Why Explanations Lie:
Idealization in Explanation

Michael Strevens

Draft of March 2007
The final version is chapter 8 of my book *Depth*

Abstract

On the causal approach to explanation, explaining a phenomenon is telling the actual causal story as to why it occurs. Science is full of idealizing explanations that deliberately falsify the relevant causal story. Therefore, either (a) the causal approach to explanation is mistaken, or (b) idealizing explanations, however convenient, are deeply flawed. Correct? I don't think so. I show that on an enlightened causal account of explanation, idealization is not merely tolerable; done right, it enhances the explanatory power of a causal model.

This paper is based on chapter 8 of Strevens (2008).

1. Introduction

Science proceeds, so often, using idealized models. Why? Let me narrow this question to a particular feature of idealization and a particular scientific endeavor. The endeavor is *explanation*; the feature of idealization—common, if not ubiquitous—is the deliberate falsification of the causal workings of the system whose behavior is to be explained.[1]

---

[1] For an illuminating discussion of a different aspect of idealization applied in the service of different scientific ends, see Cartwright (1999), especially chapter 4.

The explanatory use of causally inaccurate models is of special interest to an advocate of the causal approach to explanation. On the causal approach, to explain a phenomenon is to tell the actual causal story as to why it occurs. Idealizing explanations distort the relevant causal story, or worse. In the context of the causal approach, then, idealization looks like a mistake, to be avoided if at all possible.

Some autobiography: I am an advocate of the causal approach to explanation. But I am also an apostle of idealization: I believe that idealization has the power to make an explanation better, more enlightening, a greater contribution to scientific understanding not only as grasped by mere mortal minds, but in the abstract.

The aim of this paper is to reconcile these two apparently inconsistent faiths, by showing that a suitably sophisticated causal account of explanation not only makes space for causal idealization, but dictates precisely which causal details can be distorted, how, and to what explanatory end.

Idealizations, I will argue, convey explanatorily essential information, that is, information that must be grasped in order to understand perfectly the phenomena to be explained. Once the nature of this information becomes clear, it will be seen that

1. A good idealizing explanation is always better than its veridical counterpart, where by the veridical counterpart, I mean the explanation that corrects every one of the idealizing explanation's distortions, and

2. A good idealizing explanation is in one important respect explanatorily optimal: it cannot be improved (though there are other respects in which it is not optimal).

2. Explaining Boyle's Law

Before continuing, let me give an example of a causally distorting explanation, in fact, the example I will later use to develop an account of idealization: the explanation of Boyle's law using the ideal gas model.

Boyle's law states that, when kept at a constant temperature, the pressure of a fixed amount of gas varies inversely with its volume. Increase the volume and the pressure drops accordingly; decrease the volume and the pressure rises. In symbols,

$$PV = k$$

where P is pressure, V is volume, and the constant k's value is determined by the amount of gas and its temperature.

Boyle's law holds only under certain circumstances; in particular, the gas must be reasonably dilute. What's more, even inside this range it holds only approximately. Still, it is a striking enough relation that every textbook on physical chemistry explains why Boyle's law holds when it does—that is, approximately and within a certain range of conditions. This approximate holding within a range is the explanandum, then.

Later in this paper, I will examine the textbook explanation in some detail. For now, I will simply remind you of a few of the idealizations made by the ideal gas model in the course of its explanation of Boyle's law. The model fails to correspond to the causal structure of a real gas in the following three ways, among others:

1. It ignores the long range attractive forces between molecules,
2. It represents molecules as not colliding (indeed, as infinitely small), when in fact they do collide, and
3. It is invokes classical physics, yet the underlying processes are quantum processes.

3. Two Approaches to Idealization

It may be useful to look over two well-known approaches to thinking about the explanatory role of causal idealizations. I will not criticize either, but I will note that each violates one of my points of faith: either that the causal approach to explanation is correct, or that causal idealizations have intrinsic explanatory virtue.

3.1 The Pragmatic Approach to Idealization

On the pragmatic approach, idealization does not add to an explanation's intrinsic worth, and on most accounts detracts from it, but the explanatory loss is more than made up for by various non-explanatory advantages of the idealization, such as computational tractability. In the case of Boyle's law, for example, the inaccuracy of the ideal gas model due to its failure to represent the presence of long range attraction between molecules is compensated for by the relative ease of the derivation of the law from a model in which there are no such forces. Pragmatism, then, denies that idealization enhances the explanatory worth of an explanation; in doing so, it breaks with the second my two tenets.

Though the pragmatic view incorporates a view of idealization on which it is at best explanatorily neutral, even some of idealization's most prominent friends are pragmatists. For example, the central figure of the Poznań school, Leszek Nowak, maintains that the understanding provided by an idealized model would be enhanced if the model's distortions were replaced with the true causal story. Idealized models are stepping stones only: science eventually removes, or ought to remove, every "counter-actual" idealization in a process of "concretization" (Nowak 1992). An idealized model provides an "approximate explanation", on Nowak's view; only a fully concretized model—a model free of all idealization—can produce a "perfect explanation".[2]

---

[2] Nowak does refer to the factors that are falsified in an idealizing explanation as "inessen-

Pragmatists are not, of course, committed to the view that idealizing explanations are in principle misleading; a good idealized model, they will hold, clearly indicates which of its claims are intended literally and which are not (Railton 1981, 243). They do believe that such a model, however clearly it identifies its idealizations, is explanatorily only second best.

3.2 The Empiricist Approach to Idealization

According to the kind of empiricist view on which the aim of science is to save the phenomena, making up stories about the way that the phenomena are produced may not count against an idealized explanatory model. What's more, insofar as an idealization allows the model to save the phenomena more economically or more elegantly, idealization may be counted a good thing (see, e.g., Hempel (1965), 344–345). This is especially true on a unificationist approach to explanation, for which elegance and economy are themselves explanatory virtues.

Let me say a few words, then, about the unificationist treatment of explanation, on the understanding that, even if it does not have the allegiance of every empiricist, it does capture some of the most characteristic empiricist concerns.

The unificationist approach to explanation holds that a phenomenon is to be explained by the most unifying model covering that phenomenon, where the unifying power of a model is proportional to the number of actual phenomena modeled and the simplicity of the model.[3] A good explanatory

---

tial", but in his theory, there seems to be no qualitative difference, for explanatory or any other purposes, between inessential and essential factors: "essentiality" is a matter of degree. The quality of an explanation, on Nowak's view, is proportional to the essentiality of the factors cited; however, since he holds that idealization implies a failure to cite some factors of non-zero essentiality, idealizing explanations can always be improved by removing the idealizations. Hence his doctrine that the perfect explanation of a phenomenon will be idealization-free.

