THREE KINDS OF IDEALIZATION*

Philosophers of science increasingly recognize the importance of *idealization*: the intentional introduction of distortion into scientific theories. Yet this recognition has not yielded consensus about the nature of idealization. The literature of the past thirty years contains disparate characterizations and justifications, but little evidence of convergence towards a common position.

Despite this lack of convergence, consensus has clustered around three types of positions, or three kinds of idealization. While their proponents typically see these positions as competitors, I will argue that they actually represent three important strands in scientific practice. Philosophers disagree about the nature of idealization because there are three major reasons scientists intentionally distort their models and theories; all three kinds of idealization play important roles in scientific research traditions.

The existence of three kinds of idealization means that some classic, epistemic questions about idealization will not have unitary answers. We cannot expect a single answer to questions such as: What exactly constitutes idealization? Is idealization compatible with realism? Are idealization and abstraction distinct? Should theorists work to eliminate idealizations as science progresses? Are there rules governing the rational use of idealization, or should a theorist's intuition alone guide the process? However, the three kinds of idealization share enough in common to allow us to approach the answers to these questions in a unified way. The key is to focus not just on the practice and products of idealization, but on the goals governing and guiding it. I call these goals the *representational ideals* of theorizing. Although they vary between the three kinds of idealization, attending to them will help us better understand the epistemic role of this practice.

* Many thanks to Peter Godfrey-Smith, Stephan Hartmann, Paul Humphreys, Steve Kimbrough, Ryan Muldoon, Michael Strevens, Ken Waters, and Deena Skolnick Weisberg for extremely helpful comments and advice. I am also grateful for the thoughtful questions and comments from audiences at the Minnesota Center for Philosophy of Science, Tilburg University, the University of Pennsylvania, and Washington and Lee University, where earlier versions of this paper were presented. Thanks also to students in my graduate seminars at Penn for ongoing, stimulating discussion about idealization. The research in this paper was partially supported by National Science Foundation grant SES-0620887.

0022-362X/07/0412/639–59

I. THREE KINDS OF IDEALIZATION

Since the early 1980s, philosophers of science have paid increasing attention to the importance of idealization in scientific inquiry. While earlier literature acknowledged its existence, the pioneering studies of Nancy Cartwright,[1] Ernan McMullin,[2] Leszek Nowak,[3] William Wimsatt,[4] and others paved the way for the contemporary philosophical literature on the topic. Through much of my discussion, I will follow Cartwright's characterization and talk about theoretical representation in terms of *modeling*, the indirect representation of real world phenomena with models.[5] But many of the ideas in this paper are not essentially tied to modeling, so my reliance on the model-based idiom should not be seen as affirming this connection.

One of the most important insights of the modern idealization literature is that idealization should be seen as an activity that involves distorting theories or models, not simply a property of the theory-world relationship. This suggests that in order to distinguish between the three types of idealization we will need to know what activity is characteristic of that form of idealization and how that activity is justified. These activities and justifications can be grouped into three kinds of idealization: *Galilean idealization*, *minimalist idealization*, and *multiple-models idealization*.

Galilean Idealization. Galilean idealization is the practice of introducing distortions into theories with the goal of simplifying theories in order to make them computationally tractable. One starts with some idea of what a nonidealized theory would look like. Then one mentally and mathematically creates a simplified model of the target.

Galilean idealization has been thoroughly characterized and defended by McMullin who sees the point of this kind of idealization as "grasp[ing] the real world from which the idealization takes its origin," by making the problem simpler, and hence more tractable (*op. cit.*, p. 248).[6] Galileo employed the technique both in theoretical

[1] Cartwright, *How the Laws of Physics Lie* (New York: Oxford, 1983), and *Nature's Capacities and Their Measurements* (New York: Oxford, 1983).

[2] McMullin, "Galilean Idealization," *Studies in History and Philosophy of Science*, XVI (1985): 247–73.

[3] Nowak, "Laws of Science, Theories, Measurement," *Philosophy of Science*, XXXIX (1972): 533–48.

[4] Many of Wimsatt's most important papers on idealization and related topics are collected in Wimsatt, *Re-engineering Philosophy for Limited Beings: Piecewise Approximations of Reality* (Cambridge: Harvard, 2007).

[5] For more detail about the practice of modeling, see Michael Weisberg, "Who Is a Modeler?," *British Journal for the Philosophy of Science*, LVIII (2007): 207–33.

[6] A similar account is developed by Nowak; see Nowak, "The Idealizational Approach to Science: A Survey," in Jerzy Brzezinski and Nowak, eds., *Idealization III: Approximation*

and experimental investigations. Although this paper is concerned with the former, Galileo's vivid description of the experimental version is useful for conceptualizing the basic notion of Galilean idealization. When discussing the determination of gravitational acceleration in the absence of a medium devoid of resistance, Galileo suggests a kind of experimental idealization:

> We are trying to investigate what would happen to moveables very diverse in weight, in a medium quite devoid of resistance, so that the whole difference of speed existing between these moveables would have to be referred to inequality of weight alone.... Since we lack such a space, let us (instead) observe what happens in the thinnest and least resistant media, comparing this with what happens in others less thin and more resistant.[7]

Lacking a medium devoid of resistance, Galileo suggests that we can make some progress on the problem by initially using an experimental setup similar to the envisioned situation. After understanding this system, the scientist systematically removes the effect of the introduced distortion. The same type of procedure can be carried out in theorizing: introduction of distortion to make a problem more tractable, then systematic removal of the distorting factors.

