the *process* of abstraction itself without assuming that the artwork, artefact, model, or target system that is the output of such a process is abstract in some metaphysically heavy sense. Whether feelings are more or less abstract than fairies, functions, or fruit flies, we do not say. Targets of abstract art might always be abstract in the sense that they are significantly (if not entirely) non-figurative. But this does not require such targets always be abstract in the sense of being non-concrete, non-specific, or existing in Plato's heaven. (For examples of concrete abstractions, see Knuuttila, Johansson, and Carrillo, this volume.)

What we are calling "generative abstraction" is a process of creating a representation. It may begin by representing some particular concrete target. It may involve subtracting features from that target in creating the representation. But then it moves on to become a representation of something other than the target that initially inspired it. It is generative in the sense that in the process of creating it, a new target is generated, which is different from the initial target. We will complicate this idea in a moment. But first, consider Costello's helpful discussion of types of abstract photography. One is called "weak" abstraction, in which a photograph contains no easily recognizable objects, though it is clear that one is looking at everyday things (like a detail of a wall). "Strong" abstraction works like weak abstraction, only it is no longer possible to tell what one is looking at, beyond lines, shapes, and colours. Next is “constructed” abstraction, which interferes with the photographic process directly (e.g., in a darkroom, using light, shadow, and chemicals) to produce images that are not “of” anything, but which still might resemble or call to mind certain material textures or natural phenomena we recognize. Finally, we have “concrete” abstraction, which produces something entirely new, “from scratch” (2018, 399), and which refers to nothing outside the processes of photography and the image itself. An artist might go through each of these “stages” of abstraction, either in their career or in the course of creating a single artwork. Obviously, an artist might directly begin by producing “concrete” abstractions that were inspired by no target system outside the artwork, without going through the other “stages”. Still, thinking of it as a process that moves away from a concrete target system will be helpful in what follows.

We have distinguished two abstractive processes, and now we want to suggest that each requires its own epistemological account. We called the process of intentionally leaving certain details out of a representation “subtractive” abstraction. This we find in both scientific model-building and abstract art (as well as non-abstract art). The epistemology of such a process has been accounted for by philosophers primarily using what we might call a “preservative” epistemology. A representation is created and used in an argument for a particular conclusion about a particular target system. For example, a particular pendulum is presented, a model is built of that pendulum, which subtractively abstracts certain features but retains important truths. We find something holds in that model (for example, that the pendulum’s period is proportional to the square root of the length of the string of the pendulum), and we extend this finding to the real pendulum. And this extension is thought to be justified because the model already contains accurate information about that very target. Thus, the model is epistemically preservative. The justificatory force behind the conclusion was just the empirical observations and justified theoretical background knowledge that we already had. What justifies extending our knowledge to *all* pendulums is a separate inductive argument, which says that what we've learned about *this* particular concrete pendulum should hold (roughly) for *all* pendulums, because pendulums are similar in a way that is relevant for induction (for discussion, see Norton 2021). While we want to draw attention to another notion of abstraction, we recognize that there is much more to be said about the practice and epistemology of subtractive abstraction (see, e.g., Suárez, this volume).

We called the second kind of abstractive process "generative abstraction", and this process is more complex, at least insofar as it can contain processes of subtractive abstraction. Abstraction in this sense is a process of creating representations that mostly or completely leave the initial target system behind, to produce artefacts that become a more central focus than the initial target system. These created artefacts may still represent something, and what they represent tends to be expressed in a more "universal" language that can be interpreted from a range of different perspectives. "As David Bohm has observed, the abstract images of Kandinsky's maturity rely for their visual effect only on what is immediately presented: they are considered complete creations in and of themselves by virtue of their inherent structure and qualities” (Berry 2005, 101).

There are at least two ways in which the process of generative abstraction might go: first, in a stepwise manner, moving further and further away from an initial, concrete, inspiring target. In this case, generative abstraction might begin as subtractive abstraction, but it goes beyond this when it severs its ties to the initial system and draws attention to itself and the new target. Second, a generatively abstract representation can be built directly, without any initial, concrete, inspiring target system. What qualifies instances of the second type as instances of abstraction is that the finished product has the same set of epistemological features as the first, specifically, a lack of reference to any initial concrete target system. For example, consider Mondrian’s *Composition B (No. II) with Red*.[6] This painting aimed to represent “the dynamic equilibrium of true life” (Mondrian 1986, 283). We may fairly assume that it was not inspired by a concrete initial target and was created to represent the dynamic equilibrium of life directly. Both processes of generative abstraction are processes of *abstraction* because they take us to “a more abstract place”. And neither can be understood wholly as subtractive abstractions, because in the first case, we eventually cease subtracting and start building, while in the second case, we were never subtracting at all.

The epistemology of abstractive representations cannot be exhausted by a preservative account: it must be supplemented by a *generative* account (for a related distinction between preservative and generative accounts, see Miyazono and Tooming 2022). The question is not about justifying conclusions concerning a single target, but about producing new and epistemically valuable targets, about which our representation can teach us. The goal of the rest of this chapter is to discuss the epistemology of generative abstraction as it appears in science. To do this, we first identify some cases.