[3] I bracket the question of what it is for a model to cover, "save", or model a phenomenon. On Kitcher's influential unificationist account of explanation (Kitcher 1981, 1989), it is

model, then, is one that is relatively simple and that covers many actual phenomena.

It follows that a model's representation of the world's causal structure can be inaccurate without compromising its explanatory power. Some writers ask for a certain degree of causal accuracy,[4] but a full-blooded empiricist version of the unification account requires of models only that they conform to the appearances, that is, that their predictions about observable matters of fact be correct. It is these more empiricist variants of unificationism that especially interest me here.

How would a unificationist make sense of the idealizations that appear in the standard explanation of Boyle's law? Simple: they would argue that correcting the causal idealizations in the ideal gas model would increase its complexity without bringing any commensurate increase in the number of phenomena modeled, making for a decrease in unifying power. For the purpose of explaining Boyle's law, then, the ideal gas model is explanatorily superior to a more accurate representation of the inner workings of a real gas. Idealization, in this case, increases explanatory power; it is compulsory.

My aim here is not to argue against unificationism, but I should pause to note that there is a deep difficulty with the unificationist treatment of the Boyle's law case. The ideal gas model correctly predicts certain behaviors of gases: their adherence, approximately, to Boyle's law, Gay-Lussac's law, Avogadro's law, and so on. But it fails to predict a number of other behaviors, for example, the relatively slow speed of gaseous diffusion, which is due to intermolecular collisions. A causally more accurate model of a gas is able to account for all the former phenomena and the latter as well. The claim above that moving to a more accurate model results in a decrease in unifying power is therefore called into question: certainly, the move decreases the model's

---

simply for the occurrence of the phenomenon to be deducible using an instance of the model, where a model is conceived of as a kind of schematic argument.

[4] See, for example, Cartwright (1983). Cartwright is not offering a unification account of explanation, but her criteria for a good explanatory model are close to the unificationist's.

simplicity, but at the same time it increases the number of phenomena unified. It is unclear which model is the more unifying, then. (According to (Friedman 1974), it is the more accurate, unidealized model that unifies better .) Thus the unificationist account of idealization appears to fall through: our best idealizations seem sometimes not to result in the greatest degree of unification. That said, I imagine that a committed unificationist or other empiricist might find a way around this problem and so preserve a distinctive empiricist approach to idealization, perhaps by opting for a less holism in the determination of explanatory models. My own approach will be rather different.

4. A Causal Account of the Explanation of Laws

4.1 Introduction

My next task is to sketch a causal approach to the explanation of laws and other generalizations. This account is based on a causal theory of explanation that I have developed elsewhere, which I call the *kairetic account* (Strevens 2004, 2008). But it takes from the kairetic account only what is needed to support the account of idealization that I present in this paper; thus, what follows is a sketch, rather than a complete account, of the explanation of laws. In other circumstances, I would emphasize what is original and different about my approach to scientific explanation. Because I want my account of idealization to be self-standing, however, I will take a different tack here, doing my best to reassure you that what is assumed by the account of idealization is a set of features that anyone would want in a causal theory of law explanation. (Incidentally, though I talk about the explanation of laws, what I have to say applies also to the explanation of ongoing states of affairs, "effects", and other generalizations. I use the term *law*, then, in the broadest possible way.)

There is no canonical causal account of law explanation, but there is something of a consensus as to how such an account should look, which I wish to join: on the causal approach, laws are to be explained by describing the *relevant aspects* of the *underlying mechanism* that makes them true.

A causal account of the explanation of laws, then, will have two parts: a criterion for determining the causal mechanism underlying any given law, and an account of explanatory relevance that picks out the aspects of an underlying causal mechanism relevant to explaining the law. Some proponents of the causal approach to explanation will hold that all parts of the underlying mechanism are always relevant: they deny the need for a non-trivial account of relevance; however, as you will see, it is relevance that does much of the heavy lifting in the account of idealization.

The next two subsections correspond to the two parts of the account of the explanation of laws: they are an examination of underlying mechanisms, and an account of explanatory relevance. For the sake of brevity, the first part, the examination of underlying mechanisms, will be pursued only insofar as is necessary to set up the account of relevance.

4.2 Causal Mechanisms

To explain a law, the causalist says, exhibit the underlying mechanism. However, it is not entirely straightforward, in many cases, to decide exactly what the underlying mechanism consists in, and what supporting materials—generalizations about initial conditions or correspondence rules connecting higher and lower level vocabularies, for example—need to be added to a description of the mechanism in order to complete the explanation.

Rather than try to settle these questions here, I will consider only the most straightforward case, the explanation of a causal law of the form "if *S* then *T*", where *S* and *T* are themselves connected by a causal process. The development of this example will provide a good enough basis for the characterization of explanatory relevance, which as just remarked is what is really important for understanding idealization.

Consider a simple example of an if/then causal generalization:

> The apple law: If an apple falls from a tree at time t, it will hit the ground at time $t + r\sqrt{2d/GM}$.

where d is the distance between the apple and the ground, G is the gravitational constant, and M is the mass and r is the radius of the Earth. I take it that this generalization has a few implicit riders that are understood by any competent reader: the tree is on the Earth's surface, nothing out of the ordinary interferes with the apple's flight, and so on. What's more, since the time taken for real apples to fall is not exactly equal to the time given by the formula in the apple law, I assume that what's being explained is an *approximate relation* between time taken and distance fallen. (The unpalatable alternative is that a false generalization is being explained.)

On the causal approach, in order to explain this generalization, you must describe the causal mechanism in virtue of which any particular apple detached from its tree under the assumed circumstances in fact falls in the stated time. The mechanism is, of course, the operation of the Earth's gravity pulling the apple to the ground.[5] The rest of this section suggests a certain approach to describing the operation of the underlying mechanism; the advantages of the descriptive system will become clear when relevance is discussed in the next section.

In the case where the underlying causal process is deterministic, I represent the operation of the mechanism using a deductive argument (or argument schema) that has the antecedent of the law as one of its premises and the consequent as its conclusion. The premises of the argument representing the mechanism of the apple law might be as follows:

1. The apple becomes detached at time t.
2. The distance of the apple from the ground is d.

---

[5] For the sake of this example, I will assume a Newtonian rather than an Einsteinian conception of gravity.

