De-Idealization - No Easy Reversals

Tarja Knuuttila
University of Vienna
Mary S. Morgan
London School of Economics

Abstract

De-idealization as a topic in its own right has attracted remarkably little philosophical interest despite the extensive literature on idealization. One reason for this is the often implicit assumption that idealization and de-idealization are, potentially at least, *reversible* processes. We question this assumption by analyzing the challenges of de-idealization within a menu of four broad categories: de-idealizing as *re-composing*, de-idealizing *as re-formulating*, de-idealizing as *concretizing*, and de-idealizing as *situating*. On closer inspection, models turn out much more inflexible than the reversal thesis would have us believe—and de-idealization emerges as a creative part of modeling.

1. Introduction

Implicit in the philosophical notion of idealization is a certain idea of model construction: that models are arrived at, among other ways, through a *process*, or processes, of idealizing. The presence of such a process becomes manifest as soon as the question of de-idealization is raised. Namely, idealization and de-idealization are assumed to be, potentially *reversible* processes. Models are thought by philosophers of science to be built by making use of processes of distortion, omission, abstraction and approximation that are variously included in the category of idealization, depending on the philosophical theory in question. The different theories tend to assume that the idealizing features that make models less accurate representations of their targets can be, at least in principle, reversed, setting up de-idealization as an opposite process (or set of processes) to that of idealization. Consequently, one central question within recent philosophical discussions of idealization concerns precisely the desirability of de-idealization—thus giving de-idealization a pivotal role in distinguishing between different notions of idealization (Nowak 1992, 2000; Weisberg 2007; Elliott-Graves and Weisberg 2014).

Yet, despite the extensive literature on idealization and despite the fact that the question of de-idealization proves critical for many accounts of idealization, de-idealization as a topic in its own right has not succeeded in attracting much explicit philosophical interest. In this article, we zoom in on this lacuna and study the challenges modelers face when attempting to de-idealize their models. We have grouped such challenges within four broad categories: de-

idealizing as *re-composing*, de-idealizing as *re-formulating*, de-idealizing as *concretizing*, and de-idealizing as *situating*. Analyzing these challenges suggests that the idea of de-idealization as a reversal process is both overly simplified, and frequently misguided.

Our account, by taking the processes of de-idealization seriously, highlights certain representational, conceptual and methodological complexities of modeling that are often overlooked in treatments that focus only on idealization. De-idealization is crucial for different kinds of attempts to apply models to the world—by empirical use with statistics, in designing experiments, or in making arguments about concrete events. Such attempts unquestionably involve de-idealization. But de-idealization is frequently involved also in the use of models for theorizing in different contexts and different ways. In the following, we analyze the processes of de-idealization by drawing on examples from economics and the existing literature in the philosophy of economics on model application problems (e.g. Alexandrova 2006; Boumans 2005; Reiss 2008; Svetlova 2013). Concentrating on one discipline provides a more synoptic view than a collection of examples drawn from a multitude of disciplines. Economics provides, furthermore, a good subject for studying de-idealization by bestowing a rich repository, and relatively long history, of examples of idealization and de-idealization. Also, and equally importantly, economics as a discipline faces the expectations of both policy makers and citizens, making de-idealization of economic models both habitual and challenging.

2. What Idealization Implies about De-Idealization

It is generally agreed among philosophers of science that most if not all models involve idealizations.[1] Being idealized, models give us inexact, partial, inaccurate, or distorted depictions of reality, and due to such limitations, one influential defense of idealization is provided precisely by the possibility of de-idealization. According to this defense, as a science 'advances', the various simplifications and distortions effected by idealizations will be corrected, thus making the theoretical representations more concrete or realistic (e.g. McMullin 1985; Nowak 2000). In these discussions, one can notice a subtle slide between the notion of idealization as a process of model formation (by distorting, simplifying or abstracting, etc.) and idealization as a quality of models (models are simple, inaccurate or distorted, mathematically formulated etc.). This ambiguity, between whether one is talking about the process or an outcome object called 'the model', serves to hide the implicit assumption that models gain the quality of being idealized because of the processes that formed them.

Earlier discussion was more explicit. Nowak (e.g. 1992, 2000) offers a classic example of the supposed reversibility of both the processes of idealization and de-idealization, with the

---

[1] We are focusing on formal models, but much of our discussion also applies to material models.

de-idealizing move being called concretization. In Nowak's view, scientific breakthroughs were brought about by the method of idealization that seeks to set apart, and study, the dependencies between the most relevant magnitudes or essential components. He analyzed the work of e.g. Galileo, Darwin and Marx, and claimed that the success of these theories was due to their proficient use of idealization. However, the application of an idealized theory would require a return from the ideal world to the real-world phenomena through the procedure of concretization. By eliminating step by step the idealizing conditions, more realistic statements could be achieved. For Nowak, this concretization process is usually completed by approximation: "Normally, after introducing some corrections the procedure of approximation is applied. That is, all idealizing conditions are removed and their joint influence is assessed as responsible for the deviations up to a certain threshold ε." (Nowak 1992, 12). In other words, Nowak assumed that in the mature natural sciences the approximative structure eventually replaces the idealized structure.

In these accounts, idealization, conceived as a process, was considered to cover various different kinds of strategies of simplifying, of leaving aside some components, and of abstraction into mathematical form.[2] In the more recent discussion, these terminologies have

[2] The classic starting point is usually considered to be McMullin, 1985. Nowak (2000), however, does distinguish between idealization as distortion and abstraction as a conceptualizing move (see 3.iii below) in his application of the dialectical method to the

changed their sense such that now a clear distinction is often made between idealization and abstraction. While idealizations are thought to distort the features of the (real-world) target system in the simplification process, abstraction has been interpreted in terms of subtraction, that is, of omitting some of these features, or causal factors (e.g. Cartwright 1989; Nowak 2000; Godfrey-Smith 2009; Jones 2005; Levy and Bechtel 2013). 'Abstractions', then, are thought to give veridical, although partial, representations, whereas 'idealizations' depict something differently than what is known, or assumed to be, the case. Yet, even in this more articulated framework, the processual character of idealization is preserved. For example, Levy (2018) addresses the "process/product ambiguity" with respect to idealization, claiming that in understanding idealization the intentions of the modeler and the process of model construction are crucial; idealization involves "a deliberate introduction of falsehood into a representation".[3]

The distortions introduced by idealization have been motivated and justified on different grounds, yet hinging simultaneously on the possibility/desirability of de-idealization. Michael Weisberg (Weisberg 2007; Elliott-Graves and Weisberg 2014) has distinguished between 'Galilean idealization' (which relies on the possibility of de-idealization) and 'minimalist

---

question of idealization.