Galilean idealization is justified pragmatically. We simplify to more computationally tractable theories in order to get traction on the problem. If the theorist had not idealized, she would have been in a worse situation, stuck with an intractable theory. Since the justification is pragmatic and tied to tractability, advances in computational power and mathematical techniques should lead the Galilean idealizer to de-idealize, removing distortion and adding back detail to her theories. With such advances, McMullin argues, "models can be made more specific by eliminating simplifying assumptions and 'de-idealization', as it were. The model then serves as the basis for a continuing research program" (*op. cit.*, p. 261). Thus the justification and rationale of Galilean idealization is not only pragmatic, it is highly sensitive to the current state of a particular science.

Galilean idealization is important in research traditions dealing with computationally complex systems. Computational chemists, for example, calculate molecular properties by computing approximate wavefunctions for molecules of interest. At first, all but the simplest

---

and Truth, Volume 25 of *Poznan Studies in the Philosophy of the Sciences and the Humanities* (Atlanta, GA: Rodopi, 1992), pp. 9–63.

[7] Quoted in McMullin, p. 267.

systems were intractable. When electronic computers were introduced to computational chemistry, calculated wavefunctions remained crude approximations, but more complex, chemically interesting systems could be handled. As computational power has continued to increase in the twenty-first century, it has become possible to compute extremely accurate (but still approximate) wavefunctions for moderate sized molecules. Theorists in this tradition aim to develop ever better approximations for molecular systems of even greater complexity.[8] These techniques are still approximate, but research continues to bring computational chemists closer to the goal of "[calculating] the exact solution to the Schrödinger equation, the limit toward which all approximate methods strive."[9]

This example nicely summarizes the key features of Galilean idealization. The practice is largely pragmatic; theorists idealize for reasons of computational tractability. The practice is also nonpermanent. Galilean idealization takes place with the expectation of future de-idealization and more accurate representation.

Minimalist idealization. Minimalist idealization is the practice of constructing and studying theoretical models that include only the core causal factors which give rise to a phenomenon. Such a representation is often called a *minimal model* of the phenomenon. Put more explicitly, a minimalist model contains only those factors that *make a difference* to the occurrence and essential character of the phenomenon in question.

A classic example of a minimalist model in the physical sciences is the Ising model. This simple model represents atoms, molecules, or other particles as points along a line and allows these points to be in one of two states. Originally, Ernst Ising developed this model to investigate the ferromagnetic properties of metals. It was further developed and extended to study many other phenomena of interest involving phase changes and critical phenomena. The model is powerful and allows qualitative and some quantitative parameters of substances to be determined. But it is extremely simple, building in almost no realistic detail about the substances being modeled. What it seems to capture are the interactions and structures that

[8] There are principled reasons why the exact wavefunction for multi-electron systems can not be computed. However, there are no general, in-principle reasons why approximations of arbitrarily high degrees of accuracy and precision cannot be computed.

[9] J.B. Foresman and A. Frisch, *Exploring Chemistry with Electronic Structure Methods* (Pittsburgh: Gausian, 1996), p. 95. For a discussion of the relevant philosophical issues, see Paul Humphreys, "Computer Simulation," in Arthur Fine, Micky Forbes, and Linda Wessels, eds., *Philosophy of Science Association 1990, Volume 2* (East Lansing, MI: PSA, 1992), pp. 597–609, and *Extending Ourselves* (New York: Oxford, 2004).

really make a difference, or the core causal factors giving rise to the target phenomenon.

Among recent discussions of idealization in the philosophical literature, minimalist idealization has been the most comprehensively explored position. As such, there is some diversity among the articulations of this position. One view is Michael Strevens's *kairetic* account of scientific explanation. Strevens's account of explanation is causal; to explain a phenomenon is to give a causal story about why that phenomenon occurred. What makes Strevens's account distinct is that the explanatory causal story is limited to only those factors that made a difference to the occurrence of the phenomenon. "Making a difference" is a fairly intuitive notion, but Strevens defines it explicitly in terms of what he calls "causal entailment,"[10] which involves logical entailment in a causal model. A causal factor makes a difference to a phenomenon just in case its removal from a causal model prevents the model from entailing the phenomenon's occurrence. A causal model of the difference-making factors alone is called a *canonical explanation* of the target phenomenon.

For Strevens, idealization is the introduction of false but nondifference-making causal factors to a canonical explanation. In explaining Boyle's law, for example, theorists often introduce the assumption that gas molecules do not collide with each other. This assumption is false; collisions do occur in low-pressure gases. However, low-pressure gases behave as if there were no collisions. This means that collisions make no difference to the phenomenon and are not included in the canonical explanation. Theorists' explicit introduction of the no-collision assumption is a way of asserting that collisions are actually irrelevant and make no difference.[11] Even with this added, irrelevant factor, the model is still minimalist because it accurately captures the core causal factors.

Other accounts of minimalist idealization associate minimalism with generation of the canonical explanation alone. Robert Batterman's account of asymptotic explanation is an example of such a view. Asymptotic methods are used by physicists to study the behavior of model systems at the limits of certain physical magnitudes. These methods allow theorists to study how systems would behave when certain effects are removed, which allows the construction of "*highly idealized minimal models* of the universal, repeatable features of a sys-

[10] Strevens, "The Causal and Unification Accounts of Explanation Unified—Causally," *Noûs*, XXXVIII (2004): 154–76.

[11] Strevens, "Why Explanations Lie: Idealization in Explanation" (unpublished manuscript, September 2004), p. 26.

tem."[12] These minimal models have a special role in physics because they can be used to explain universal patterns, common behaviors across material domains such as pressure, temperature, and critical phenomena. Adding more detail to the minimal model does not improve the explanations of these patterns; more details only allow a more thorough characterization of a highly specific event.