3. A causal law, that the Earth exerts a gravitational force of magnitude GMm/r^2 on an object of mass m on its surface, causing a constant rate of downwards acceleration equal to GM/r^2.

4. A mathematical fact: an object experiencing a constant acceleration a will cover a distance d in time $\sqrt{2d/a}$.

The conclusion is, of course, that the apple reaches the ground at time $t + r\sqrt{2d/GM}$. Note that each step in the deduction (in this simple example, there is only one step) corresponds to a causal process by which the states of affairs and laws asserted by the premises together causally produce the state of affairs asserted by the conclusion.

Let me generalize this technique for representing causal mechanisms, introducing three pieces of terminology along the way.

1. To explain a causal law of the form "if S then T", you must exhibit the causal mechanism by which a given S causes a given T. Call the description of the mechanism a *causal model.*

2. In the case where the law to be explained is deterministic, the mechanism can be described by a deductive argument taking an occurrence of S as one of its premises and having the resulting occurrence of T as its conclusion. Note that the conclusion of the relevant argument does not assert the truth of the explanandum, the law, but rather the instantiation of the consequent of the law. Call the state of affairs asserted by the consequent the *target* of the causal model.

3. If the deductive argument is to represent the causal mechanism by which the explanatory target is introduced, then each step in the argument must correspond to a causal step in the operation of the mechanism. It is only in virtue of its satisfying this requirement that the argument can be said to represent the operation of the mechanism. When the requirement is met, say that the argument *causally entails* the

> occurrence of the target (noting, first, that this is a requirement on the structure of the argument, not only on its premises, and second, that the requirement is not satisfied in virtue of the argument's structure alone, but in virtue of a relation between that structure and the causal structure of the world). The details of causal entailment are worked out in Strevens (2008).

In summary, a deterministic causal mechanism is to be represented by a causal model, which is a deductive argument that causally entails the explanatory target.

4.3 Difference-Making and Explanatory Relevance

A causal factor may satisfy all of the conditions stated in the last section for being a part of the causal mechanism underlying a given law, yet may be counted as explanatorily irrelevant.

Let me give an example, using the case of the apple law. It seems right to say that a part of the causal mechanism determining the time taken for an apple to fall from its tree is the air resistance that slows its fall. Another, far slighter causal influence is the gravitational pull of Earth's moon. The inclusion of these influences in the falling mechanism is reflected in the fact that propositions asserting the presence of both factors could be inserted in the deductive argument laid out in the last section, and the argument would not only remain deductive, but more importantly, would continue to causally entail the explanatory target: each step in the deduction would correspond to the operation of a real causal influence on the time taken for the apple to fall, hence on the the event realizing the explanatory target. Both our everyday explanatory practice and the criterion stated in the last section concur, then, that air resistance and lunar pull are a part of the causal mechanism determining the time taken for the apple's descent.

Nevertheless, it seems wrong to cite either as a part of the explanation of the apple law. Let me remind you exactly what the apple law asserts. It says that a falling apple takes a time of approximately $r\sqrt{2d/GM}$ to reach the ground. Air resistance and lunar pull play no part in explaining the fact that the time taken has this approximate form. You do not understand any better why, say, the time taken varies in proportion with the square root of d by taking into account air resistance or the position of the moon. Though these factors causally influence the time taken for the apple's fall, they do not explain the dependences that constitute the law to be explained. Thus they are explanatorily irrelevant.

In the light of these observations I propose, as foreshadowed above, that a causal account of the explanation of laws will have two parts: an account of causal mechanisms, and an account of relevance that determines which aspects of a given causal mechanism are relevant to the explanation of a given law.

What determines explanatory relevance, then? Let me begin with the following familiar suggestion. Suppose that T is the explanatory target for the law to be explained, that is, the state of affairs asserted to obtain by the conclusion of the deductive argument representing the relevant causal mechanism. Then an aspect of the causal mechanism is explanatorily relevant to the law just in case it *makes a difference* to whether or not T obtains.

Consider, for example, the apple law. The state of affairs T asserted by the consequent is the apple's hitting the ground at time approximately $t + r\sqrt{2d/GM}$. A part of the causal mechanism influencing the fall is explanatorily relevant to the apple law, then, if it makes a difference to the fact that this approximate equality holds. Air resistance and lunar pull make a difference to the time taken for the fall, but they do not make a difference to the fact that the time taken is close to the time entailed by the apple law. Thus they are irrelevant to the explanation of the apple law.

As you will see, this difference-making criterion for explanatory relevance has its bite because the law to be explained, the apple law, picks out some aspects, but not others, of the complete state of affairs produced by the relevant mechanism, the complete state of affairs being the exact time taken for the apple's fall. Though there is (let's say) an exact relation that holds between the time taken for the fall and all the physical parameters having any causal influence on the fall, the apple law asserts only a high level property of the relation: the approximate form of its dependence on a subset of the parameters characterizing the mechanism, namely, d, r and M. This is, I note, typically true of the laws of nature: they tell us not about every detail of the phenomenon they cover, but only about some relatively high level properties of the phenomenon.

I have explicated explanatory relevance in terms of difference-making, but how to explicate difference-making? Other approaches to explanation that have in various ways appealed to facts about difference-making have tended to look to some sort of counterfactual criterion: a part of the causal mechanism producing an event makes a difference to an aspect of the event—its exhibiting some high level property—just in case, were the mechanism part not present, the event, though presumably different in its specifics, would still have exhibited the given high level property (Lewis 1986; Woodward 2003).

The kairetic account of explanation makes a somewhat different suggestion. To test whether a causal factor A makes a difference to the explanatory target T, take a causal model for T that includes a premise asserting the presence of A. The model represents a part of the causal mechanism, including A, sufficient to causally produce T. Now remove the premise asserting the presence of A. If the premises that remain are no longer sufficient to causally entail T, then A is a difference-maker, otherwise not. In short, then, A makes a difference to T—it plays an essential role in the causal production of T—just in case a premise asserting the presence of A plays an essential role in causally entailing T in the corresponding deterministic causal model. Note that, by removing the premise asserting the presence of A, you are not considering a new causal model in which *A* is not present, but rather, a model in which it is not specified whether or not *A* is present.