[3] Levy (2018) also mentions "concretization" but does not expand on this theme.

idealization' (which does not, and so has to be justified on different grounds).[4] The pivotal role de-idealization occupies within this distinction becomes apparent once we consider Galilean and minimalist idealizations a bit more closely.

Galilean idealizations are primarily introduced to make a model tractable for computational and other purposes, the ultimate goal being the de-idealization of the model. The simple, hypothetical model, arrived at by idealization, is in the subsequent research rendered more accurate by de-idealizing it. Thus, Galilean idealizations are thought to be corrigible in that they are supposed to be, at least in principle, reversible by adding back real-world details and correcting the distorted features.[5] From the epistemic point of view, however, as Batterman (2009, 16) points out, there is something paradoxical about such strategies of idealization, as the justification of idealizations is due to their (future) eliminability.

In minimalist idealization—according to Weisberg—simple models are not to be de-

---

[4] Weisberg (2007) also introduces a third kind of idealization: the multiple models idealization.

[5] Consequently, Galilean idealizations have also been called tractability assumptions (e.g. Hindriks 2005). For McMullin, however, Galilean idealization does not only boil down to making the representation more tractable. He points out that for Galileo idealization also imposes order "in an attempt to grasp the real world from which the idealization takes the origin" (1985, 248).

idealized because this will not increase their explanatory or epistemic value. Batterman calls such minimalist notion of idealization "the non-traditional view" and claims that "(t)he adding of details with the goal of 'improving' the minimal model is self-defeating—such improvements are illusory" (2009, 430). Weisberg (Weisberg 2007, Elliott-Graves and Weisberg 2014) considers mainly the case where the minimal idealized model is thought to contain only the "core causal features that give rise to a phenomenon of interest" (Elliott-Graves and Weisberg 2014, 178). This idea is articulated in the work of Strevens (2008, 2016), who puts forth a causal difference-making approach, where idealization serves to distinguish between difference-makers and non-difference-makers. The latter are set aside through assigning to the variables representing them extreme or default values. However, Batterman (Batterman 2000; Batterman and Rice 2014) explicitly discards the idea that minimalist idealization depends on the assumption that models are explanatory by virtue of being able to isolate some core causal factors. Batterman underlines the positive explanatory role of idealizations: they may demonstrate what details are irrelevant instead of merely being used to isolate the difference-makers from the irrelevant features—and assigning the explanatory work to the difference-makers.

Despite these differences in accounts of minimalist idealization, Weisberg's proposal effectively captures a distinction that had already been built into the discussion of idealization—and the centrality of the question of de-idealization for different strategies of idealization. Further distinctions proposed also serve to highlight this centrality point. Sklar

has distinguished between controllable and uncontrollable idealizations in discussing when scientists are justified in isolating the system they are studying from the interferences of other factors, and assuming them to be negligible (Sklar 2000). In the case of controllable idealizations the theory, or a background theory, informs scientists "in what ways, and to what degree, the conclusions we reach about the idealized model can be expected to diverge from the features we will find to hold experimentally in the real system in the world" (Sklar 1993, 258). Sklar points out, however, that scientists do not always know how to compensate for such limit-style idealizations, raising the question of how to legitimize such less tractable and thus 'uncontrollable' idealizations (2000, 63). In somewhat similar vein, Elgin and Sober write about 'harmless' idealizations, where "a causal model contains an idealization when it correctly describes some of the causal factors at work, but falsely assumes that other factors that affect the outcome are absent" (2002, 448). Such idealizations "are *harmless* if correcting them wouldn't make much difference in the predicted value of the effect variable" (ibid.)[6]

What interests us in these distinctions—Galilean idealization versus minimalist idealization, and uncontrolled versus controlled or harmless idealization—is that they *crucially depend*, although in different ways, *on the possibility and desirability of de-idealization in modeling*. Galilean idealizations are considered reversible and correctable by de-idealization. Minimalist idealizations, on the other hand, dispense with de-idealization on the assumption that such idealizations could either be harmless or controllable, and/or de-

[6] Elgin and Sober do not talk about harmful idealizations.

idealizing them would be explanatorily counterproductive. However, the texts presenting these distinctions rarely mention de-idealization, let alone trying to articulate it. The situation is curious: how has such a central notion as de-idealization escaped the explicit attention of philosophers so far?

In what follows, we will consider de-idealization separately from the discussion on idealization, in contrast to the major philosophical contributions on idealization that have approached de-idealization as if it were a relatively uninteresting question, either conceptualizing it as a reversal, or questioning its desirability altogether. Even though the existent philosophical discussion on (de-)idealization has been orbiting around the question of whether to de-idealize or not, it has not examined what de-idealization entails, and accomplishes in actual scientific practices. We focus instead on de-idealization directly, drawing some inspiration from the literature on model application that points toward the complexities and wider functions of de-idealization, as well as to the creative and constructive processes involved (e.g. Alexandrova 2006; Morgan and Knuuttila 2012, Miyake 2015, along with this paper). De-idealization turns out to be central for the practice of modelling, and illuminative of what it encompasses.

3. The De-Idealization Menu: Forms, Aims and Heuristics

In order to address de-idealization on its own, freed from the traditional assumptions and

distinctions, we do not begin from the processes of idealization or even model-making, but rather study how models are made usable in certain domains. Models function for scientists as both representing devices and artefacts on which experiments of a particular form—'model experiments'—can be undertaken. On the one hand, then, scientists use models to represent some system of interest—real, hypothetical, or fictional—that they want to investigate. On the other hand, scientists work with models to learn more about their performance and implications. They treat models as experimentable, or explorative, devices: they ask questions of them, manipulate them and even 'play' with them to study their properties and so, directly or indirectly, the possible targets that they might be used to represent. Starting with these functions of models: representing and experimenting, we see that scientists are engaged in a variety of constructive activities when de-idealizing.