Arguing in a similar vein, Stephan Hartmann describes cases where highly complicated systems are characterized using physical models "of (simple) effective degrees of freedom," which help to give us "partial understanding of the relevant mechanisms for the process under study." This plays a cognitive role by allowing theorists "to get some insight into the highly complicated dynamics" of such systems.[13]

Cartwright's account of *abstraction* is also an example of what I call minimalist idealization.[14] On her view, abstraction is a mental operation, where we "strip away—in our imagination—all that is irrelevant to the concerns of the moment to focus on some single property or set of properties, 'as if they were separate'." If the theorist makes a mathematical model of this abstract, real phenomenon, then she is in possession of a minimal model. Such a model can reveal the most important causal powers at the heart of a phenomenon.[15]

Despite the differences between minimalist idealization and Galilean idealization, minimalist idealizers could in principle produce an identical model to Galilean idealizers. For example, imagine that we wanted to model the vibrational properties of a covalent bond. A standard way to do this is to use a harmonic oscillator model. This model treats the vibrating bond as spring-like with a natural vibrational frequency due to a restoring force. This is a very simple representation

[12] Batterman, "Asymptotics and the Role of Minimal Models," *British Journal for the Philosophy of Science*, LIII (2002): 21–38. See also Batterman, *The Devil in the Details* (New York: Oxford, 2001).

[13] Hartmann, "Idealization in Quantum Field Theory," in Niall Shanks, ed., *Idealization in Contemporary Physics* (Atlanta, GA: Rodopi, 1998), pp. 99–122.

[14] Cartwright distinguishes this view from what she calls idealization, which is closer to Galilean idealization. In a more recent defense of this distinction, Martin Jones has cogently argued that abstraction is best seen as a kind of omission, whereas idealization is the assertion of falsehood. Cartwright's and Jones's proposal is perfectly reasonable—omission and distortion are distinguishable practices. However, since I am arguing for pluralism about the nature of idealization, I see no reason why we should not treat minimalist modeling as a form of idealization. See Jones, "Idealization and Abstraction: A Framework," in Jones and Cartwright, eds., *Idealization XII: Correcting the Model: Idealization and Abstraction in the Sciences* (New York: Rodopi, 2005), pp. 173–217, for a careful defense of the alternative view. Also see Humphreys, "Abstract and Concrete," *Philosophy and Phenomenological Research*, LV (1995): 157–61, for a criticism of Cartwright's view and an argument that idealization (in Cartwright's sense) will almost always.come along with abstraction in real scientific contexts.

[15] Cartwright, *Nature's Capacities*, p. 187.

of the vibrational properties of a covalent bond, but one that is commonly used in spectroscopy. Galilean idealizers would justify the use of this model by saying that it is pragmatically useful for calculating energies, thus avoiding having to calculate the many-dimensional potential energy surface for the whole molecule. Minimalist idealizers, however, would justify the use of this model by suggesting that it captures what really matters about the vibrations of covalent bonds. The extra detail in the full potential energy surface, they would argue, is extraneous.

As this example illustrates, the most important differences between Galilean and minimalist idealization are the ways that they are justified. Even when they produce the same representations, they can be distinguished by the rationales they give for idealization. Further, while Galilean idealization ought to abate as science progresses, this is not the case for minimalist idealization. Progress in science and increases in computational power should drive the two apart, even if they generate the same model at a particular time.

Just as there is no single account of minimalist idealization, there is no single account of its justification. However, all of the influential accounts described above agree that minimalist idealization should be justified with respect to the cognitive role of minimal models: they aid in scientific explanations. Hartmann argues that minimal models literally tell us how phenomena behave in a simpler world than our own. This gives us the necessary information to explain real-world phenomena. For Batterman, minimal models demonstrate how fundamental structural properties of a system generate common patterns among disparate phenomena. Strevens and Cartwright look at things more causally, describing the role of minimal models as showing us the causal factors that bring about the phenomenon of interest. In all of these cases, minimalist idealization is connected to scientific explanation. Minimal models isolate the explanatorily causal factors either directly (Cartwright and Strevens), asymptotically (Batterman), or via counterfactual reasoning (Hartmann). In each case, the key to explanation is a special set of explanatorily privileged causal factors. Minimalist idealization is what isolates these causes and thus plays a crucial role for explanation. This means that unlike Galilean idealization, minimalist idealization is not at all pragmatic and we should not expect it to abate with the progress of science.

Multiple-Models Idealization. Multiple-models idealization (hereafter, MMI) is the practice of building multiple related but incompatible models, each of which makes distinct claims about the nature and causal structure giving rise to a phenomenon. MMI is similar to minimalist idealization in that it is not justified by the possibility of de-idealization back to the full representation. However, it differs from both Galilean and minimalist idealization in not expecting a single best model to be generated. This type of idealization is most closely associated with a distinctive kind of theorizing called *modeling*[16] or *model-based science.*[17]

One most commonly encounters MMI in sciences dealing with highly complex phenomena. In ecology, for example, one finds theorists constructing multiple models of phenomena such as predation, each of which contains different idealizing assumptions, approximations, and simplifications. Chemists continue to rely on both the molecular orbital and valence bond models of chemical bonding, which make different, incompatible assumptions. In a dramatic example of MMI, the United States National Weather Service employs three complex models of global circulation patterns to model the weather. Each of these models contains different idealizing assumptions about the basic physical processes involved in weather formation. Although attempts have been made to build a single model of global weather, the NWS has determined that the best way to make high fidelity predictions is to employ all three models, despite the considerable expense of doing so.[18]

The literature about MMI is less well developed then the others, so there is less of a clear consensus about its justification. But one especially important justification of MMI is the existence of *tradeoffs*, a position closely associated with biologist Richard Levins and his philosophical allies.[19] This justification begins by noting that theorists have different goals for their representations, such as accuracy, precision, generality, and simplicity. Levins further argues that these desiderata and others can trade off with one another in certain circumstances, meaning that no single model can have all of these properties to the highest magnitude. If a theorist wants to achieve

[16] Weisberg, "Who Is a Modeler?"