6.4 Scientific Models Can Be Generatively Abstract

Nancy Nersessian's book *Interdisciplinarity in the Making: Models and Methods in Frontier Science* (2022) presents the results of more than 15 years of ethnographic research into how scientists make models. The case we want to focus on concerns a series of connected models built in a neuroengineering lab. Very roughly, we might describe their work as follows. The scientists wanted to understand how the brain "learns", which they operationalized in terms of the construction and stabilization of networks of neurons in response to external stimuli and feedback. To investigate this sense of "learning", they turned to studies on the brains of mice. Rather than performing *ex vivo* studies on mice brains, they "harvested" cortical neurons from mouse embryos, separated them to break any existing neural connections, and placed them on top of an 8 × 8 grid outfitted with 60 electrodes. These electrodes were able to provide electrical inputs, receive outputs, and make possible the tracking of neural activity as the neurons established synaptic connections. This model, which could be metaphorically understood as a “brain” on a dish, provided data that the scientists were not able to characterize using known concepts and theory. In response, they built computational models of the dish models. This is characteristic of much scientific work: in response to epistemological and practical problems, the targets and models change together. With each iteration, the target of inquiry shifts, and overall, we claim, the scientists participated in a process of generative abstraction (see Fig. 6.1).

Fig. 6.1

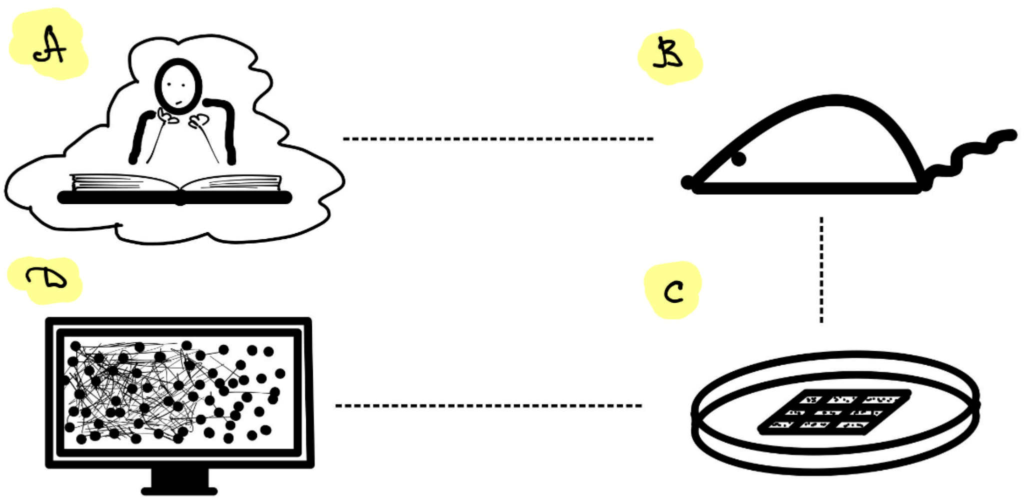

Figure 6.1. A process of generative abstraction. Each transition (from A to B, C, and D) involves abstraction in the subtractive sense, but also introduces some idealizations and additional constraints to create the next vehicle of modelling.[7]

In more detail, the neurons in the dish would fire in response to electrical stimuli. To understand how the dish "learned", some kind of meaningful patterns had to be discerned.

> The in silico model, which might be considered a second-order model, was constructed initially by one researcher in an attempt to understand the spontaneous, dish-wide firing of the neurons ("*burst*" phenomena) that was occurring in the in vitro model and that they assumed was an impediment to progress in the lab's research project of getting the dish to learn.
>
> (Nersessian 2022, 107)

Nersessian claims that "this kind of second-order modeling of built prototypes (which we consider the in vitro dish to be) is a common engineering investigative practice" (106). To overcome the difficulties of using the dish model, a participant in Nersessian's study decided that a new representation was necessary; in his words,

> the advantage of modeling [computational] is that you can measure everything, every detail of the network......I felt that [computational] modeling could give us some information about the problem [bursting and control] we could not solve at the time [using the in vitro dish model-system].
>
> (quoted on 115)

Nersessian points out that this scientist

> felt that to understand the phenomena of bursting he needed to be able to "*see*" the dish activity at the level of individual neurons, to make precise measurements of variables such as synaptic strength, and to run more controlled experiments than could be conducted with the physical dish.

(115)

This participant's long-term goal was to understand the behaviour of the neurons on the dish. But that behaviour, when translated into the computer model, took on a life of its own. This is at least partially because the computer model introduced different constraints from the dish model. Some constraints were built into the computer model from the dish: e.g., an 8×8 grid with 60 electrodes. Some came from the neuroscience literature, e.g., given 1000 neurons, theory predicted there should be about 50,000 synaptic connections. In addition, values for other parameters came directly from the literature, including values for conduction velocity, delay, noise level, and action potential effects, as well as information about which types of synapses there should be, how they should be connected, what percentage of the neurons should be excitatory and what percentage should be inhibitory, and so on. Some constraints came from the modelling software, and finally, some became part of the model via the iterative process of model-building as the algorithm was run over and over again until it produced behaviour similar to the behaviour observed in the dish.

Before going further, it is instructive to ask how we would characterize what the scientist was doing in terms of the traditional definitions of abstraction and idealization. Clearly, the scientist was not merely abstracting in the subtractive sense. Of course, at various stages, details are removed. For example, empirical values are converted into ranges, and a three-dimensional brain structure is collapsed into a twodimensional dish. But other details (e.g., from the neuroscience literature) are added. So the model is both more and less abstract in the subtractive sense, insofar as it contains more and less information about the target system. Are the scientists also idealizing? Presumably some idealizations had to be made, though Nersessian only mentions a few potential cases. For example, when building the computational model, a participant assumed that the neurons would be randomly distributed over the dish, and he admits not being sure if this is the case in the actual dish, though it looks “pretty random” through the microscope (117). Assumptions like this, made for computational tractability, are at least one potential source of idealizations. It therefore seems likely that we’ll be able to use the traditional concepts of abstraction and idealization to help understand what is going on here. But if we stopped there, we would be leaving out the importantly generative aspect of the story.