The criterion for difference-making can be extended in an important way. Suppose that you have a premise in your deterministic causal model that specifies some exact value for a parameter, and suppose also that the argument goes through if this premise is exchanged for one that specifies only that the parameter falls within a certain range. Then I say that the exact value of the parameter does not make a difference to the law to be explained, but that what does make a difference is the fact of the parameter's falling inside the given range. The discussion of Boyle's law will supply an example.

The idea that some variety of relevance can be decided by determining what does and does not play an essential deductive role in an argument has a long and checkered history in the philosophy of science. The problems began with Hempel and Oppenheim's requirement that a DN explanation must essentially involve a law of nature. They continued with attempts in confirmation theory to use the essential role criterion to rescue hypothetico-deductivism from certain relevance problems (Glymour 1980, chap. 2). Fetzer (1981) makes a similar proposal for explanation. We now know that all such accounts of relevance are vacuous if no constraints are placed on the deductive arguments in which an essential role must be played.[6] The causal approach has the advantage that it puts a strong constraint on the arguments, namely that they must not only entail, but must causally entail, their conclusions; this solves the problems of the "essential deductive role" approach to relevance, as explained in Strevens (2004, §4.4) and Strevens (2008).

---

[6] Though some writers hold out the hope that a purely syntactic constraint will solve the problem (Gemes 1997).

5. The Textbook Explanation of Boyle's Law

Let me begin my treatment of idealization by taking a closer look at the way that textbooks explain Boyle's law, starting with a very simple explanation of the sort that appears in an introductory text such as Ebbing (1987), and then augmenting it to obtain a more sophisticated textbook explanation of the law.

The simple explanation has two parts. First, there is a physical picture of a gas as a collection of very small, very fast-moving molecules. Second, there is a set of assumptions about the behavior of the molecules from which the law is derived. The assumptions are:

A-1. The pressure that a gas exerts on a container is proportional to the frequency of collisions between the gas's molecules and the container's walls per unit area of wall. (It also depends on the force exerted by the colliding molecules.)

A-2. The frequency of molecule/wall collisions per unit area of a container's walls is proportional to the density of the gas. (It also depends on the velocity of the local molecules.)

A-3. The density of a gas is inversely proportional to its volume. (It also depends on the number of molecules in the gas.)

If pressure is proportional to collision frequency, collision frequency is proportional to density, and density is inversely proportional to volume, then pressure is inversely proportional to volume, as Boyle's law states.

The explanation has two weaknesses. First, the assumptions A-1, A-2, and A-3 are not themselves accounted for. Assumption A-2, in particular, is not obviously true, and is thus itself in need of explanation. Second, the assumptions include parenthetical complications that are ignored without any justification. More sophisticated explanations of Boyle's law, such as that offered by McQuarrie and Simon (1997), §25–4, fill both gaps. In what follows, I show how the gaps are filled, and I spell out some of the idealizations that the sophisticated textbook explanation makes along the way.[7]

Begin with the parenthetical complications. The explicit conditions attached to Boyle's law, namely, the fact that the temperature and the quantity of the gas remain fixed, provide some justification for ignoring these qualifications. In particular, if the quantity of gas is fixed, then the number of molecules in the gas remains constant, so A-3's assertion of the dependence of density on volume *and* molecule number can be collapsed to an assertion of the inverse proportionality of density to volume alone, as required for the derivation of the law.

In order to ignore the parenthetical complications in A-1 and A-2, by contrast—in order to assume (roughly) that the velocity of the molecules near the container's walls remains constant as the gas's volume changes—something more than a constant temperature must be assumed. Constant temperature entails that the overall mean velocity of the molecules in the gas remains constant. This does not rule out the possibility that the mean velocity in the vicinity of the container's walls varies with, say, the container's shape. The natural bridging assumption is, then

D-1. The velocities of the gas molecules near a container's walls have the same statistical profile as in the gas as a whole.

From this and constant temperature it follows that the force and frequency of the collisions remains the same under a change in volume, so that the change in pressure is determined solely by the change in the density, hence by the change in volume.

The statistical posit D-1 is not obviously true. (In fact it is not true at all, because long range attractive forces between gas molecules exert an inhibitory effect on molecules heading away from the rest of the gas, as a result of which

---

[7] The sophisticated treatments more or less follow Clausius (1857), although Clausius's demonstration of A-2 is not entirely general; see Clausius (1857), §15.

molecules on the edge of the gas—molecules, that is, near the container walls—will tend to hit the walls with a lower average velocity than if D-1 were true. D-1 is, however, approximately true, which is good enough for what follows.)

Since it is nonobvious, more sophisticated derivations of Boyle's law justify, rather than merely stating, D-1. This justification invokes further assumptions about the distribution of the properties of gas molecules. All such assumptions are consequences of the Maxwell-Boltzmann probability distribution over the properties of the molecules of a gas in thermal equilibrium, a distribution that yields the probability that a molecule will be found in a specified region, or traveling in a given direction or at a given speed (i.e., velocity magnitude) when there are no complicating factors (e.g., gravitational fields).

The property of the Maxwell-Boltzmann distribution that founds D-1 is

D-2. The speeds, directions of travel, and positions of a gas's molecules are stochastically independent; they are not correlated.

From D-2 it follows immediately that a molecule close to a container wall is just as likely to have a given speed or direction of travel as a molecule anywhere else, and so that D-1 is true; likewise, D-1's approximate truth follows from D-2's approximate truth. This concludes the justification for ignoring the parenthetical complications in A-1, A-2, and A-3.

The other improvement made by sophisticated explanations of Boyle's law is, you will recall, the justification of A-2. This is achieved by showing that A-2 follows from D-2 and two further properties of the Maxwell-Boltzmann distribution:

D-3. A gas's molecules are (with high probability) at all times evenly distributed throughout its container.

D-4. The directions of travel of a gas's molecules are (with high probability) evenly distributed among all the possibilities; that is, all directions are equally probable.

The demonstration proceeds by calculating the number of molecules that will collide with a small portion of the container wall over a short period of time, given the distribution assumptions and some assumptions about molecular dynamics. One major idealization is made in the course of the calculation, that nothing interferes with the molecules as they travel around the container, and in particular, that molecules do not collide with one another.