[17] Peter Godfrey-Smith, "The Strategy of Model Based Science," *Biology and Philosophy*, XXI (2006): 725–40.

[18] Details about the three primary models, as well as a number of others employed by the NWS can be found at http://www.meted.ucar.edu/nwp/pcu2.

[19] Levins, "The Strategy of Model Building in Population Biology," in Elliott Sober, ed., *Conceptual Issues in Evolutionary Biology* (Cambridge: MIT, 1984), pp. 18–27; Jay Odenbaugh, "Complex Systems, Trade-Offs and Mathematical Modeling: A Response to Sober and Orzack," *Philosophy of Science*, LXX (2003): 1496–507; Weisberg, "Qualitative Theory and Chemical Explanation," *Philosophy of Science*, LXXI (2004): 1071–81, and "Forty Years of 'The Strategy': Levins on Model Building and Idealization," *Biology and Philosophy*, XXI (2006): 623–45; John Matthewson and Weisberg, "The Structure of Tradeoffs in Scientific Modeling" (manuscript). For a critique of these ideas, see Steven H. Orzack and Sober, "A Critical Assessment of Levins' 'The Strategy of Model Building in Population Biology'," *Quarterly Review of Biology*, LXVIII (1993): 533–46.

high degrees of generality, accuracy, precision, and simplicity, she will need to construct multiple models. Levins summarizes his discussion of these issues as follows:

> The multiplicity of models is imposed by the contradictory demands of a complex, heterogeneous nature and a mind that can only cope with few variables at a time; by the contradictory desiderata of generality, realism, and precision; by the need to understand and also to control; even by the opposing esthetic standards which emphasize the stark simplicity and power of a general theorem as against the richness and the diversity of living nature. These conflicts are irreconcilable. Therefore, the alternative approaches even of contending schools are part of a larger mixed strategy. But the conflict is about method, not nature, for the individual models, while they are essential for understanding reality, should not be confused with that reality itself (*op. cit.*, p. 431).

Our cognitive limitations, the complexity of the world, and constraints imposed by logic, mathematics, and the nature of representation, conspire against simultaneously achieving all of our scientific desiderata. Thus, according to Levins, communities of scientists should construct multiple models, which collectively can satisfy our scientific needs.

Several other justifications for MMI can be found in the literature. Wimsatt argues that highly idealized models are important because, taken together, they help us develop truer theories.[20] Population biologists Robert May and Joan Roughgarden argue that clusters of simple models increase the generality of a theoretical framework, which can lead to greater explanatory depth.[21] Finally, Strevens's account of idealization can also be used to justify MMI. For Strevens, a theorist first finds a minimal causal model for a phenomenon of interest. She idealizes when she makes this highly abstract model more concrete, and in doing so introduces (nondifference making) distortions. The processes of filling in the minimal causal model with concrete details can be carried out in different ways, hence this process can yield multiple, idealized models.

Some of these motivations suggest strong parallels between MMI and minimalist idealization. In some cases, one cannot build a single minimal model that contains all of the core causal factors for a class of phenomena. Yet it may be possible, in such cases, to build a small set of models, each of which highlights a different factor and which

[20] Wimsatt, "False Models as a Means to Truer Theories," in M. Nitecki and A. Hoffmann, eds., *Neutral Models in Biology* (New York: Oxford, 1987), pp. 23–55.

[21] Jonathan [Joan] Roughgarden, *Theory of Population Genetics and Evolutionary Ecology: An Introduction* (New York: Macmillan, 1979); May, *Stability and Complexity in Model Ecosystems* (Princeton: University Press, 2001).

together account for all of the core causal factors. This motivation for MMI is parallel to the motivation for minimalist idealization, even though the practice itself is different.

However, there are additional motivations for engaging in MMI that do not parallel the motivation for minimalist idealization. For example, modelers may engage in MMI strictly for the purpose of maximizing predictive power, as do the forecasters at the National Weather Service. Another instance of MMI may involve building a set of models that gives maximum generality, at the expense of capturing all of the core causal factors. Still another is the synthetic chemist or engineer's motivation for MMI: to find the set of idealized models that is maximally useful for creating new structures. There are thus many motivations for MMI. Some are pragmatic, where scientists are focused on prediction and structure construction, while some are explanatory and nonpragmatic.

MMI also gives a complex, mixed answer about the permanence of idealization as science progresses. In some domains, MMI may abate with the progress of science. The National Weather Service may one day discover a single model that makes optimal predictions. However, if tradeoffs exist between theoretically important desiderata in a particular domain, then we should not expect MMI to abate with further progress. These tradeoffs are consequences of logic and mathematics and thus present a permanent justification for MMI.

From the discussion so far, it may seem that the literature on idealization describes a hodgepodge of disparate practices, leaving no hope for any further analysis of idealization simpliciter. This worry is not without merit because the methods, goals, and justifications of these three forms of idealization are quite distinct. Although a fully unified account of the three kinds of idealization is impossible, some progress can be made towards developing a unified framework with which to understand the practice of idealization in general. This framework focuses on the *goals* associated with idealization, rather than the activities or products of it. I call these goals the *representational ideals* of idealization.

II. REPRESENTATIONAL IDEALS

Representational ideals are the goals governing the construction, analysis, and evaluation of theoretical models. They regulate which factors are to be included in models, set up the standards theorists use to evaluate their models, and guide the direction of theoretical inquiry. Representational ideals can be thought of as having two components: *inclusion rules* and *fidelity rules*. Inclusion rules tell the theorist which kinds of properties of the phenomenon of interest, or *target* *system*, must be included in the model, while fidelity rules concern the degrees of precision and accuracy with which each part of the model is to be judged.