One way to characterize the generativity of this process is to focus on the different affordances of each model in the chain. The dish model, unlike *ex vivo* brain slices, was dynamic: it changed over time depending on the inputs it received. The computational model was also dynamic, but in a different way: it could be run in infinitely many different configurations, paused, replayed, and restarted, at will. For example, synaptic connection strength, which is a measure of “learning” as they operationalized it, could be measured in the computer model at any time, though it could not be measured in the dish. Additionally, running experiments on the computer model came without any great cost of time or danger of killing the neurons living on the dish, which had to be painstakingly cared for.

Another way to appreciate the generativity here is to focus on how data from each model was visualized (see also Bolinska 2013, 2016; Vorms 2010, 2011; Kulvicki 2010). For example, to understand the outputs of the computer model, a visualization was built. As Nersessian points out, this could have been done in “any number of ways” including some that were very familiar to the research group. However, the participant built a new visualization that was not yet used in the lab: he visualized the model as a network.

The behaviour of the *in silico* dish could now be shown to the entire research group, who quickly recognized that its behaviour was “novel and distinct from anything they had thus far understood about in vitro dishes” (121). This mode of visualization made new phenomena visible because the computer model tracked the activity of individual “neurons”, which made the propagation of neural activity more clearly visible. A number of “burst types” were then identified: “*you get some feeling about what happens in the network – and what I feel is that... the spontaneous activity or spontaneous bursts are very stable*” (quoted on 121). This transformed the target of their research: before, bursts were noise; now, they are patterns to be investigated and employed. Around ten kinds of bursts were identified, and a new concept, the burst vector, was introduced. This became the new target: directional “waves” of “neural” “activation”. Importantly, the scientists “had the information always... the information was always there" (quoted on 126–7), but it was hidden in the raw data. The computer model made it visible. This might be thought of analogously to an abstract artist who sees something worth investigating in the lines or shapes of a scene, and who produces a painting that brings that aspect out, aided (not frustrated) by the fact that the inspiring scene is no longer visible in the painting. The dynamic, functional qualities of the *in vitro* neural behaviour were brought out by the computer model in a way that allowed the researchers to "*look inside the dish*" (127).

At this point, someone might object that abstract art is supposed to be non-representational, and this computer model (like the dish model) is clearly representational. However, to repeat, abstract art is abstract in that it leaves behind the *original* inspiring visual target (if there was one). It can and often does make reference to new targets or systems of interest, which it might do by inventing targets of its own. So the question is not whether this chain of models represents *something* but whether they represent *the same thing* all the way through.

Someone could claim that they all represent the same thing: "learning". Despite this being the way scientists might frame their work in grant proposals or paper abstracts, "learning" is clearly not the main, or only target represented in all the models. We might think of the first target system as stable patterns of neural signalling including feedback loops in real brains. The second might be the same, in rat brains. The dish they used drew on single-neuron work as well as work on rat brain slices. They produced a physical dish model that was only one layer of neurons thick and fed by a bath of chemicals and kept to a uniform temperature. It only used cortical neurons, since these are the most adaptable. In the dish model, electrodes were attached to the neurons, and after about two weeks, neural circuits grew around these electrodes. The input the dish neurons received might simulate perception and haptic feedback, as the lab would connect the input and output to computer models of virtual environments or physical robots. For example, one dish was modelled after moths, as they tried to teach it to focus on a central "light" source (Nersessian 2022, 75, 128). Another was modelled after a human arm, which they gave a pen and a camera and tried to teach it to colour in between the lines. The target of the computer model was the behaviour of the dish model. This is clearly a different target than the behaviour of real brains. The computer model was further used to model the behaviour of the moth-dish, or the arm-dish, and was used successfully to "program" both. Clearly, this is an episode of scientific progress, despite it not being exhaustively characterizable as growing understanding about a single, original, target. The target moves, and this explains the difficulty that the scientists had in describing what the computational model was a model of. Their answers ranged from a model of learning, to a model of cortical neurons, to a model of itself (Nersessian 2012).

Further, the universalizing ambitions of early abstract artists can also be seen here, as the computer model helped to

> form a global perspective on the phenomena – a perspective that cannot be obtained from the more limited in vitro and real-world experimental possibilities of the target

> system. This global perspective is what informs the "*feeling for the model*," that [the participant] expressed, and that is ubiquitous among modelers.
>
> (Nersessian 2022, 134)

Nersessian calls the perspective "global" because the computational model offers something that can be applied much more broadly than to a particular brain-on-a-dish. It aims to reveal features of neural burst behaviour and neuronal networks more generally, just as abstract art aims to reveal features about emotion, experience, or aesthetic relationships more generally.