Call the explanation presented above the textbook explanation of Boyle's law. Let me summarize the idealizations made by the textbook explanation. First, there is the assumption that there are no long range forces between gas molecules, implicit as already noted in D-1 and therefore in the other distribution assumptions from which D-1 is derived. Second, there is the assumption that there are no intermolecular collisions, invoked as already noted in the derivation of A-2. A third idealization is that the combined volume of the gas molecules themselves is zero, that is, that molecules take up no space in the container. This is, strictly speaking, redundant: it is not required if it is already assumed that there are no collisions. I include it because it is well known that correcting for this idealization—given that there are intermolecular collisions—leads to a more accurate version of Boyle's law. (I am referring of course to van der Waals' law, which also corrects for the effect of long range forces.) Fourth, the justification of A-1 assumes that molecule/wall collisions are perfectly elastic. A fifth and final idealization is the assumption of classical, rather than quantum, mechanics throughout.

The five idealizations, then, are as follows (reordered for later reference).

I-1. The volume occupied by the gas molecules is vanishingly small compared to the volume of the container.

I-2. The molecules exert no long range forces on one another.

I-3. Collisions between the molecules and the walls are perfectly elastic.

I-4. The molecules do not collide with one another (or interact in any other way at short range).

I-5. The behavior of the gas is as described by classical mechanics.

(This list is not intended to be an exhaustive.)

When the density of a gas is low enough, I-1 and I-2 are approximately true, in the sense that the volume occupied by the molecules is almost vanishingly small and the long range intermolecular forces are almost zero. In this sense, they are "minor" idealizations. I-3 might also perhaps be said to be minor (although it will turn out to have more in common with I-4).

I-4, by contrast, is nowhere near true. I-5 is also a major idealization, because the quantum behavior of molecules is quite different in many respects from their classical behavior. In other explanations using kinetic theory, where collisions are taken into account, the physics used is not only non-quantum, it is not even classical, in that molecules are assumed to bounce off one another as billiard balls appear to do, rather than, as they actually do, repel one another at very short distances.

6. The Canonical Explanation of Boyle's Law

I will contrast the textbook explanation of Boyle's law with what I call the *canonical explanation* of Boyle's law, the canonical explanation being the explanation recommended by my own causalist account of explanation, sketched in section 4. Whereas the textbook explanation contains idealizations, the canonical explanation contains none. A comparison of the two explanations will reveal the explanatory role of idealization.

This section sketches the canonical explanation of Boyle's law; the next three sections develop the theory of idealization. One conclusion that will be possible at the end of this section, before putting forward any particular thesis about the nature of idealization, is as follows: if the distortion of the causal details perpetrated by the ideal gas explanation of Boyle's law were corrected—that is, if the distorted description were replaced by a true description of the same details—there would be no improvement in the quality of the explanation.

Suppose, then, that your task is to construct the canonical explanation of the approximate truth of Boyle's law. Following the account of the explanation of laws developed in section 4, you should create a causal model that has as its explanatory target the approximate inverse proportionality of any gas's pressure to its volume, which I will call the gas's *Boylean behavior*.

The underlying mechanism—the causal process responsible for producing the explanatory target in any given gas—is the mechanism by which a given volume of gas in a given container exerts a given pressure on the container walls, namely, the steady bombardment of the walls by the gas's molecules. A complete description of this mechanism for a particular gas would have to document the trajectory of every molecule in the gas. The kinetic theory of gases allows you to give a statistical description of this mechanism using the Maxwell-Boltzmann distribution that is, for all practical purposes, just as good. Take this statistical description of the bombardment mechanism as your starting point. (I argue in Strevens (2008) that the statistical description is in fact explanatorily superior to the more complete deterministic description.)

Having found and described the underlying mechanism, the next step is to prune the mechanism description so that it mentions only aspects of the mechanism that make a difference to the explanatory target. This is where things will get interesting.

A full description of a gas in modern kinetic theory accommodates all of the causal complications omitted in the ideal gas model: long range forces, collisions, and so on. Include all of these factors in the initial causal model, then, along with whatever probabilistic assumptions are required to determine changes in pressure given changes in volume. The result is a description of a causal mechanism from which you can derive not only Boyle's law, but much else besides, such as the more sophisticated and more accurate law relating a gas's pressure and volume (and other properties) formulated by van der Waals.

Your job is to remove from this description everything that makes no difference to the inverse proportionality of pressure to volume, that is, everything that is not essential to the causal entailment of Boylean behavior. It does not matter that the stripped down model will be a far less rich theory of gases; your aim is not a theory of gases—you have one already—but the explanation of a very particular behavior of gases.

Without invalidating your model's causal entailment of Boylean behavior, you can discard the following elements. First, you can remove the specification of the gas molecules' volume and of the long range intermolecular forces between them, substituting just the statement that the molecules are very small and the forces very weak. You are replacing the exact values for the volumes and the forces, then, with the specification that they fall into certain ranges, beginning at zero and ending at some value that is not too high. What is important for Boyle's law is that the actual volumes and forces are somewhere within these ranges, so that, for example, D-1 is approximately true.

Second, you can remove certain aspects of the Maxwell-Boltzmann probability distribution over the various properties of gas molecules. For example, kinetic theory ascribes a certain probability distribution over the speeds, or velocity magnitudes, of the gas molecules. The shape of this distribution makes no difference to Boyle's law. As you can see from the discussion in section 5, what is important are the properties required for D-1 to hold, namely, D-2, D-3, and D-4.

Third, the details of intermolecular collisions make no difference to the derivation. Or almost no difference: it must be specified that, say, a collision between two molecules cannot double the speed of both, but the necessary limitations, such as the conservation of energy, are already spelled out elsewhere in the model, for example, in the physics of molecule/wall collisions. All the physics that is particular to the collisions, then, can be removed: whether collisions are billiard-ball style bounces (as in "hard sphere" models of gases) or are really short range repulsions (as in real gases); if repulsions, the nature of short range forces, and much more. As so often in applications of the criterion for difference-making, particular parameters will be replaced by ranges of parameters, but in the case of collisions, these parameters can take on just about any values at all compatible with kinetic theory's "big picture" of gases as composed of large numbers of small, fast-moving molecules. Some of these values will imply that there are no collisions at all (short range forces of zero strength in a repulsion physics, or molecules with zero size in a hard sphere physics). The abstracted model will not even require, then, that collisions occur.[8]

Fourth, you can remove various details of the physics of molecule/wall interactions. For example, in a real gas, there will be transfers of energy from molecules to walls, and from walls to molecules. These will balance out (since the walls are made up of molecules at the same temperature as the gas). The rates of transfer back and forth can be ignored; all that need be specified is that the transfers have no cumulative effect.

Fifth, many properties of a gas's container can be left unspecified. Most interestingly, the container can be any shape you like. That shape does not matter is already implicit in the very statement of Boyle's law, so it comes as no revelation, but it is worth remarking on.