An important, albeit very simple, representational ideal is called COMPLETENESS, which is associated with classic accounts of scientific method. As such, it forms an important background against which every kind of idealization can be discussed.

COMPLETENESS. According to COMPLETENESS, the best theoretical description of a phenomenon is a complete representation. The relevant sense of 'completeness' has two components associated with its inclusion rules and fidelity rules, respectively. The inclusion rules state that each property of the target phenomenon must be included in the model. Additionally, anything external to the phenomenon that gives rise to its properties must also be included in the model. Finally, structural and causal relationships within the target phenomenon must be reflected in the structure of the model. COMPLETENESS's fidelity rules tell the theorist that the best model is one that represents every aspect of the target system and its exogenous causes with an arbitrarily high degree of precision and accuracy.

The description of COMPLETENESS given so far is accurate, but potentially misleading. With very few exceptions, the inclusion and fidelity rules of COMPLETENESS set a goal that is impossible to achieve. Unless extremely self deceived, or in possession of an extremely simple and abstract target system, no theorist thinks that complete representation is actually possible. Given the impossibility of achieving complete representation, how can COMPLETENESS play a guiding role in scientific inquiry?

Despite its unattainable demands, COMPLETENESS can guide inquiry in two ways. First, COMPLETENESS sets up a scale with which one can evaluate all representations including suboptimal ones. If a theorist wants to rank several representations of the same phenomenon and has adopted COMPLETENESS, she has a straightforward way to do so. The closer a representation comes to completeness, the better it scores. I call this the *evaluative function* of the representational ideal because it sets the standards for evaluating sub-optimal representations.

The second and more important way that COMPLETENESS can guide inquiry is through its *regulative function*. Regulative functions are similar to what Kant called *regulative ideals*. They do not describe a cognitive achievement that is literally possible, rather, they describe a target or aim point. They give the theorist guidance about what she should strive for and the proper direction for the advancement of her research program. If a theorist adopts COMPLETENESS, she knows that she should always strive to add more detail, more complexity, and more precision to her models. This will bring her closer to the ideal of completeness, although she will never fully realize this goal.

COMPLETENESS is a unique representational ideal because it directs theorists to include everything in their representations. All other ideals will build in some aspect of approximation or distortion. In thinking about ideals other than COMPLETENESS, we can begin to see the outline of a framework for characterizing the three kinds of idealization. Different kinds of idealization will be associated with different representational ideals. Before we carry this analysis forward, let us consider several additional representational ideals.

SIMPLICITY. After COMPLETENESS, the next most straightforward ideal is SIMPLICITY. The inclusion rule for this ideal councils the theorist to include as *little* as possible, while still being consistent with the fidelity rules. The fidelity rule for SIMPLICITY demands a qualitative match between the behavior of target system and the properties and dynamics of the model.

SIMPLICITY is primarily employed by working scientists in two contexts.[22] The first is pedagogical. Students are often introduced to the simplest possible model that can make sense of the data, even where scientists believe that the model contains serious problems. One example of this is in the Lewis electron pair model of chemical bonding. This model is not even quantum mechanical, yet it can be used to account for many canonical molecular structures. Beginning students are introduced to this model as a way of building intuitions about chemical structure and reactivity.

The second scientific context where SIMPLICITY is employed is when theorists construct models to test general ideas. "A minimal model for an idea tries to illuminate a hypothesis .... [It] is not intended to be tested literally, any more than one would test whether the models for a frictionless pulley or a frictionless inclined plane are wrong."[23] This second use represents a motivation and justification for a particular kind of modeling in scientific practice. Theorists often begin a project by trying to determine what kind of minimal structures could generate a property of interest. They do not need to know, at first,

[22] There is also a long tradition which investigates the epistemic role of simple models. In some circumstances, it seems that simple models ought to be preferred because they are more likely to be true. This is a different kind of justification for the use of simple models than I am discussing in this article. For a recent defense of the possible epistemic significance of simplicity, see Malcom Forster and Sober, "How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions," *British Journal for the Philosophy of Science*, XLV (1994): 1–35.

[23] Joan Roughgarden, *Primer of Ecological Theory* (Upper Saddle River, NJ: Prentice Hall, 1998), p. x.

how a specific target system actually works. Once the dynamics are understood in simple models, theorists examine more complex models and empirical data to assess the plausibility of the simple model's explanation of a real system's behavior.

1-CAUSAL. This representational ideal instructs the theorist to include in the model only the core or primary causal factors that give rise to the phenomenon of interest. Put in the language of the causation literature, this ideal tells the theorist to only include the factors that made a difference. The theorist constructs a mathematical model of a much simpler system than the one actually being studied, one that excludes higher order causal factors. These are the factors which make no difference to the occurrence of the phenomenon, but control the precise way in which the phenomenon occurs.[24] This is closely related to SIMPLICITY, but unlike SIMPLICITY, 1-CAUSAL restricts the level of simplicity that is allowed. If we are trying to construct the simplest possible model that can make predictions qualitatively compatible with our observations, there is no restriction on the kind or number of causal factors that must be included. SIMPLICITY, for example, may allow us to neglect all quantum mechanical effects and use the Lewis model. 1-CAUSAL, however, would not sanction the use of such a model because it requires the theorist to include the quantum mechanical interactions that compose the core physical explanation of the structure.