We mentioned above that there are at least two ways to practice generative abstraction. The first is by gradually leaving the initial concrete target system behind, ending with a new artefact and a new target that is typically more general (or more conceptual, or more universal). The second way is to develop such a representation directly, without any initial concrete system to serve as the starting point from which information would be subtracted. Above we presented an example of the first sort. Examples of the second sort include some of those with no original concrete target system, like those in synthetic biology in which computational or material systems are built to have certain functions that are found in no living system (see, e.g., Knuuttila 2021; Knuuttila and Koskinen 2021; Knuuttila and Loettgers 2021). One interesting case is the repressilator, which is a circuit of genes that turn each other on and off using proteins in a way that mimics the game of rock-paper-scissors. As in the case above, in attempting to understand this model scientists built a computational model, as well as an electrical analogue that uses voltages to represent protein concentrations (Knuuttiila 2021). Other examples include minimal cells and alternative genetic systems. These systems likewise represent general ways things could be, without having been inspired by any particular concrete system (Knuuttila and Koskinen 2021). Perhaps we could also include exploratory models that target possible or hypothetical systems, like Maxwell's ether model and supersymmetric particle models (Gelfert 2016, Massimi 2019). We think that all of these cases are well described as instantiating processes of generative abstraction. They are clearly abstract in some way, but not because they omit information about a particular system.[8]

One final point of clarification. We have defined generative abstractions in terms of new artifacts and new targets, but we want to be clear that each can be "new" in different ways. Thus, the new artifact is typically new in the sense of "previously not existing". The target of representation (what the model "points at") might be new in that sense, as in the minimal cell, but it could instead be new merely in the sense of being different from the original target. For example, abstract artists might produce a generative abstraction that refers to a feeling or state of being, while the neuroengineers above produced a generative abstraction that refers to neural behaviours, each of which already existed.[9]

6.5 Discussion

6.5.1 General Epistemological Considerations

In considering the epistemology of abstraction, philosophers have focused on a subtractive notion of abstraction. As a result, philosophers have pursued a preservative epistemology of abstraction. There are several ways this could go. Here is one: abstraction is done properly when the only features not abstracted away are the true “difference makers”. Thus, models should be as abstract as possible (as long as they capture all the relevant difference makers), so that they are maximally cognitively tractable for humans (e.g., Strevens 2008). On such an account, the best abstractions are those that don’t interfere with the truth, and which can easily be un-done.

We agree with Nersessian (2008), Leonelli (2008), Carrillo and Martínez (2023), and many of the contributors to this volume that subtractive abstraction is not the only kind of abstraction relevant to scientific model-building. One other important kind of abstraction is generative abstraction. This kind of abstraction is easy to spot when there are chains of connected models, especially in interdisciplinary contexts. Given the fact that generative abstraction is very different from subtractive abstraction, we should expect generative abstraction to require a different epistemology.

For one thing, generative abstractions, unlike subtractive abstractions, cannot be un-done. Adding information from the initial target system “back” into a model is not a sensible thing to do once the target has changed. This is even clearer in the case of generative abstract models that were not inspired by a particular concrete target system. For example, when scientists are trying to build cells that reveal how minimal a genome can be while maintaining core cellular functions, it would not be helpful to introduce information from a particular cell, like a human liver cell, into that minimal cell. Such information would not make the minimal cell a better representation of its target, which is the *minimal* cell. It would make it a *worse* representation of a minimal cell, because the extra information would make it less minimal.

Rather than staying true to some inspiring target system, generatively abstract models should be interesting as artefacts in themselves. They may still (come to) refer to things, and these things should be of scientific interest. In the neuroscience case discussed above, the target shifted from the human brain to the mouse brain to the dish model to the computational model. Unlike subtractive abstraction, which ideally increases understanding about the initial target system, generative abstraction explores new features, like bursts and burst vectors, which may or may not be found in the original target.

Generative abstraction is successful to the extent that it improves our epistemic standing with respect to the new target. Just as abstract artists manage to explore musical harmony (Kandinsky), feelings (Mondrian, Rothko), or religious themes (af Klint) via lines, shapes, and colours, scientists invent new systems with interesting properties, just as the computer model discussed above enabled scientists to “see significant system behaviors” (Nersessian 2022, 137). But how is the process of generative abstraction done *well*? What could we say to a scientist embarking on such a process? This is a difficult question. Often the abstractor might desire to leave the initial target behind without knowing what the final artefact should be or what the final target system should be. Or they might know what they want the final target to be, but they don’t know what to build in order to explore it.

Generative abstractions are useful because they enable scientists to break away from what may be a limiting focus on a particular target system. Because of this, the practitioner has a lot of flexibility in creating them. This might be frustrating for someone who wants to follow a set of rules to build a generative abstraction, but there cannot be such a set of rules. For Kandinsky, the process should be intuitive, as well as slow, careful, and rational: “Reason, consciousness, purpose, and adequate law play an overwhelming part. Yet, it is not to be thought of as a mere calculation, since feeling is the decisive factor” (Kandinsky 1977, 108, 109, 117, 123). For Malevich, it should be a completely rational process, carefully planned out from the beginning: “In constructing painterly forms it is essential to have a system for their construction, a law for the constructional inter-relationships of forms” (Malevich 1969, 100). However, Malevich never gave rules for such a system.

Unlike subtractive abstraction, where scientists can (in simple cases) chip away irrelevant information bit-by-bit until an explanatory kernel of dependency relations is revealed, scientists abstracting generatively must be permitted to move playfully, adding detail here, removing detail there, building up and breaking down, as they try to create something new that is interesting and useful. Nersessian argues that the choices they make are not necessitated: there are always many equally rational moves to make. All that is required is that each step of the model-building process must be justifiable from the present perspective. This way, the process may be rational, as Kandinsky, Mondrian, and Malevich demand, even without a foolproof method that could be specified in advance.