Sixth and finally, any other elements of physical theory not needed to derive A-1 and A-2 can and should be left unspecified in the model. As a result, the model may well encompass both classical and quantum physics,

---

[8] Collisions might appear in a yet more detailed explanation of Boyle's law that characterizes the process by which a gas reaches equilibrium. The explanations I consider in the main text assume that the gas is already at equilibrium.

and many other possible, but non-actual, physics of the world.

Once all the factors listed above have been abstracted away, what remains in the model are just those high level properties of the mechanics of kinetic theory that make a difference to gases' Boylean behavior. It is by appreciating that these and only these properties are the difference-makers, that you understand Boyle's law.

A suggestive conclusion about idealization follows immediately. A model that makes all of the idealizations listed in the previous section—each of I-1 to I-5—instantiates the canonical causal model for the law's underlying mechanism. Though non-veridical, it contains all the elements of the best explanation of the law.

In this respect, it is no worse than the veridical, overly detailed model to which you applied the optimizing procedure. Both models state or entail all the explanatorily relevant properties of gases, but then go on to state further irrelevant properties as well. It does not matter, for the purposes of explanatory power, whether the irrelevant details specified by a model are veridical or non-veridical. Either way they contribute nothing positive to understanding, and they detract from understanding in the same way, by falsely claiming relevance for irrelevancies. The difference-making conception of explanatory relevance explains, then, why the idealized model does not suffer for its falsehoods.

What is not explained is why the idealized model does not suffer for its irrelevancies. That will require a novel proposal about the meaning and purpose of idealization that is independent of the account of difference-making.

7. Idealization in the Textbook Explanation of Boyle's Law

One half of the theory of idealization is in place. I have shown (for the case of Boyle's law, but the lessons are of course supposed to generalize) that the causal factors distorted by idealized models are details that do not matter to the explanatory target—they are explanatory irrelevancies. The distortions of the idealized model are thus mitigated. But how, if at all, do they make a positive contribution to the explanation?

I suggest that the idealizations in the textbook explanation of Boyle's law are not to be taken at face value. This is in part because the explainers, the textbook writers, know they are not true, but also in part because they know they are irrelevant.

Immediately after deriving Boyle's law, the authors of one textbook write

> We . . . assumed that the molecules of the gas do not collide with each other . . . But if the gas is in equilibrium, on the average, any collision that deflects the path of a molecule from [the path assumed in the derivation] will be balanced by a collision that replaces the molecule (McQuarrie and Simon 1997, 1015).

The no-collision assumption is justified, then, by an argument, or rather an assertion, that collisions or not, the demonstration of Boyle's law goes through.

This is, I suggest, advice on how to interpret the idealizing explanation. McQuarrie and Simon are not merely warning their readers not to take the no-collisions assumption literally. If that were their aim, then they would be using a proposition (no collisions) to do some work, then adding that the proposition is false and cannot do the work but that fortunately another proposition (collisions balance out) is true and can do the work instead. Such a strategy would increase the complexity of the exposition without reducing the ultimate cognitive load on the reader.

The authors must have something different in mind, then. My proposal: the assumption that there are no intermolecular collisions is supposed to communicate to the reader that collisions have no net effect on the relationship between volume and pressure. One way to say that collisions have no net effect is to say that it is *as if* there are no collisions; this is how the idealizing explanation is to be understood. If I am right, then on the subject of collisions, the canonical model and the idealized model say precisely the same thing: collisions make no difference to Boylean behavior, and thus are irrelevant to the explanation of Boyle's law. The idealized model, then, adopts the optimal explanatory policy on collisions.

All idealizations, I suggest, work in the same way: an idealization does not assert, as it appears to, that some non-actual factor is relevant to the explanandum; rather, it asserts that some actual factor is irrelevant. The best idealized models will be equivalent in one explanatorily central sense to the corresponding canonical models: when understood correctly, both kinds of models cite the same relevant factors and no irrelevant factors. However, they represent these explanatory facts by way of two different conventions: the fact that certain pervasive causal influences play no role in bringing about the explanatory target is left implicit in a canonical model, while it is made explicit in an idealizing explanation's flagrant introduction of fictional physical factors.

This interpretation of idealization will be bolstered by examining the other idealizations in the explanation of Boyle's law, showing that they are constructed to assert—as the corresponding canonical model does in its own way—that certain factors make no explanatory difference to the target.

First, consider I-1 and I-2, the assumptions that the gas molecules have no volume and that there are no long range intermolecular forces. To set the volume and the forces to zero is to say, on my view, that neither the volume nor the forces play any role in producing Boylean behavior.

Second, consider I-3, the assumption that molecule/wall collisions are perfectly elastic, that is, that during a collision, none of a molecule's kinetic energy is transferred to the walls as heat energy. Setting the energy transfer to zero indicates that energy transfer is not a difference-maker for Boylean behavior.

Compare the following textbook justification of the elasticity assumption:

> Since pressure is a property averaged over many wall collisions, we assume that in any one wall collision, there is no change in the molecule's translational kinetic energy. Although this assumption is false, it is "true" averaged over all the molecules, and hence gives the correct result for pressure (Levine 2002, 458).

The point is that only the average energy transfer affects the pressure, so that individual transfers make no difference to the pressure provided that they cancel out. For the purpose of explaining pressure, the idealization implies, it is *as if* there is no energy transfer at all.

Third, what about I-5, the use of classical mechanics? Again the intent of the idealization is to state that much of the underlying physics makes no difference to Boylean behavior, and in particular, that the quantum aspects of our world's physics make no such difference. Classical mechanics is selected for the idealized model not because it sets some "quantum parameter" to zero, but because it epitomizes for us the non-quantum—it is the natural way for physics to be non-quantum. That this is so is in part a matter of historical contingency, of course, but what is important for the idealization to play its role are not its detailed workings, but the simple fact that it is the non-quantum default.

8. The Nature and Purpose of Idealization

To generalize the discussion of Boyle's law: the role of an idealization is to assert the explanatory irrelevance, that is, the failure to make a difference, of a salient causal factor. The three characteristics of an idealization are

1. It is evidently false. This signals that its function is to make a claim about what does not make a difference, rather than about what does make a difference.

2. The false claim fills out certain details left unspecified by the canonical explanatory model. Thus the claim is explanatorily irrelevant, and so cannot stand in the way of the causal entailment of the explanatory target.