Our analysis begins with greater emphasis on the experimentable qualities of using models and moves toward greater emphasis on their representing qualities. As we find, the idea of de-idealization as reversal seems more apt—perhaps surprisingly—when modeling work is considered in an analogy to experimentation. Yet, the turn to representing issues shows that the idea of de-idealization as a set of reversals is very difficult to sustain. Of course, the two functions of models can never be fully disentangled. The conceptual distinction between the two functions furnishes, however, a handy analytic tool for the introduction of the processes of de-idealization under four categories: (i) re-composing, (ii) re-formulating, (iii) concretizing, and (iv) situating. *Re-composing* refers to the reconfiguration of the parts of the model with

respect to the causal structure of the world; the supposed links between the parts of the model and real elements, or causal forces, highlight the experimentable qualities of models. *Re-formulating* and *concretizing* deal more directly with the issues of representing in focusing on the two different sides of the abstractness of models: their symbolic and conceptual formulation. Finally, *situating* addresses the applicability of models to particular situations, either in the real world or in theorizing. It is concerned not just with how a model can be de-idealized to represent some determinable target situations, but how such processes enhance their use in theorizing, also stressing their mobility across different uses and disciplines. The proposed classification, and the associated labels, aim to render visible the positive, creative, and use-oriented aspects of de-idealization—and not just the challenges involved.

3.1. De-Idealizing as Re-Composing

One primary use of models is investigative, they are vehicles for gaining new knowledge. When scientists come to use models for investigative purposes, they treat them as experimentable objects, without of course waiving their representational status. This points us to the quality of models as experimental set-ups: simple and focused situations in which a very small number of causes or elements are considered and all other elements/causes are ‘shielded off’ outside the model’s boundaries. This is consistent with minimalist modelers’ viewpoint, but the point we make here is that models are not simplified representations because scientists necessarily believe in the simplicity of the world, but because these assumptions are needed

for models to function as experimentable devices. As for any laboratory experimental set-up, a modeller focuses on the relevant small number of factors/ elements, and shields the set-up from other factors including disturbances (Morgan 2005; Mäki 2005).

But to what extent does an analogy between modelling and experimental practice imply that such 'shielding off' idealizations could be undone in a reverse process of de-idealization? To reverse the (quasi)experimental set-up of a model is no simple task; it would not be in the laboratory either. One way to appreciate the difficulties of this set of de-idealization processes is to think of them as reversals of the various *ceteris paribus* conditions. They are conditions that generally remain implicit rather than explicit. Boumans (1999) has argued that there are three separate conditions here, not just one. Following his division, such processes of de-idealization entail adding back (a) those factors that are normally assumed absent yet that do have an influence (i.e. the *ceteris absentibus* factors); (b) those factors normally assumed of so little weight that they can be neglected in the idealized model (the *ceteris neglectis* condition); and (c) variability in those factors that are present but whose effect in the model is neutral as they are assumed to be held constant (i.e. the *ceteris paribus* factors).

There are practical difficulties in using the reversal processes for *ceteris absentibus* factors, as the set of likely causal factors to be taken back into account might be very large, not able to be fully specified, or dependent in complex ways on one another. If so, adding back these other causal factors will alter the existing contents of the model.[7] Such a model cannot be

[7] An early philosophical discussion of adding back in such omitted factors is found in

simply de-isolated, it can only be re-composed by knowledge of the rest of the elements. And just as the world is unlikely to be neatly decomposable, neither is the model.

Ceteris neglectis, in turn, concerns things so small that they can be neglected—providing, as we have seen, one important defense of minimal modelling strategy. But even if small individually, when added together, the neglected factors might make a difference to the model in application.

Within the context of economics, reversing the *ceteris paribus* conditions (those that hold things constant) has been more discussed than relaxing the two earlier conditions. Models have often embedded assumptions that have been made to smooth out variety to create stability and so enforce homogeneity, and it is not always obvious how that squashed-out variety is to be re-constituted. This might—for example—mean replacing distributions for averages, messy empirical values for simplified hypothetical values, or bringing in time-dependent variations rather than assuming a static world. Such reversing may include correcting factors which have been set to *ideal values* for which there is no evidence of any *real values*, either because of absence of knowledge or because there are no possible equivalent de-idealized values. But notice that de-idealization may involve something easy to do that complicates the model only a little such as replacing average values by probability distributions, or something very different such as replacing perfect knowledge with partial ignorance.

Hausman 1990.

The ultimate problem of laboratory experimental work is that the world in the test tube is so restricted and isolated that it cannot immediately and easily be fitted up to be usable in the world—think only of the incredible scientific investment in pharmacology to test whether a synthesized 'cure' developed in the lab will work in patients. But the problems of model experiments extend beyond that, and are also of a different nature—an experiment in the lab is not the same as an experiment on a model (see Morgan 2003, 2012). Models are usually accomplished by representing the system to be investigated in a different medium than their real-world target systems (for instance, a real life economic action is represented in mathematical form). These differences in material media take us to the realm of representing. The issues involved concern both the constraints of the representational languages used that become visible in the attempts to re-formulate the model (3.ii) and those of concretizing the theoretical concepts (3.iii).

3.2. De-Idealizing as Re-Formulating

The diversity of scientific models is astounding, they are formulated in many different modes of representation in order to convey their content. These representational means impose their own constraints on modeling that can be both enabling and limiting (Knuuttila 2011). If the model is diagrammatic, for example, it offers certain possibilities and imposes certain limitations on what can be represented, and these will be different if the model form is algebraic, or geometric. There are three major considerations here: *integration issues*,

tractability issues, and *translation issues*, all of which provide challenges to any process of de-idealization.

Models must hold together, there must be some form of *integration*, a process of giving overall form to a model. Such integration may be achieved quite subtly, by what Boumans (1999) aptly calls 'mathematical molding' that amounts to, for example, making mathematical formulation choices that integrate a set of elements in a certain way. Mathematical molding is a central feature of the model, yet it might not be noticed or seen as such—once the choices have been made. These choices related to mathematization cannot often simply be 'undone': de-idealization involves re-formulating the model, taking into account that the model might fall apart without that particular construction.

Integration operates as a strong constraint, but it is not always obvious which side it is on: the side of idealization or de-idealization. For example, should the sequence of equations in a model embed a simultaneity requirement or be block recursive (modular)—a choice with very different consequences for processes of de-idealization, for the former cannot easily be taken apart, where the latter can be. This particular problematic sits at the heart of economics. In theoretical terms, it marks the difference between a statement that the world is in a state of equilibrium in all markets at all points of time, from the alternative view that the system only has a tendency to equilibrium. It is a contrast that permeates both theoretical modeling and applied modeling. Economists may believe that the modular system best represents how people plan and act, in a very complicated set of co-dependency relations that are also time

dependent—implying only a tendency toward equilibrium. In principle, these relations could be unraveled within a model, but in practice economists can only get aggregate data for such models at such wide intervals that the application model must necessarily be written in a simultaneous form and so as an equilibrium model (Morgan 1991). Even with the present computing power, treating every individual in the market as a member of a simultaneous system is not easy. As in climate science modelling, this is more than a big data problem, it is a complexity problem, and attempts to solve it in economics may start from another direction (such as by agent-based modeling).