1-CAUSAL's fidelity criteria make a considerable difference in determining when the theorist has constructed an adequate model because its inclusion rule (restriction to primary causal factors) is not very specific. In addition, the fineness of specification of the target phenomenon itself will make a difference to the kind of model we can build. Imagine that we wanted to build a 1-CAUSAL model for the maintenance of the sex ratio. We would need a more complex model to explain the 1.05:1 ratio of male to female *Homo sapiens*, than if we only were interested in why the sex ratio is roughly 50:50. Even holding the fidelity criteria fixed, the best model would be different in these two cases, with the former requiring greater specification of internal and external causal factors.

Models generated using 1-CAUSAL are especially useful in two contexts. Like the models generated with SIMPLE, they can be used as starting points for the formulation and analysis of more complex models. 1-CAUSAL models are typically generated when one has a

[24] Of course, which factors do and do not make a difference to the occurrence of a phenomenon must be judged with respect to how precisely the phenomenon is individuated.

reasonably comprehensive understanding of how a system behaves, since knowing the primary causal factors that give rise to a phenomenon requires knowing quite a lot about the system. Further modeling from this point is usually aimed at greater quantitative accuracy, not deeper fundamental understanding.

The second context where 1-CAUSAL is especially important involves scientific explanation. Several recent philosophical accounts of scientific explanation have pointed to the central role that primary causal factors—the factors that really make a difference—play in scientific explanation.[25] Recent work on the cognitive psychology of explanation has also emphasized the crucial role that picking out central causal factors plays in people's judgments of explanatory goodness.[26] In their methodological discussions, a number of other scientists have commented on this connection. For example, chemist Roald Hoffmann emphasizes that "... if understanding is sought, simpler models, not necessarily the best and predicting all observables in detail, will have value. Such models may highlight important causes and channels."[27] These accounts all suggest that models generated with 1-CAUSAL seem to be at the heart of theorists' explanatory practices.

MAXOUT. We now move from an ideal which looks superficially like SIMPLICITY to one that looks superficially like COMPLETENESS, the ideal called MAXOUT. This ideal says that the theorist should maximize the precision and accuracy of the model's output. It says nothing, however, about how this is to be accomplished.

One way to work towards this ideal is by constructing highly accurate models of every property and causal factor affecting the target. This is the same approach taken in COMPLETENESS, although the goal of MAXOUT is to achieve maximum output precision and accuracy, not a complete representation. A second option, one more commonly associated with MAXOUT, is to engage in *model selection*,[28] a process of using statistics to choose a functional form, parameter set, and parameter values which best fit a large data set. The model selected by these techniques is then continually optimized as further data comes

[25] James Woodward, *Making Things Happen: A Theory of Causal Explanation* (New York: Oxford, 2003); Strevens, "The Causal and Unification Accounts of Explanation Unified—Causally."

[26] Tania Lombrozo, "The Structure and Function of Explanations," *Trends in Cognitive Science*, x (2006): 464–70.

[27] Hoffmann, V.I. Minkin, and Barry K. Carpenter, "Ockham's Razor and Chemistry," *Bulletin de la Société Chimique de France*, CXXXIII (1996): 117–30.

[28] Forster, "The New Science of Simplicity," in Arnold Zellner, Hugo Keuzenkamp, and Michael McAleer, eds., *Simplicity, Inference and Modelling* (New York: Cambridge, 2001), pp. 83–117.

in. Finally, MAXOUT also sanctions the use of black box models, the sort that have amazing predictive power, but for unknown reasons. These may be discovered using model selection techniques, or may be discovered in a more serendipitous fashion.

At first blush, it may seem unscientific to adopt an ideal that values predictive power over everything else. Most scientists believe that their inquiry is aimed at more than raw predictive power. While scientists want to know how a system will behave in the future, they also want an explanation of why it behaves the way that it does. MAXOUT ensures that we will generate models which are useful for predicting future states of the target system, but gives no guarantee that the models will be useful for explaining the behavior of the system.

Nevertheless, representations generated by MAXOUT have their place in scientific inquiry. Explanation and prediction are clearly both important goals of scientists, but there is no reason that they must both be fulfilled with the same model. Theorists can adopt a mixed representational strategy, using different kinds of models to achieve different scientific goals. It may also be rational to elevate predictive power above all other considerations in some situations. Following his reflection on the importance of simple models quoted above, Hoffmann argues that "If predictability is sought at all cost—and realities of marketplace and judgments of the future of humanity may demand this—then simplicity may be irrelevant."[29]

P-GENERAL. Generality is a desideratum of most models. This desideratum really has two distinct parts: *a-generality* and *p-generality*. A-generality is the number of actual targets a particular model applies to given the theorist's adopted fidelity criteria. P-generality, however, is the number of possible, but not necessarily actual, targets a particular model captures.[30] The representational ideal P-GENERAL says that considerations of *p-generality* should drive the construction and evaluation of theoretical models.

While a-generality may seem like the more important kind of generality, theorists are often interested in p-generality for several reasons. P-general models can be part of the most widely applicable theoretical frameworks, allowing real and nonreal target systems to be compared. P-generality is also often thought to be associated with explanatory power. This can be seen in both the philosophical literature on explanation and in the comments of theorists. An excellent example of

[29] Hoffmann, "Ockham's Razor and Chemistry."

[30] For further discussion, see Weisberg, "Qualitative Theory and Chemical Explanation," and Matthewson and Weisberg, "The Structure of Tradeoffs in Scientific Model Building."

this can be found in R.A. Fisher's discussion of modeling the nonactual. He begins by quoting Arthur Eddington:

> We need scarcely add that the contemplation in natural science of a wider domain than the actual leads to a far better understanding of the actual.[31]

Fisher goes on to argue:

> [for] a biologist, speaking of his own subject, [this] would suggest an extraordinarily wide outlook. No practical biologist interest in sexual reproduction would be led to work out the detailed consequences experienced by organisms having three or more sexes; yet what else should he do if he wishes to understand why the sexes are, in fact, always two (*ibid.*, pp. vii–ix)?