Perhaps this coheres best with a consequentialist epistemology. Such an account would judge a process of generative abstraction based on the epistemic quality of the output (Stuart 2022a, 2022b). What makes one output better than another? Generative abstraction is a way to build models, and models have many epistemic uses. So a process of generative abstraction will be better to the extent that it contributes to some epistemic aim, for example, providing a good starting point, providing a proof-of-principle demonstration, generating a potential explanation, leading to an assessment of the suitability of a target, delivering knowledge of causal possibilities, or delivering knowledge of objective possibilities for hypothetical entities (see Gelfert 2016 and Massimi 2019). An internist version of this idea would claim that a process of generative abstraction has more epistemic value to the extent that, as far as the abstractors can foresee, it would best promote some of the above epistemic aims. An externalist version would claim that a process of generative abstraction has more epistemic value to the extent that it really turns out to best promote such epistemic aims. It might also be possible to formulate a deontic epistemology of generative abstraction, such that a process of generative abstraction is epistemically correct when each act that makes it up respects duties of maintaining representational accuracy or staying consistent with background knowledge. But given the artistic, experimental, and imaginative nature of generative abstraction, perhaps developing and obeying strict rules would not be the main way that scientists (should) perform and justify their work (Stuart 2020).

Another way to explore the epistemic powers of generative abstraction is to make reference to existing epistemologies of scientific representation, which explain how representations produce new knowledge or understanding of their targets. While it would be interesting to see how this might go in detail for each account, doing so would not give any definite answer about how we should understand the epistemological contributions of generative abstraction until it was clear which of these accounts was the correct one. Thus, structuralists (e.g., da Costa and French 2003; Bueno, French, and Ladyman 2002) can explain successful generative abstractions by reference to homomorphisms, monomorphisms, isomorphisms, or partial isomorphisms that obtain between the model and the new target. Inferentialists (e.g., Suárez 2004) can argue that generatively abstract models succeed when they enable useful inferences to be drawn about the new target. Interpretationalists (e.g., Contessa 2007, 2011) can argue that generatively abstract models are those that can be interpreted to be about the new target. Each of these accounts produces explanations concerning how generative abstractions work, but which explanation is to be preferred depends on which account is correct, and that is still very much an open question.

However, it might be interesting to turn the question around and use the existence of generative abstraction as a test for accounts of representation. If generative abstraction is a genuine and important part of science, then accounts that have difficulty accommodating it face a challenge. Consider the two main kinds of fictionalist accounts of scientific representation. Fictionalists claim that a model is a fiction in the sense that it constrains “games of make believe” that we can play. To play such a game, we recognize certain implicit and explicit rules, as well as “props” around which the game is focused. The model, or the model description, serves as a prop in the game, and our goal is to determine what else is true in the fiction. *Indirect* fictionalists claim that model descriptions inspire the creation of imaginary systems which can be explored and compared to a target system. These accounts are indirect in the sense that they claim we learn about the real target system by means of a third thing, the imaginary system. We think that indirect fictionalist accounts like Frigg and Nguyen’s (2016, 2020) can handle generative abstract models since the last stage of model-based reasoning in their view freely “keys up” properties instantiated in the imaginary system with properties that are to be attributed to the target, and as far as we can tell, nothing in their account prevents that target from being different than the initial target system on which the model was originally based. However, there are also *direct* fictionalists, who claim that the model is always about some real world target system. There is no “third thing”, no fictional system, through which our investigation detours. For example, a mathematical model of a pendulum is always and only about a specific desk pendulum, or the set of all actual pendulums (Toon 2012; Levy 2012, 2015). This seems to require that models created via abstraction must only tell us about some particular real world target system. Generatively abstract models still count as representations on this account, as they prescribe imaginings in the context of a game of make believe. However, as scientists go through a process of generative abstraction in their model-building, they change target or produce new targets. When this happens, the resulting model becomes either a *bad* representation of the initial target, or we must ignore the process of model-building and simply say of the finished model that it represents the new target. The first option is unattractive because the process of generative abstraction can create *better* models, not just worse ones. And this improvement is substantial: generative abstraction helps scientists to achieve particular epistemic aims, like explanation, prediction, and opening up new theoretical possibilities. To ignore this and claim that any changes in the target inevitably make the model worse would be to ignore all the good reasons that scientists have for making generative abstractions. The second option is unattractive because attention to scientific practice makes it clear that very few models are really “complete” such that we can say once and for all what the “real” target is or should be. We want an account of scientific representation that can accommodate the flexibility and open-endedness of models and model-building practices. Genuinely moving targets therefore present a challenge to direct fictionalism as a descriptive and normative account of scientific practice.

6.5.2 Considerations for the Scientist

What consequences, if any, does the existence of generative abstraction have for the practicing scientist? One main consequence is that generative abstraction should be kept distinct from subtractive abstraction, even in the mind of the scientist. This is because the different kinds of abstraction are justified in different ways, and scientists should keep track of which past scientific actions were justified and how.

Recall that subtractive abstraction can be used for generative purposes. This is because subtraction has a capacity to *reconfigure* a model, its workings, internal commitments, constraints, and representational relations to the target. Scientists should be aware that with enough subtractive abstraction, they might find themselves *generating*. Once they have embarked down this path, a generative rather than preservative justification will be required, and employing only a preservative epistemology will yield incorrect evaluations of past practice.