These representing issues become intertwined with experimentability issues when we recognise that models are built to be *tractable.* Frequently, this is thought to mean merely setting certain variables to certain values (e.g. to zero) in order to make the mathematics work easily. But it is often difficult to know how many of those model assumptions could be translated back into statements about real entities and processes. Alexandrova (2006) calls such assumptions 'derivation facilitators' and asks whether it is more realistic for agents to have discretely as opposed to continuously distributed valuations, given that it is already questionable whether people form their beliefs of a value by drawing a value from a probability distribution.

At a granular level, then, it is difficult to see how one could easily tease apart the individual assumptions of a model and de-idealize them separately. Indeed, according to Cartwright (1999), economic models are over-constrained by which she means that the modeled situation

is constructed in order to yield certain kinds of results. Morrison (2009) pays attention to this same feature of mathematical abstractions in physics claiming that they are needed to make the model work.

Tractability of course impinges on the investigative function of models. For example, economists' infamous 'overlapping generations' model designed to get at the relationships between consumption and savings in an economy imagined a world of two generations, who work and save in their first period and who use their savings to consume in retirement (see Hausman's 1992 analysis). The model relates the two 'generations' so that each new working generation transfers resources to the current retired generation. Restricting the model made it tractable, indeed, economists often begin with modeled worlds which have only two dimensions (two consumers, two goods, or two factors of production), for ease of the mathematics. De-idealizing to increase the number of dimensions (for example, an overlapping model of three generations: children, workers, retirees) would make their models somewhat more realistic of course, but also more difficult to manage.

The process of de-idealizing mathematical models may also involve *translations*, for different scientific uses may require a formulation that is more convenient for that particular use, that is, de-idealizing may involve making a choice of different representational modes, frequently a switch from one formal language to another. Whatever formal language the model is presented in, it cannot straightforwardly be translated into another formal language, for both will likely have a different semantics and syntax. Even in those cases where the various

mathematical versions of a model are ‘formally equivalent’, implying easy switching between ‘equivalent’ formal representations, scientists’ own subject-based understanding of, and use of, the model is likely to be different (Vorms 2011; Morgan 2012). As an example, game theory in its early years had three different representations, in three different mathematical formulations, to describe or instantiate game structures: a matrix structure of payoffs (which depicts the outcome of choices), a branching tree diagram of possibilities (which depict the decision process of choosing), and a spatial solution set (depicting the set of possible solutions) (Luce and Raiffa, 1957). These different formulations focused on different aspects of the relevant game for different purposes, and imply different processes of de-idealization for use.

These difficulties of arriving at mathematical representations and holding model elements together may explain in part the enormous success of some mathematical formulations that are applied across different disciplines. Examples of such cross-disciplinary templates (Humphreys 2004) are general mathematical forms and computational methods underlying such simple mathematical models as the Ising model and the Lotka-Volterra model, or network methods more generally, all of which have been applied to various problems within economics. But, of course, de-idealizing models built on cross-disciplinary formal templates is problematic almost by definition since their application is precisely based on the tractability of their particular syntactic configurations. Moreover, the semantics are also important: the template needs to be translated from the theoretical framework of the source field to the new

discipline, as well as the new target, in question. Such translation typically involves considerable theoretical effort. For example, it is not a trivial question how a formal template designed for the phenomenon of ferromagnetism can be applied to neural networks, or peer pressure in socio-economic systems (Knuuttila and Loettgers 2014, 2016). Accordingly, many problems of translation point to the issues faced by the attempt to concretize the concepts incorporated into models.

3.3. De-idealizing as Concretizing

On the representing end of the spectrum, de-idealizing involves (apart from re-formulating) concretizing the conceptual core of the model that may be needed for specific purposes in theorizing or in application. The idealized model is likely to embed a scientist's theoretical or conceptual commitments about either the system or the elements of that system. While there has been some consideration of concept-formation associated with modeling (e.g. Wartofsky 1968; Nersessian, 2008), there has been little on the problem of de-idealizing those conceptual abstractions (except for Nowak 1992, 2000). This means figuring out how such conceptual abstractions about the system, or the elements in it, are made concrete; the conceptual elements can be de-idealized in different ways for different sites and for different purposes.

The concept of an economy, the economy as a whole unit, has been made concrete in many different ways, for example: as following a dynamic path with cyclical oscillation in business

cycle research; as a system that relates all the inputs to all the outputs of each productive sector in 'input-out analysis'; or as a 'macro-economy' which uses an accounting framework to examine the relations of aggregates of consumption, investment, etc. These different concretized models enable both reasoning and analysis at a theoretical level, and substantive empirical investigation.

'Utility' in turn, is concept more usually associated with the individual. It is one of the most abstract and ubiquitous concepts used in economics, referring to the unobservable relationship between people and the goods that they consume, and is the conceptual starting point for much of modern economics of individual behavior. There are various versions of the concept: focusing on human need, or satisfaction, or enjoyment, and other relational notions. It was developed in both textual and mathematical accounts in nineteenth-century economics and in the twentieth century there have been various attempts to de-idealize it for specific and interventionist usages. One of many examples has been in the development of QUALYS: quality-adjusted life years, in order to capture subjective (i.e. patient-experienced) utility from extended life following a medical intervention. These processes are not just ones of measurement operationalization, nor of replacing the *symbolic* abstraction 'U' (for 'utility'), but are designed to fill in the content that would fit that conceptual abstraction 'utility' to equivalent *substance*: such as a quality of physical life (see section 3.v. below).