The key to understanding this actual system, Fisher argues, is to understand a possible, but nonactual one. In the behavior of this nonactual three-sex system lies the key to understanding why the two-sex system evolved. Some recent philosophical accounts of scientific explanation also stress the importance of p-generality to explanation.[32]

P-General can also play a subtler regulative role. Instead of trying to understand specific targets, theorists may wish to understand fundamental relationships or interactions, abstracted away from real systems. For example, ecologists may wish to study predation or competition, far removed from the interactions of particular species. In such cases, P-General is often adopted, guiding theorists to develop models that can be applied to many real and possible targets. This exploratory activity is a very important part of modern theoretical practice, although we do not yet have good philosophical account of how it works.[33] One thing we do know, however, is that there is a delicate balance between achieving deep and insightful p-generality and low-fidelity, uninformative p-generality, generated by overly simplistic models.

We have now looked at a number of representational ideals, the goals that guide theoretical inquiry. As I mentioned at the beginning of this section, representational ideals are at the core of the practice

[31] Fisher, *The Genetical Theory of Natural Selection* (New York: Oxford, 1930), pp. viii–ix, quoting Eddington's *The Nature of the Physical World*.

[32] Strevens, "The Causal and Unification Accounts"; Woodward, *Making Things Happen*.

[33] Some aspects of this exploratory mode of theorizing are discussed in Levins, "The Strategy of Model Building"; Wimsatt, "Robustness, Reliability and Overdetermination," in M. Brewer and B. Collins, eds., *Scientific Inquiry and the Social Sciences* (San Francisco: Jossey-Bass, 1981), pp. 124–63; Weisberg, "Robustness Analysis," *Philosophy of Science*, LXXIII (2006): 730–42; Patrick Forber, "On Biological Possibility and Confirmation" (unpublished manuscript).

of idealization and a systematic account of them can ultimately lead us to a more unified understanding of idealization. To that end, we now turn back to the three kinds of idealization and consider which representational ideals are associated with them.

III. IDEALIZATION AND REPRESENTATIONAL IDEALS

Recall that Galilean idealization is the practice of introducing distortions into theories in order to simplify them and make them computationally tractable. It is justified pragmatically, introduced to make a model more computationally tractable, but with the ultimate intention of de-idealizing, removing any distortion, and adding detail back to the model. Models generated by Galilean idealization are thus approximate, but carry with them the intention of further revision, ultimately reaching for a more precise, accurate, and complete model. The ultimate goal of Galilean idealization is complete representation; its representational ideal is thus COMPLETENESS.

Minimalist idealizers are not interested in generating the most truthful or accurate model. Rather, they are concerned with finding minimal models, discovering the core factors responsible for the target phenomenon. Minimalist idealizers thus adopt the representational ideal 1-CAUSAL, the ideal that says the best model is the one that includes the primary causal factors that account for the phenomenon of interest, up to a suitable level of fidelity chosen by the theorist. While minimalist idealizers may sometime look like they are adopting SIMPLICITY, this is almost always inaccurate, because theorists engage in minimalist idealization to really understand how the target phenomena work and why they behave the way that they do. This requires finding the causal factors that really do make a difference, not a model that simply can reproduce the phenomenon qualitatively.

Like Galilean idealization's representational ideal, minimalist idealization's ideal also demands the construction of a single model for a particular target or class of target phenomena. One typically engages in minimalist idealization in order to generate explanatory models. Such models tend to be ones that simultaneously unify many target phenomena into a class and identify the causal factors which really make a difference. For the class of phenomenon of interest, this will mean finding a single model, despite the fact that it will leave out quite a lot of detail which accounts for the uniqueness of each target.

Finally, we can consider MMI. The biggest difference between MMI and the other kinds of idealization is that there is no single representational ideal which is characteristic of it. Pretty much any representational ideal—including 1-CAUSAL and in rare cases

COMPLETENESS—can play a role in this form of idealization. MMI arises because of the existence of tradeoffs between different theoretical desiderata. This suggests that not all desiderata are simultaneously maximizable, at least in a single model. Thus the most significant aspect of MMI is that it instructs theorists to construct a series of models which pursue different desiderata and are guided by multiple representational ideals.

Consider, for example, the ecological research program that is concerned with understanding predation. A cursory look at the ecological literature on predation, reveals little in the way of the search for a single, best model of predation. Instead, one finds a series of models, some of which are more precise and accurate, some of which are more qualitative, some of which are very well suited for populations that are homogenously distributed in space, and some of which are flexible enough to deal with complex spatial structure. This situation is the norm in theoretical ecology. As John Maynard Smith explained, "For the discovery of general ideas in ecology ... different kinds of mathematical description, which may be called models, are called for."[34]

For modern ecologists pursuing MMI, a full understanding of the ecological world is going to depend on multiple, overlapping, possibly incompatible models. How might we justify this kind of pluralism? One possible approach is anti-realist. We could argue that maximizing empirical adequacy in some cases requires the use of multiple models. Since anti-realism only requires that models be empirically adequate, the use of different kinds of idealized models is unproblematic.

This line of response is available to anti-realists, but neglects some of the motivations for building multiple models that theorists have discussed in the literature. The same ecologists who champion the use of multiple models very explicitly describe this practice as aimed at having a more complete understanding of the phenomena of interest, not simply making accurate predictions. As Levins puts it, "[O]ur truth is at the intersection of independent lies" (*op. cit.*, p. 20). This is clearly a realist sentiment. To understand if it is justified we must ask whether the use of multiple idealized models, or the use of any idealized models at all, is compatible with scientific realism.