3. The details are filled out in a certain way: the relevant parameters are assigned a zero, an infinite, or some other default value. This is the idealization's way of asserting that the actual details do not matter. (More exactly, the details either do not matter at all, or they do not matter provided that they lie in the vicinity of the default value. I will return to this equivocation shortly.)

The content of an idealized model, then, can be divided into two parts (cf. Railton 1981, 243). The first part contains the difference-makers for the explanatory target, and if the model is perfect, is identical to the canonical model. The second part is all idealization; its overt claims are false but its role is to point to parts of the actual world that do not make a difference to the explanatory target.

At the beginning of this paper, I promised to single out two respects in which idealization could be said to be explanatorily virtuous:

1. An idealizing explanation is always better than its veridical counterpart, and

2. An idealizing explanation is as good in one important sense as the corresponding canonical explanation.

I will establish these two claims in turn.

First, the comparison with the veridical counterpart. An idealized model's veridical counterpart is the model that corrects the idealized model's causal distortions. Where the idealized model sets a parameter, such as the strength of long range intermolecular forces, to zero, the veridical counterpart states the correct value for the parameter. All falsified causal details in an idealizing explanation are explanatorily irrelevant: they are non-difference-makers. The veridical counterpart, by correctly describing these details, falsely implies that they are difference-makers. It is therefore explanatorily inferior to the idealized model, which, properly interpreted, correctly identifies them as non-difference-makers.

Second, the comparison with the canonical model. I have already shown that an idealized model in one sense does the same explanatory work as the corresponding canonical model: both models correctly specify all the difference-makers and all the non-difference-makers. Let me now introduce a well-known distinction between two senses of explanation, ontological and communicative. An explanation in the ontological sense is a collection of explanatory facts, namely, the causal laws and initial conditions picked out by an explanatory model, whereas an explanation in the communicative sense is a representation of such facts. One way to individuate explanations in the ontological sense is by the difference-makers they cite: two explanatory models for the same target are identical explanations of that target if they cite precisely the same difference-makers. On this view, it would appear that an idealized model represents exactly the same explanation of its target as the corresponding canonical model. If objective explanatoriness is your sole concern, there is consequently nothing to choose from between the idealizing explanation and the corresponding canonical explanation of a phenomenon.

Although this position is consonant with my greater goal of finding an explanatory virtue in idealization, it must be resisted, I think, for two reasons. First, the proposed individuation criterion for ontological explanations attends only to an explanatory model's statement of the matters of fact and underlying laws from which it derives the explanatory target. But a causal model has another part besides this assembly of materials, namely, the derivation itself. The first part, the list of causal factors, specifies the difference-makers. The second part, a deduction of the explanatory target from the listed factors, represents the way in which they make a difference. This second part of the model surely ought to contribute to explanation individuation. If the derivation is taken into account, however, it can be argued that an idealizing explanation is different from, and inferior to, a canonical explanation, for the following reason. Although idealized models specify, and often highlight in a useful way, what does not make a difference, they are much less clear—quite often, completely silent—about why it does not make a difference. For example, although McQuarrie and Simon add a parenthetical comment on the balancing out of intermolecular collisions to their explanation of Boyle's law, they do not provide enough information to see for sure that balancing out will occur. Contrast this with the canonical model, which by demonstrating that Boylean behavior can be derived without making any specific assumptions about the nature of collisions, shows why the physics of collisions makes no difference to the behavior, thus is irrelevant to the explanation of Boyle's law. The canonical model in this way provides more objectively explanatory information than the idealized model.

Second, notwithstanding what I have said above, an idealized model is sometimes less clear than the corresponding canonical model when it comes to indicating what makes a difference. Suppose that a parameter in the idealized model is falsely assigned a value of zero. It follows that the exact value of the parameter makes no difference. But does this mean that the value of the parameter makes no difference at all—that it can take any value without affecting the occurrence of the explanatory target—or that, while its exact value makes no difference, it makes a difference that the parameter falls into a certain range? If the latter, what is the extent of the range? The idealized model on its own answers neither question (except to say that the range includes zero).

These are important differences between an idealized model and the corresponding canonical model. When an idealized model is presented in an informative context, however, they may disappear: appropriate annotations, or the reader's knowledge of the subject matter or of the writer's intentions, may supply the explanatory information that is missing from the idealized model considered in isolation. In other words, in context the canonical and idealizing explanations may well communicate precisely the same objectively explanatory information, namely, the information made explicit in the two parts of the canonical model. There is a sense, then, in which the two explanations are identical. There is also a sense—the communicative sense, in which an explanation's means of representation are taken into account—in which they are rather different, since the canonical model identifies non-difference-makers by not mentioning them, whereas an idealized model identifies non-difference-makers by blatantly distorting their properties. Finally, there is a sense in which the two explanations are very similar but not quite identical, namely, the sense that attends to the objective explanatory facts communicated by the models in isolation.

Why idealize, rather than giving a canonical explanation? Here, at last, I will turn to considerations that are not intrinsic to the explanatory enterprise. There are three reasons to construct idealized models.

First, there are some causally salient factors that are, for all their salience, irrelevant to the explanation of certain phenomena. It is important to appreciate their irrelevance. If the factors are idealized, their irrelevance is underlined in a dramatic way—more dramatic, certainly, than a canonical explanation's knowing silence.

The second reason to idealize is also a practical one, familiar from the utilitarian approach to idealization (section 3.1), but unlike the first, it has nothing to do with communication. Because idealizations assign zero or default values to their idealized parameters, an idealized model is simpler than a model that fills out the same details veridically; it is also often simpler, because it replaces ranges with definite values, than the canonical model. As a consequence, the derivation of the explanatory target using an idealized model is typically more straightforward than the derivation using any other veridical causal model. The explanation of Boyle's law is a good example: it is easier to derive Boyle's law from an assumption of zero long range forces and no collisions than to derive it from the assumption that the parameters governing long range forces and collisions fall into certain ranges.

The third reason to idealize stems from the fact that, because an idealized model distorts only non-difference-makers, it is as effective as an instrument of prediction as the canonical model, provided that the prediction is of the kind of phenomenon relative to which the distorted details make no difference. (The ideal gas model is good for predicting changes in pressure due to changes in volume, but bad for predicting diffusion behavior, since collisions make a difference to diffusion behavior.) Given that idealizations often, as just explained, make a model simpler to work with, idealized models are good scientific all-rounders: they make very good explanations, but also have virtues that are more important to the predictive enterprise, such as simplicity and tractability. A certain economy is gained by using the same model for both explanatory and predictive work; this helps to account for the prevalence of idealization in explanatory models.