Being more concrete does not necessarily mean being more realistic or accurate to any particular observable objects in the world, for the concretized versions of these elements

remain wedded to their conceptual framing. The implications of concretization and the choices they require are well demonstrated in Frank Knight's idealized interpretation of rational economic man. The main ancestor of 'rational economic man' is usually taken to be 'homo economicus', a character associated with John Stuart Mill's mid-nineteenth century recipe for making economics a doable science: a wealth-seeking miser who was nevertheless held back by his desires for luxury, and dislike of work. For Mill, this was an abstraction in two senses: these were the economic characteristics of man to be found universally, but they were also understood conceptually. In the late nineteenth century, economists' economic man was portrayed as a consumer, seeking to maximize his utility according to his preferences. In order to fully explore the rationality of that notion of economic behavior, Knight endowed him with the virtues of perfect knowledge and perfect foresight, so that his economic model man had no ignorance of the present and no uncertainty about the future. The de-idealization might seem obvious—return uncertainty and lack of foresight to the account of man's behavior. Changing this assumption would not have been that straightforward for Knight (1921), who thought that lack of knowledge could be divided into two chunks: risk, for which we could write down a probability distribution, and genuine uncertainty about which we had no knowledge. Moreover, Knight later described the implications of his assumptions to mean that his idealized economic man was actually just a slot machine, with no reasoning power and no intelligence (Morgan 2006). So, de-idealizing Knight's model would mean introducing into it

intelligence and reasoning power—a considerable problem in artificial intelligence, rather than a relatively contained, if difficult, task of forecasting the economic future.

3.4. De-Idealizing as Situating

Models, by virtue of their simplified, ideal or abstract qualities are not immediately applicable to any real concrete situation in the world. This final important category refers to how models often need to be explicitly situated back into the world, and not just into the world in any general sense, but rather to be made usable for *specific* situations in the world. One obvious place we can see this happening is when a simple mathematical model used in theorizing is de-idealized into a statistical model as it becomes fitted to data. This is not just a matter of change in language (i.e. 're-formulating' as 3.ii), but of positive fitting to specific case situations. For example, in economics in the early days of modeling markets for goods, it was often assumed that statistical data could just be fitted to the equations, for the data issues and measurement requirements were considered hardly relevant to questions about the difference between corn and hog markets. More recently, such economic modeling has taken the lessons from the problems of fitting models to data and begun to retrofit mathematical models so that they are already geared towards the statistical data available (for example, by embedding the probability assumptions into the micro-choice structures faced by specific sets of individuals in economic labour markets).

The aim of de-idealization as situating might be to locate a model in many different but perhaps superficially similar specific sites, either using statistical work, or experimental work in lab or field. There is no reason to expect any 'general' de-idealization, i.e. one that will work everywhere: any model is likely to need a different de-idealization for every different situation: time, place and topic. For example, poverty alleviation field experiments are usually based on some idealized model of behavior, which may be successfully situated (de-idealized) for application in a particular site, but then often prove not so successful when applied at another geographical site. The critical point to note here is that models made relevant by de-idealizing (situating) for just one site may require *idealizing again* in some respects (i.e. de-situating) and then de-idealizing again (re-situating) in order to be relevant to another site (see Cartwright 2012 and Morgan 2014). Such processes, ie. the transfer of a model-based experimental designs from one site to another different site, bear remarkable similarities with the transfer of model templates within and between different disciplines that we discussed in 3ii.

What these similarities between the application of model-based experimental designs and the transfer of theoretical model templates show is that de-idealization as (re-)situating is not just a challenge for applied work, but equally relevant for theoretical work where models need to be partially de-idealized to situate them in a particular domain of theorizing. These processes offer an equally open-ended and challenging agenda. A telling example is provided by the supply and demand model, probably the most iconic model in economics. This model

in its original diagrammatic representation was a simple cross or ‘scissors’ diagram of the supply and demand curves cutting to capture the price and quantities exchanged in the market (there was also a version in an algebraic format). Even before the end of the nineteenth century, the diagram could be found in different versions as appropriate for thin markets like the market for race horses (where the demand and supply are both limited and value/price are difficult to determine), and for markets for spoiling goods, like fish. That diagrammatic model was also developed at the same time to picture different shapes for the supply curves as appropriate for different industrial structures on the supply-side (monopolists or competitors). Thus, that very simple iconic general model was re-formulated to be appropriate for categories or kinds of things in the economic world. These forms of de-idealization create various generic models: applicable to particular classes of things, remaining still theorizing objects as direct offspring of the original simpler model. Each one of them formulated for kinds of markets, yet no one version of the model applying to all and every market.

Although de-idealizing supply and demand model may seem to involve steps towards more concrete accounts of markets, this does not amount to simply adding back factors, or reversing assumptions, or even reaching something like a realistic account of markets. Rather, de-idealization involves shaping the ideal model in particular ways so that it becomes relevant to a subset of the domain for both theorizing about, and applications in, those sub-domains. Thus, the de-idealizing process may involve a move from abstract and general to a still

abstract, formal model appropriate for a generic class of phenomena—or to a level of model which is evidentiary specific.

Sometimes this situating process involves radical change, particularly changes in concepts. The supply and demand model had to be reinterpreted when it was moved from the market for goods to the market for labour, prompting the concept of 'voluntary unemployment'—i.e. that the unemployed chose unemployment because they valued leisure (regardless of whether the choice was real in the sense that job vacancies were available). The appropriation of the model in a very different domain required a re-conceptualization of the nature of unemployment. Once again, such processes of de-idealization may well accompany the transfer of models for use between subfields within a discipline, or even between fields. Game theory models have been moved from economics to apply in political science and evolutionary biology, but not without changing conceptual interpretations and usages in these theoretical domains, hinting towards the often neglected conceptual dimension in de-idealization—and underlining the close entanglement between concretizing and situating.

3.5. The De-idealizing Menu: Example

Above we categorized de-idealization into four distinct processes related to the investigative and representational functions of models. Our analysis recognizes strong limitations in understanding de-idealization as processes of reversal, and suggests an alternative way of thinking about it. Our four categories—of re-composing, re-formulating, concretizing, and

situating—provide not only a useful analytical framework, but now offer an array, or menu, of processes of de-idealization that can be applied to a model according to the purposes at hand. These processes are exemplified in the Figure 1 with respect to the utility function, and discussed below through the example of QUALYS, quality adjusted life years for a given medical procedure.