To see how an abstraction can reconfigure the model landscape, let us look back at the dish model. From a theoretical perspective, moving from an intact mouse brain to a dissociated mouse brain seems like a mere abstraction of constraints or features that define the functioning of a brain, such as the quantity of connected neurons and their spatial organization. On the dish, we are dealing with a smaller number of neurons arranged in two rather than in three dimensions, so it is easier to study. Here, the problem is that subtraction in the theoretical sense may be not the same as in the material sense: what theoretically looks like removal of some constraints or features in fact results in a substitution of one set of constraints for another, which may reshape the behaviour of the model in an unanticipated way. On the positive side, new or unnoticed effects or phenomena may surface, such as the phenomenon of bursting; however, on the negative side, it may produce contingencies and artefacts that may not have anything to do with the original target. (They may still be interesting effects to study by themselves, and this is one source of generativity.)

One may reply that such a danger is only pertinent to abstractions in the case of material models. We disagree: subtractions can turn into reconfigurations also in the case of conceptual, mathematical, or computational models, because even if abstraction in the model may be theoretically tractable, its effect on what the target is, generally speaking, isn’t always foreseeable. There probably are no cases where we could *guarantee* that further subtractive abstraction would not lead to generative abstraction. Perhaps, instead, there are some contexts in which this is not a very serious epistemic risk. For example, as long as we have a firm intention to fix the target, and the model is relatively simple, the effects of greater and greater subtractive abstraction can be tracked. Common pedagogical uses of the mathematical model of the pendulum are one example. But such a situation seems to be an exception, not the norm, in the scientific practice of model-building and model-using. And this is why supposedly small subtractions in mathematical models or simulations might not work as innocently as one expects them to. Consequently, their effect cannot be just “un-done”: one may not know in advance if they abstract contingent features of some real difference makers. Thus, subtraction should always come with evaluations of the effect of this subtraction, as cutting off the “wrong wire” may lead to (good or bad) unpredicted epistemic consequences, which require a new kind of epistemological justification. We are not suggesting that all generative abstractions must be planned in advance or done with conscious foresight. We are merely pointing out that generative abstraction and subtractive abstraction have different epistemic features and yet one can easily lead to the other, so it would be epistemically irresponsible to pretend that subtractive abstraction alone exists, especially when the stakes are high. Scientists should keep track of their targets, even, or especially, when those targets are being brought into and out of existence.

6.6 Conclusion

Our main goal in this chapter has been to add to the existing repertoire of concepts for describing scientific practice. Generative abstraction is something that scientists do, and it is worth looking at more closely. Focusing on this sense of abstraction is helpful in celebrating the complexity of the practice of crafting scientific models. A secondary contribution is to consider the epistemology of this way of model-building. One upshot is that “abstraction” is not always reversible, since only subtractive abstraction is (arguably) reversible. We could recover the traditional way of speaking by disqualifying generative abstraction as a kind of abstraction. But given the intuitively abstract nature of its outputs and its connection to abstract art, this would require argument.

Generative abstraction raises new questions. One is how generative abstraction relates to idealization. Do generative abstractions introduce intentional misrepresentations of their targets? In some cases, the model system will no longer function as a representation of a particular inspiring target system. Instead, it will *become* a target system. In that case, like the concrete abstractions discussed above, the model cannot misrepresent, since it only represents itself. For example, consider genetic variants of a model organism. Each new genotype has a specific (and different) part of the wildtype phenotype as its original target, yet it is studied for its own features and is not clearly a misrepresentation of anything. In other cases, a generative abstraction will require building a representational analogy base that starts from a particular concrete system, and as we noted above, this can include the use of idealizations. However, as the target of the model changes, what were once idealizations (in the sense of misrepresentations) can become accurate representations.

What this suggests is that generative abstraction is not identical to idealization; however, much more can and should be said about this connection.

A second question concerns the differences between generative abstraction in art and science. One interesting historical difference is that abstract art was sharply criticized for its idea that shapes and colours could really serve as a "universal language". In science, the idea that generative abstractions are more universal might be appealing, since such models are typically more formal, more mathematical, and more likely to "travel" across disciplinary boundaries. There are surely other informative disanalogies between the two contexts that would be worth exploring.

A final question concerns whether generative abstraction arose first in art and then travelled into science, or vice versa, or whether it arose independently in both.[10] A natural thought is that generative abstraction arose first in art. However, there are scientists who made generative abstract visualizations already in the 1890s, like W.E.B. DuBois, whose visualizations were said to "anticipate Kandinsky's famous Bauhaus color and shape tests administered to his students decades later" (Battle-Baptiste and Rusert 2018, 97; Phull forthcoming).[11] If the arrow of historical connection runs from science to art, this would help to explain why generative abstractions are found so readily in science. If it arose independently in both, this would suggest that similar problem-solution pairs arise in both science and art, which could support continuum theorists about science and art. In any case, exploring and comparing the history of generative abstraction in both science and art will be imperative for learning more about the kinds of problems that generative abstraction has been used to solve, and where it has been, and can be, more or less successful.

Acknowledgements

The authors would like to thank Julia Sánchez-Dorado and Chiara Ambrosio for extensive comments on a draft of this chapter, as well as the other contributors to this volume who attended a preparatory workshop in April 2023. Mike Stuart also thanks the audience at Philosophy of Science Association Around the World Conference held in November 2023, Nancy Nersessian, and Michela Massimi for comments.

References

Barr, Alfred H. Jr. 1936. *Cubism and Abstract Art*. London: Routledge.

Battle-Baptiste, Whitney, and Britt Rusert, eds. 2018. *W. E. B. Du Bois's Data Portraits: Visualizing Black America*. Illustrated edition. Amherst, MA: Hudson, NY: Princeton Architectural Press.

Berry, Kenneth. 2005. "The Paradox of Kandinsky's Abstract Representation." *Journal of Aesthetic Education* 39 (1): 99–104.