IV. IDEALIZATION, REPRESENTATIONAL IDEALS, AND THE AIMS OF SCIENCE

Peter Godfrey-Smith gives the following helpful formulation of scientific realism: "One actual and reasonable aim of science is to give us accurate descriptions (and other representations) of what reality

[34] Maynard Smith, *Models in Ecology* (New York: Cambridge, 1974), p. 1.

is like. This project includes giving us accurate representations of aspects of reality that are unobservable."[35] The realist thus believes that scientists aim and sometimes succeed at representing this external, independent reality, while anti-realists demur, at least when it comes to unobservables.

Prima facie, idealization looks like it might cause problems for scientific realism. All three forms of idealization involve the willful distortion of scientific representations. Willful distortion and approximation appears to militate against Godfrey-Smith's conception of realism, because the theorist is not even aiming to give an accurate description of what mind-independent reality is like. Despite this, I think all three kinds of idealization are compatible with the sort of realism sketched by Godfrey-Smith, if his definition is understood in a broad and sophisticated way.

Galilean idealization is the most straightforwardly compatible with scientific realism. Galilean idealizers often fall short of their representational ideal of COMPLETENESS and may even do so willingly. However, in the long run, the Galilean idealizer does aim to give complete, nondistorted, perfectly accurate representations. In order to accommodate the possibility of Galilean idealization, scientific realists need to understand that achieving accurate representations of complex phenomena is an ongoing process. Even when the short-term practice involves the willful introduction of distortion, the long-term aim can still be to give an accurate representation of what reality is really like. Thus scientific realism is perfectly compatible with Galilean idealization, if the realist aim is understood to be long-term or ultimate.

Minimalist idealization and MMI present more serious challenges to scientific realism. It will not be possible for minimalist and MMI idealizers to assent to at least one interpretation of Godfrey-Smith's formulation because they do not ever aim to give a *fully* accurate representation of reality. However, defenders of minimalist idealization aim to uncover real causal structure, or fundamental patterns in common between multiple phenomena. This suggests that a weaker reading of Godfrey-Smith's formulation, which does not require fully accurate representations, is compatible with minimalist idealization.

There are other respects in which minimalist idealization is compatible, and indeed demands a kind of realism. Consider the goals and justification of minimalist idealization: Minimalist idealizers are trying to model the most important causal factors that underlie the

[35] Godfrey-Smith, *Theory and Reality* (Chicago: University Press, 2003), p. 176.

properties and behaviors of target phenomena. That is, they often recognize that real scientific explanation involves the identification of the core causal factors giving rise to the system, not all of the details. This recognition is surely a realist one. While minimalist idealizers are decidedly not interested in the truth, the whole truth, and nothing but the truth, they want to know the truth about what really matters. For their explanatory interests, representation of just a few key factors is what matters. This representation must be accurate.

Finally, consider multiple models idealization. As this constitutes a more diverse set of practices, it is much harder to make a unified judgment about the degree of realism embodied by MMI. Some kinds of representational ideals are clearly not realist. For example, the ideal MAXOUT tells the theorist that she should seek maximal precision and accuracy in the output of her model. However, this ideal provides no guidance about the internal structure of the model and is compatible with black-box models. MAXOUT is also compatible with models that are willfully distorted with the sole aim of making the predictions more accurate. So clearly a practice of idealization that only uses MAXOUT is incompatible with realism. In such a case, theorists do not aim to give accurate representations of the underlying reality of their target phenomena.

As I described it, however, MMI transcends relying on any one kind of representational ideal. It is a strategy for investigating phenomena when complexity and tradeoffs preclude the accomplishment of this in a single model. When a theorist chooses to engage in MMI because her system is complex, but nevertheless wants to develop an accurate description of her target phenomenon, she is acting in a realist fashion. At least one aim of her practice is the development of a mind-independent picture of a real-world phenomenon. This attitude is also clearly realist in spirit, despite the fact that it will be strictly incompatible with Godfrey-Smith's definition of realism.

Indeed, there can be several realist dimensions of MMI. Typical episodes of MMI employ a package of representational ideals, which taken together, aim to give an accurate representation of real world systems. While no single model may contain the complete picture of the properties and behavior of a complex system, a collection of them can. Levins and Maynard Smith advocate the use of multiple models precisely because a collection of ecological models will give a more accurate representation of the behavior of real world ecosystems then any single one does.

The recognition that multiple models can give a more accurate and informative representation of real world systems is itself another realist dimension of MMI, in fact one might call it a *higher order* realist motivation for MMI. Multiple models idealization is justified by the existence of tradeoffs between theoretically important desiderata such a simplicity, accuracy, precision, and generality. If these tradeoffs exist in the way many scientists and philosophers believe that they do, adopting theoretical strategies that recognize them is the proper realist response. Where tradeoffs exist, the realist should not be content to choose a single, most accurate model. Such a strategy ignores important discoveries about the world, in this case about our representational capacities. When faced with tradeoffs and complex systems, the realist should surely follow the Levinsonian strategy of multiple-model use.

The goal of this discussion of realism is to show that, despite *prima facie* concerns about the incompatibility of idealization and realism, all three kinds of idealization can be made compatible with sophisticated forms of realism. A more detailed study of these issues would look more carefully at each representational ideal and consider the extent to which its fidelity criteria and inclusion rules are compatible with realism and other scientific desiderata.

What, then, have we learned about idealization and its justification? I have endeavored to show that the three kinds of idealization recognized in the philosophical literature are not competitors, but reflect three practices important to scientific inquiry. What distinguishes them is not the product of their application, but rather the representational ideals which guide theorists in using them. There is no single, over-arching justification for idealization. Differing representational ideals respond to the demands of a complex world in different ways. This precludes a single justification for idealization.

MICHAEL WEISBERG

University of Pennsylvania