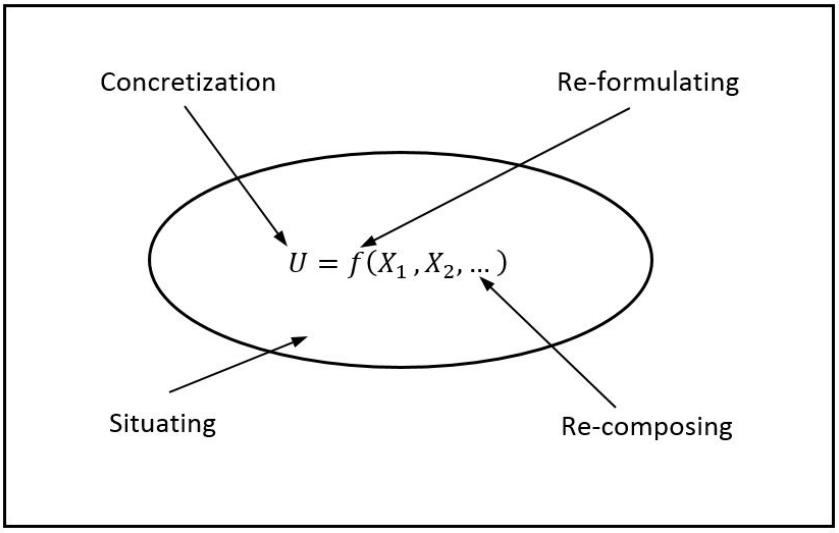

Figure 1: De-idealizing the utility function

Concretization: The symbolic abstraction U stands for the conceptual abstraction 'utility': the relation between a person and a 'good' understood as the value they gain from consuming the good (where the notion of a 'good' includes anything that a person values, such as a musical performance or a hot shower or even a replacement limb, not just a box of chocolates or a cold

drink). The notion of utility has been the subject of deep analysis and debate over a very long period. Its conceptual meaning has varied over history and according to specific institutional and problem situations in which it is used. At the current time, it primarily features as both a mathematical construct in equations, and as a theoretical entity with psychological implications underlying choice behavior. That theoretical entity refers to something that economists believe "exist(s) independently of scientists and the scientific conventions of the scientific community" (Cohen 1997, 178).

Given the ontological status of the notion of utility, economists have—with a large investment in social research—made it concrete for specific purposes in practical domains such as developing the notion and measurement of QUALYS: Quality Adjusted Life Years. And in this context, health-related utility functions can include both needs and preferences.[8] One can measure a patient's preferences with respect to a particular treatment – ie. how much does it contribute to improved quality of life, convenience of use, fewer side effects – and also to their needs – ie. survival.

Re-formulating: Models are made for reasoning with, and so need to be formulated to make them workable instruments to reason with, in this case involving decisions eg. on the shapes of

[8] Economists have followed two alternative strands of measurement according to two notions of utility, one which frames thinking about utility valuations as stemming from individuals' preferences and the other in terms of their needs (Cohen 1997).

people's utility curves, their choice behaviour, and what it means to maximize. In the simple form shown in Figure 1, it is not clear what functional form the model takes, nor what holds the elements together: the model is not ready for use until these commitments are made. In the case of QUALYS, one of the complications for these decisions is that in order to normalize experience across a range of patients and medical interventions, QUALYS are valued from 0 to 1, with 0=death, 1=full healthy life; and (arbitrarily set) interval scales are used between these. The interval scale, still idealized, is useful for policy making and administrators in allowing for comparative measurements, say, between different types of medical interventions. But, the presumption that increments along the interval scale from 0 to 1 are all equivalent is problematic, and such scales may not make much sense at the beginning and end of the scale. All of these questions have implications for the way the model is formulated for theoretical reasoning, and for practical usages.

Re-composing: The valuations, and weighting of needs and preferences, that people make about their utility in judging the future quality of their life under certain circumstances of less than perfect health (in QUALYS) will vary with age, gender, nationality, family circumstances, etc.. So, if the economist's purpose is to understand the factors that influence those valuations, such factors should be introduced separately into the model. Their introduction is challenging because the added factors are not likely to be independent of each other; and there may be unknown disturbing factors that cannot be taken into account. This marks the problem of reversing the *ceteris absentibus* and *neglectis* conditions. However, the

economist can render such re-compositional issues into statistical measurement problems, interrogating the extent to which the variation in unknown variables account for the measured differences in utility valuations. In this way, economists will gain further information about what has been left out via what are called 'the residuals' in such a measurement equation. Such residuals can often be very illuminating about missing factors, or the forms given for those included factors, and this kind of learning from recomposing has long been a standard practice in econometrics (the statistical branch of economics). Similarly, in discussing earth sciences, Teru Miyake (2015) has recently related the use of residuals to what he calls 'active' de-idealization: the generation of new observations through the comparison of the actual observations with predictions of simple and "false" reference models.

Situating: Before QUALYS can be used in decisions by individuals, health providers, and payers to decide on actions that will affect individuals with a certain condition (eg. kidney dialysis or hip replacements), the model and measurements of utility have to be situated for that group of people and their particular decisions. This situating could concern patients thinking about the impact of a particular, or of alternative treatments, but it could also result in a decision model for health service providers who aim to compare costs of treating alternative conditions in view of their gain for patients as a group. The model will have to be 'peopled' with the specific QUALYS values based on survey data from a set of patients, and the specific costs of different treatments. For treatment decisions at the system level, the model will also need to be shaped and adjusted for specific country and institutional health service system (eg.

a national health insurance-based system vs a private health care system).

Our four categories—re-composing, re-formulating, concretizing, and situating, fleshed out above—provide a framework for analyzing de-idealization. Other kinds of generic processes, or finer-grained distinctions are surely possible. But these categories do offer a menu of processes of de-idealization that can be applied to a model according to the purposes at hand. It is important to note, moreover, that this array of de-idealization strategies is in many cases independent from any original idealizing strategies or purposes of the idealized model. Model users de-idealize in view of their aims in particular theoretical and practical contexts—and such processes likely involve different combinations of de-idealizing strategies.

The de-idealizing menu, along with our economic examples, show how problematic the notion of de-idealization as a set of reversals can be. The problems do not only boil down to our limited knowledge concerning the omitted factors, or practical problems concerning tractability, admittedly often demanding—they are more endemic in nature. Examining the range of details posed by the four kinds of de-idealizing processes showed that there is no easy 'adding back', or 'correcting' for previous idealizations. And just as there are no easy reversals, neither are there self-evident movements between the more or less idealized state of a model—the idea (spelled out by Nowak 1992, 2000) that scientists can go up and down the ladder from idealized to less idealized and back again. And, consequently, there is no precise

way to talk about the degree of (de)idealization either (cf. Levy 2018). De-idealization as a reversal is clearly an idealization of its own!