Bolinska, Agnes. 2013. 'Epistemic Representation, Informativeness and the Aim of Faithful Representation.' Synthese 190 (2): 219–34.

———. 2016. ‘Successful Visual Epistemic Representation.’ Studies in History and Philosophy of Science Part A 56 (April): 153–60.

Bueno, Otávio, Steven French, and James Ladyman. 2002. “On Representing the Relationship between the Mathematical and the Empirical.” Philosophy of Science 69 (3): 497–518.

Carrillo, Natalia, and Sergio Martínez. 2023. “Scientific Inquiry: From Metaphors to Abstraction.” Perspectives on Science 31 (2): 233–261.

Chilvers, Ian, and John Glaves-Smith. 2009. *A Dictionary of Modern and Contemporary Art*. *A Dictionary of Modern and Contemporary Art*. Oxford: Oxford University Press.

Contessa, Gabriele. 2007. “Scientific Representation, Interpretation, and Surrogative Reasoning.” Philosophy of Science 74 (1): 48–68.

———. 2011. “Scientific Models and Representation.” In *The Continuum Companion to the Philosophy of Science*, edited by Steven French and Juha Saatsi, 120–37. London and New York: Continuum Press.

Costello, D. 2018. “What is Abstraction in Photography?” British Journal of Aesthetics 58 (4): 385–400.

Da Costa, Newton, and Steven French. 2003. *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*. Oxford: Oxford University Press.

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

Frigg, Roman, and James Nguyen. 2016. “The Fiction View of Models Reloaded.” The Monist 99 (3): 225–42.

———. 2020. *Modelling Nature: An Opinionated Introduction to Scientific Representation*. Synthese Library. New York: Springer.

Gelfert, Axel. 2016. *How to Do Science with Models*. Dordrecht: Springer.

Gertsman, Elina, ed, 2021. *Abstraction in Medieval Art: Beyond the Ornament*. Amsterdam: Amsterdam University Press.

Godfrey-Smith, Peter. 2009. “Abstractions, Idealizations, and Evolutionary Biology.” In *Mapping the Future of Biology: Evolving Concepts and Theories*, edited by Anouk Barberousse, Michel Morange, and Thomas Pradeu, 47–56. Boston Studies in the Philosophy of Science. Dordrecht: Springer Netherlands.

Goodman, Nelson. 2003. *Abstraction*. In Grove Art Online. Oxford University Press. www.oxfordartonline.com/groveart/display/10.1093/gao/9781884446054.001.0001/oao-9781884446054-e-7000000257

Jansson, Lina, and Juha Saatsi. 2019. “Explanatory Abstractions.” The British Journal for the Philosophy of Science 70 (3): 817–44.

Kandinsky, Wassily. 1977. *Concerning the Spiritual in Art*. Translated by M. T. H. Sadler. Revised edition. New York: Dover Publications.

Khalifa, Kareem. 2017. *Understanding, Explanation, and Scientific Knowledge*. New York: Cambridge University Press.

Knuuttila, Tarja. 2021. "Epistemic Artifacts and the Modal Dimension of Modeling." European Journal for Philosophy of Science 11 (3): 65.

Knuuttila, Tarja, and Rami Koskinen. 2021. "Synthetic Fictions: Turning Imagined Biological Systems into Concrete Ones." *Synthese* 198 (9): 8233–50.

Knuuttila, Tarja, and Andrea Loettgers. 2021. "Biological Control Variously Materialized: Modeling, Experimentation and Exploration in Multiple Media." Perspectives on Science 29 (4): 468–492.

Kulvicki, John. 2010. 'Knowing with Images: Medium and Message'. Philosophy of Science 77 (2): 295–313.

Lawler, Insa. 2019. "Scientific Understanding and Felicitous Legitimate Falsehoods." Synthese 198 (7): 6859-6887.

Leonelli, S. 2008. "Performing Abstraction: Two Ways of Modelling Arabidopsis Thaliana". Biology & Philosophy 23: 509–28.

Levy, Arnon. 2012. "Models, Fictions, and Realism: Two Packages." Philosophy of Science 79 (5): 738–48.

———. 2015. "Modeling without Models." Philosophical Studies 172 (3): 781–98.

———. 2021. "Idealization and Abstraction: Refining the Distinction." Synthese 198 (S24): 5855–72.

Malevich, Kasimir Severinovich. 1969. *Essays on Art, 1915-1933*, Vol. 1, edited by Troels Anderson. London: Rapp & Whiting.

Massimi, Michela. 2019. Two Kinds of Exploratory Models. Philosophy of Science, 86(5), 869-881. DOI:10.1086/705494.

McMullin, Ernan. 1985. "Galilean Idealization." Studies in History and Philosophy of Science 16: 247–73.

Miyazono, Kengo, and Uku Tooming. 2022. "On the Putative Epistemic Generativity of Memory and Imagination." In *Philosophical Perspectives on Memory and Imagination*, edited by Anja Berninger and Íngrid Vendrell Ferran, 127–45. New York: Routledge.

Mondrian, Piet. 2007. *Plastic Art and Pure Plastic Art, 1937, and Other Essays, 1941-1943*. New York: Wittenborn Art Books.

Nersessian, N. 2008. *Creating Scientific Concepts*. Cambridge: MIT Press.

———. 2012. "Modeling Practices in Conceptual Innovation: An Ethnographic Study of a Neural Engineering Research Laboratory." In *Scientific Concepts and Investigative Practice*, edited by U. Feest and F. Steinle, 245–69. Berlin: DeGruyter.