4. Modeling Reconsidered

The starting point of our paper was to examine how the philosophical discussion on idealization crucially makes use of the idea of de-idealization, whilst at the same time leaving that notion largely unarticulated and unexamined. We suggested that this neglect can be explained by the implicit assumption that de-idealization amounts to a reversal of the idealizing process. In our analysis of de-idealization we did not begin from the notion of reversal, and so not from that of idealization either. Rather, we set out to study the actual processes of de-idealization. But the focus on the theoretical and practical challenges of de-idealization also proves illuminative of modelling, as well.

Our analysis of de-idealization processes opens up two critical perspectives to modeling which largely remain hidden when only the processes of idealization are considered. The first concerns decomposability of models and the second modeling heuristics—the way models are actually achieved in scientific practice.

Thinking seriously about the practices and processes of de-idealization leads us to ask to what extent models can be considered as entities whose parts can be teased apart from each other and edited, or corrected, as the reversal account would have it. The notion of de-

idealization as a reversal of idealization seems to require that models are composed of separable assumptions or components, enabling theorists to de-idealize such components in selective, controlled fashion. And, furthermore, our knowledge of the bits and pieces of the real world could then be, at least in principle, mapped onto a model in a relatively unproblematic way (and vice versa). In other words, the idea of reversing step-by-step the idealizations made in the modeling process presupposes that models are decomposable. This seems to be a generally held view of philosophical writings on modeling and idealization.[9] Yet the problems encountered by robustness analysis show that this may not often be the case (Odenbaugh and Alexandrova 2011). Of course, minimalist idealization does not need to rely for its explanatory value on de-idealization. However, the causal difference-making variant of minimalist idealization seems to be based on the decomposability assumption concerning models, supposing, moreover, that the world is causally modular, enabling the analyst to separate the difference-making causal factors from non-difference making irrelevant ones.

But our analysis of the processes of re-composing and re-formulating suggest that models are relatively inflexible to changes in their contents in many different respects. Just as an experimental protocol needs to keep the experiment shielded for it to work, so too in reasoning with models. It may not be possible to add back in certain causal factors without consequences as these factors are related to others the scientist also wants to keep in the model. And the

[9] See, however, Rice (2017) for a critique of the presumption that models could be decomposable into contributions made by their different parts.

representing issues inherent in the de-idealization processes show that a model may not be decomposable in another more serious sense—if the model consists of the integration or the molding of various elements together then it may not be possible simply to tease those elements apart without collapsing the functionality of the model. In short, models may not be robust to many kinds of de-idealizations.

Our second point concerns modeling heuristics. We espy, in the usual assumptions concerning idealization combined with the associated neglect of de-idealization problems, an insidious presumption by philosophers of science that scientists originally start from considering the real world, and then arrive at their models through idealizing (and abstracting). And because that habitual assumption is made, it also seems unproblematic to assume that scientists can reverse their modeling recipes to get back down to the more fully-blown situation they started with, fraught though that process might be. But many of the challenges we discussed stem from the fact that scientists did not start with a complex picture, simplifying it to get to idealized tractable models. As we know from other studies of modeling, scientists often begin with something that is already simple and abstract in content, and based around some conceptual elements that they believe underlie the phenomena they want to model (Morgan 2012; Knuuttila and Loettgers 2016). That is, they often just begin with an idealized simple model, not with processes of idealization to get to that model. This critical hidden point lies behind many of the challenges of de-idealization outlined in the paper.

If, and when, scientists start with already simple abstract models, one of the main challenges of de-idealizing arises from filling in the set of unknown elements, concretizing and situating the concepts, and being able to render them into a representable form. These are *ceteris paribus* conditions of various kinds, assumptions related to mathematical tractability, assumptions about what is most relevant, and challenges of definitional and conceptual content, etc. Many of these assumptions might not be spelt out because they are taken for granted by those in the community working with that group of models, typically only some of them are articulated. Moreover, as far as scientific practice goes, models are not well-defined by a set of assumptions that lie behind them, nor are they only derived from them in some determinate manner. They can rather be conceived as free standing artefacts (Knuuttila 2011, 2017), with a degree of autonomy from both theory and data regimes (Morrison and Morgan 1999). And they may be constructed in various ways—through analogy and template transfer, through putting together a list of ingredients in view of some theoretical goals, or through the use of theoretical imagination (Morgan 2012). None of these standard ways of model construction starts only with a set of assumptions in order to derive a model.

It is important to realize, then, that although models appear simple, they may not in fact be so *because* they were suitably simplified in the modeling process, but rather it is because they were chosen from the start to be ideal and abstract in certain ways. That models often 'start seemingly simple' has important consequences for de-idealization. The challenges of de-idealization we studied are generic, separable in principle, yet their difficulties may be

confounded when a model is not made by any explicit process of idealization, but rather a scientist has started with a simple or ideal hypothesis in model form, possibly making use of some familiar model template. This also means that even simple models are much more problematic objects than philosophers have noticed. They are more inflexible than the reversal thesis would have us believe, and so de-idealization emerges as creative a part of modeling as any other dimension of it.

REFERENCES

Alexandrova, Anna. 2006. "Connecting Economic Models to the Real World: Game Theory and the FCC Spectrum Auctions." *Philosophy of the Social Sciences* 36:173-192.

Batterman, Robert W. 2000. "Multiple Realizability and Universality." *The British Journal for the Philosophy of Science* 51:115-145.

———. 2009. "Idealization and Modeling." *Synthese* 169:427-446.

Batterman, Robert W., and Collin C. Rice. 2014. "Minimal Model Explanations." *Philosophy of Science* 81:349-376.

Boumans, Marcel. 1999. "Philosophy of Complex Systems." In *Models as Mediators*, ed. Mary Morgan, and Margaret Morrison, 66-69. Cambridge: Cambridge University Press.

———. 2005. *How Economists Model the World into Numbers*. London and New York: Routledge.

Cartwright, Nancy. 1989. "Capacities and Abstractions." *Minnesota Studies in the Philosophy of Science* 13:349-356.

———. 1999. "The Vanity of Rigour in Economics: Theoretical Models and Galilean Experiments." Discussion paper series 43/99. Centre for Philosophy of Natural and Social Science.

———. 2012. "Will this Policy Work for You? Predicting Effectiveness Better: How Philosophy Helps." *Philosophy of Science* 79:973-989.

Cohen, Joshua. 1997. *Utility: A Real Thing - A Study of Utility's Ontological Status*. Amsterdam: Thesis Publishers.

Elgin, Mehmet, and Elliott Sober. 2002. "Cartwright on Explanation and Idealization." In *Ceteris Paribus Laws*, ed. John Earman, Clark Glymour and Sandra Mitchell, 165-174. Dordrecht: Springer.

Elliott−Graves, Alkistis, and Michael Weisberg. 2014. "Idealization." *Philosophy Compass* 9:176-185.

Godfrey-Smith, Peter. 2009. "Abstractions, Idealizations, and Evolutionary Biology." In *Mapping the Future of Biology*, ed. Anouk Barberousse, Michel Morange and Thomas Pradeu, 47-56. Dordrecht: Springer.

Hausman, Daniel M. 1990. "Supply and Demand Explanations and their *Ceteris Paribus* Clauses." *Review of Political Economy* 2:168-187.

———. 1992. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.

Hindriks, Frank A. 2005. "Unobservability, Tractability and the Battle of Assumptions." *Journal of Economic Methodology* 12:383-406.

Humphreys, Paul. 2004. *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. New York: Oxford University Press.

Knight, Frank H. 1921. *Risk, Uncertainty and Profit*. Boston: Houghton Mifflin Company.

Knuuttila, Tarja. 2011. "Modelling and Representing: An Artefactual Approach to Model-Based Representation." *Studies in History and Philosophy of Science Part A* 42:262-271.

———. 2017. "Imagination Extended and Embedded: Artifactual and Fictional Accounts of Models". *Synthese.* doi: 10.1007/s11229-017-1545-2

Knuuttila, Tarja, and Andrea Loettgers. 2014. "Magnets, Spins, and Neurons: The Dissemination of Model Templates Across Disciplines." *The Monist* 97:280.

———. 2016. "Model Templates within and between Disciplines: From Magnets to Gases—and Socio-Economic Systems." *European Journal for Philosophy of Science* 6:377-400.

Levy, Arnon. 2018. “Idealization and Abstraction: Refining the Distinction.” *Synthese*. https://doi.org/10.1007/s11229-018-1721-z

Levy, Arnon, and William Bechtel. 2013. "Abstraction and the Organization of Mechanisms." *Philosophy of Science* 80:241-261.

Luce, Robert Duncan, and Howard Raiffa. 1957. *Games and Decisions: Introduction and Critical Surveys*. New York: Wiley.

Mäki, Uskali. 2005. "Models are Experiments, Experiments are Models." *Journal of Economic Methodology* 12:303-315.

McMullin, Ernan. 1985. "Galilean Idealization." *Studies in History and Philosophy of Science Part A* 16:247-273.

Miyake, Teru. 2015. "Reference Models: Using Models to Turn Data into Evidence." *Philosophy of Science* 82:822-832.

Morgan, Mary. 1991. "The Stamping Out of Process Analysis in Econometrics." In *Appraising Economic Theories: Studies in the Methodology of Research Programs,* ed. Neil B. de Marchi, and Mark Blaug, 237-263; 270-272. Aldershot/Brookfield: Edward Elgar Publishing.

———. 2003. "Experiments Without Material Intervention: Model Experiments, Virtual Experiments and Virtually Experiments." In *The Philosophy of Scientific Experimentation*, ed. Hans Radder, 261-35. Pittsburgh: University of Pittsburgh Press.

———. 2005. "Experiments Versus Models: New Phenomena, Inference and Surprise." *Journal of Economic Methodology* 12:317-329.

———. 2006. “Economic Man as Model Man: Ideal Types, Idealization and Caricatures” *Journal of the History of Economic Thought* 28(1): 1-27

———. 2012. *The World in the Model: How Economists Work and Think*. Cambridge: Cambridge University Press.

———. 2014. "Re-Situating Knowledge: Generic Strategies and Case Studies." *Philosophy of Science* 80:1012-1024.

Morgan, Mary, and Tarja Knuuttila. 2012. "Models and Modelling in Economics." In *Philosophy of Economics*, ed. Uskali Mäki, 49-88. Oxford: Elsevier.

Morrison, Margaret. 2009. "Models, Measurement and Computer Simulation: The Changing Face of Experimentation." *Philosophical Studies* 143:33-57.

Morrison, Margaret, and Mary Morgan. 1999. "Models as Mediating Instruments." In *Models as Mediators: Perspectives on Natural and Social Science*, ed. Mary Morgan, and Margaret Morrison, 10-37. Cambridge: Cambridge University Press.

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

Nowak, Leszek. 1992. "The Idealizational Approach to Science: A Survey." In *Idealization III: Approximation and Truth*, ed. Jerzy Brzeziński, and Leszek Nowak, 9-66. Amsterdam: Rodopi.

———. 2000. "The Idealizational Approach to Science: A New Survey." In *Idealization X: The Richness of Idealization*, ed. Leszek Nowak, and Izabella Nowakova, 109-184. Amsterdam: Rodopi.

Odenbaugh, Jay, and Anna Alexandrova. 2011. "Buyer Beware: Robustness Analyses in Economics and Biology." *Biology & Philosophy* 26:757-771.

Reiss, Julian. 2008. *Error in Economics: The Methodology of Evidence-Based Economics*. London: Routledge.

Rice, Collin. 2017. “Models Don’t Decompose That Way: A Holistic View of Idealized Models.” *The British Journal for Philosophy of Science*. https://doi.org/10.1093/bjps/axx045

Sklar, Lawrence. 2000. *Theory and Truth: Philosophical Critique within Foundational Science*. Oxford: Oxford University Press.

Sklar, Lawrence. 1993. *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press.

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

———. 2016. "How Idealizations Provide Understanding." In *Explaining Understanding: New Perspectives from Epistemology and the Philosophy of Science*, ed. Stephen Robert Grimm, Christoph Baumberger and Sabine Ammon, 37-49. New York: Routledge.

Svetlova, Ekaterina. 2013. "De-Idealization by Commentary: The Case of Financial Valuation Models." *Synthese* 190:321-337.

Vorms, Marion. 2011. "Representing with Imaginary Models: Formats Matter." *Studies in History and Philosophy of Science Part A* 42:287-295.

Wartofsky, Marx W. 1968. *Conceptual Foundations of Scientific Thought*. New York: Macmillan.

Weisberg, Michael. 2007. "Three Kinds of Idealization." *The Journal of Philosophy* 104:639-659.