———. 2022. *Interdisciplinarity in the Making: Models and Methods in Frontier Science*. Cambridge: MIT Press.

Nguyen, J. 2020. "It's Not a Game: Accurate Representation with Toy Models." The British Journal for the Philosophy of Science 71(3): 1013–1153.

Norton, John D. 2021. *The Material Theory of Induction*. Calgary: BSPS Open.

Phull, Kiran. Forthcoming. “Visualizing Hope in the Radical Data Work of W. E. B. Du Bois.” In *The Politics of Hope in the Anthropocene: Agency, Governance and Critique*, edited by Valerie Waldow, Pol Bargués, and David Chandler. Edinburgh: Edinburgh University Press.

Portides, Demetris. 2021. “Thinking about Idealization, Abstraction and Scientific Models: An Introduction.” Synthese 198 (24): 5849–53.

Strevens, Michael. 2008. *Depth: An Account of Scientific Explanation*. Cambridge, MA: Harvard University Press.

Strevens, Michael. 2017. “How Idealizations Provide Understanding.” In *Explaining Understanding: New Essays in Epistemology and the Philosophy of Science*, edited by Stephen Grimm, Christoph Baumberger, and Sabine Ammon. New York: Routledge.

Stuart, Michael T. 2020. “The Productive Anarchy of Scientific Imagination.” Philosophy of Science 87 (5): 968–78.

———. 2022a. “Scientists are Epistemic Consequentialists about Imagination.” Philosophy of Science. https://doi.org/10.1017/psa.2022.31

———. 2022b. “Sharpening the Tools of Imagination.” Synthese. https://doi.org/10.1007/s11229-022-03939-w

Suárez, Mauricio. 2004. “An Inferential Conception of Scientific Representation.” Philosophy of Science 71 (5): 767–79.

Thomson-Jones, Martin R. 2005. “Idealization and Abstraction: A Framework.” *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 86 (12), 173–218.

Toon, Adam. 2012. *Models as Make-Believe: Imagination, Fiction and Scientific Representation*. New Directions in the Philosophy of Science. Basingstoke: Palgrave Macmillan UK.

Vargish, Thomas, and Mr Delo E. Mook. 1999. *Inside Modernism: Relativity Theory, Cubism, Narrative*. New Haven, CT: Yale University Press.

Verreault-Julien, Philippe. 2022. “Representing Non-Actual Targets?” Philosophy of Science 89 (5): 918–27.

Vorms, Marion. 2010. ‘The Theoretician’s Gambits: Scientific Representations, Their Formats and Content.’ In *Model-Based Reasoning in Science and Technology: Abduction, Logic, and Computational Discovery*, edited by Lorenzo Magnani, Walter Carnielli, and Claudio Pizzi, 533–58. Studies in Computational Intelligence. Berlin, Heidelberg: Springer.

Vorms, Marion. 2011. ‘Representing with Imaginary Models: Formats Matter.’ Studies in History and Philosophy of Science Part A, Model-Based Representation in Scientific Practice 42 (2): 287–95.

Yablo, Stephen. 2020. “Models and Reality.” In *The Scientific Imagination*, edited by Arnon Levy and Peter Godfrey-Smith, 16–50. Oxford: Oxford University Press.

---

[1] As Julia Sánchez-Dorado helpfully reminded us, Kandinsky himself rarely removed *all* traces of material objects, as such traces often served his goals as an abstract artist.

[2] Each of the four backs can be viewed via the Museum of Modern Art (New York). The first: http://bitly.ws/PeBB, the second: http://bitly.ws/PeBI, the third: http://bitly.ws/PeBR, the fourth: http://bitly.ws/PeBV

[3] This image can be viewed through its current holder, Sotheby’s here: http://bitly.ws/PeDH

[4] See, for example, his *Improvisation Auf Mahagoni* (*Improvisation on Mahogany*), 1910, which can be viewed through its current holder, Sotheby’s, here: http://bitly.ws/PeE8

[5] For example, *Avond: De rode boom* (*Evening: The* red *tree*) (1908-10), *De grijze boom* (*Grey tree*) (1911), *Bloeiende appelboom* (*Blossoming Apple Tree*) (1912). All three paintings can be viewed on the Hague Art Museum website: *De rode boom* (http://bitly.ws/PeHK); *De grijze boom* (http://bitly.ws/PeIj); *Bloeiende appelboom* (http://bitly.ws/PeKm)

[6] View this artwork here: http://bitly.ws/PAFN

[7] Taking up "learning" as our first target, we move to the mouse brain; then we move away from the complexity and intricacies of the mouse brain by dissociating mouse neurons onto a dish – the resulting arrangement of cells is easier to control and less complex; finally, the dish is simulated in the computer to yield even more control of the input parameters while providing access to the internal processes that possibly underpin the neural interactions relevant for cognition. Each model shifts its target away from the original target of human learning (to learning in mice, to behaviours of the dish, to the interaction of computational variables based on mathematical premises), and in so doing generates a new landscape of affordances, allowing scientists to pose new questions that were not possible for previous models.

[8] For some other excellent examples of this second kind of generative abstraction, see Costello, this volume.

[9] Thanks to Michela Massimi for prompting us to think more about this..

[10] For excellent work on this connection generally, see the entries in this volume by Sánchez-Dorado, and Tarja Knuuttila, Hanna Johansson and Natalia Carrillo.

[11] To view these visualizations, see the Library of Congress's collection, here: http://bitly.ws/PAMw