In summary, then, though an idealizing explanation is in certain ways inferior to a canonical explanation, there are considerations of communicative effectiveness, descriptive and computational simplicity, and scientific economy that motivate the widespread use of idealization in explanation. None of these considerations would count for much, however, unless idealizing explanations were intrinsically very good explanations. It is in this respect that my view of idealization differs from the utilitarian's (section 3.1). Let me conclude this section of the discussion, then, by reminding you of how excellent an idealizing explanation can be: it is vastly superior to its veridical counterpart, and is as good, in one important sense, as the best possible explanation.

* * *

To test the view of idealization presented here, consider some other well-known idealizations in science. First, to explain the approximately parabolic trajectory of a cannonball, air resistance is set to zero. This indicates that air resistance, provided it is low, makes no difference to the trajectory's approximate shape.

Second, in explaining the appearance of a rainbow, it is assumed that raindrops are perfect spheres (Batterman (2002)'s example). In fact, local forces will tend to deform the drops slightly. By assuming zero deformation, the model asserts that deformations within the normal range make no difference to the existence of rainbows.

Third, in explanatory models in population biology, infinite populations are assumed. The effect of this assumption is to set the expected rate of genetic drift to zero. What such a model says is that drift, though present, made no difference to the feature of the gene pool to be explained, say, the extinction of an inferior trait. (More precisely, the only property of drift that made a difference was its not being too large.)

Fourth, when a genetic explanation assumes Mendel's law of independent assortment, it asserts that the correlation between any two alleles' being passed on to the same offspring is zero. Translation: the correlation, whatever it was, made no difference to the explanatory target.

Fifth, and perhaps most notoriously, much explanation in economics assumes that people are maximally rational in certain respects, for example, that they have a strictly transitive ordering of preferences. The influence of aspects of normal human psychology that may in various respects falsify this assumption is set to zero. What is asserted, I claim, is not that economic actors are inhuman, but that the normal aspects of human behavior that depart from the robotic utility maximization of *Homo economicus* make no difference to whatever phenomenon is in the course of being explained.

9. Structural Idealization

An act of idealization, I have suggested, sets one of the canonical explanatory model's parameters to zero, infinity, or some default value. This characterization should, I now argue, be expanded in the light of a case from biological modeling.

In a typical model in population ecology or evolutionary biology, individuals are represented only in the aggregate: the model consists of mathematical rules describing the dynamics of populations or sub-populations as a whole. Some modelers, however, aspire to represent elements of individual behavior; they build computer programs, called agent-based models, that keep track of the state of every organism in a system.

The explanatory power of such models has been contested; on the kairetic approach, in particular, they are optimally explanatory only when the phenomenon to be explained depends on features of the dynamics of individuals that cannot be captured by population level models, and thus for the most part, only for rather fine-grained explananda, something of a rarity on the ecological and evolutionary side of the biological sciences.

Still, such explananda exist, and their province may be rather wider than I have suggested. Suppose, then, that you have an agent-based model that captures the difference-makers for some biological event or phenomenon. What idealizations might legitimately be made? One simplification found in agent-based models is the discrete representation of a real-valued variable, as when, say, an organism's level of nutrition is allowed to take only whole-numbered values between one and ten, or its position is represented on a grid in which all creatures occupying the same patch of ground are assigned exactly the same position. (Discrete representations are of course more or less unavoidable in models built on digital computers.)

Such idealizations are allowed, I propose, when the exact value of the variable in question makes no difference to the explanatory target, that is, when the canonical explanatory model specifies ranges of values rather than exact values for the variables. Provided that the ranges are wide enough that they each include at least one of a granular model's discrete values, that model will instantiate the canonical model. It is not perspicuous to think of a granular model as having been derived from the corresponding canonical model by setting certain parameters to zero, infinity, or default values. Better to think of the model as formed by way of a structural simplification of the canonical model, and more particularly, by a kind of concretization of certain of its abstract aspects, achieved by the substitution of particular values for specified ranges. (This is rather different from Nowak's conception of concretization, then (section 3.1). Note that when idealizing by setting parameters to zero, infinity, or default values, there may be a uniquely natural choice of idealized model. This is not so for granular models, since there are usually, perhaps always, many equally good choices of sets of discrete values.)

I place structural simplification alongside the setting of parameters to zero or default values as a device for idealization. Other examples of structural idealization include the use of biological models assuming asexual reproduction to explain the dynamics of sexually reproducing populations, the use of the Ising model in statistical physics to explain certain properties of phase transitions (since the material substrate in the Ising model has a simpler physical structure than that of the materials whose behavior it is invoked to explain), and perhaps the explanatory appeal to *Homo economicus* in economic models.

I have been discussing cases in which it is the idealized model that is discrete and reality continuous; sometimes, however, it is the model that is continuous where reality is discrete. The best known example is the representational apparatus of fluid mechanics, in which fluids are assumed to have the structure of a continuum rather than their true atomic structure. So which is it: is a discrete model a structural simplification of a continuous model, or is a continuous model a structural simplification of a discrete model?

It may well be that what constitutes an appropriate idealizing simplification varies from model to model. Even given a particular model, the standards of appropriateness may vary with the scientific practice or the cognitive and social constitution of the explainers. There is no need for a single objective standard in determining this element of idealization.

The standards for judging an idealization may therefore be divided into two parts, one objective and one perhaps not so objective. The objective standard is straightforward: an idealized explanatory model for a phenomenon must instantiate a canonical explanatory model for that explanatory target, so guaranteeing that the elements of reality falsely represented by the model are non-difference-makers. The idealized model can therefore be understood as the conjunction of the canonical model and certain further, false claims about reality. The less objective standard concerns the choice of these false claims: they should set parameters of the canonical model to zero or default values, or they should amount to a structural simplification of the canonical model. It is the terms *default* and *simplification* that introduce the subjectivity into the standard.

In short, as to the accuracy of its representation of the facts about difference-making, an explanatory model answers to the facts themselves, which are observer-independent (or at least, as observer-independent as the standards for determining difference-making). But as to the efficiency with which it communicates those facts, and in particular, the efficiency with which it distinguishes the true claims that single out difference-makers from the false claims that single out non-difference-makers, an idealized explanatory model answers to everything relevant to communicative effectiveness, which will surely include standards that vary from locale to locale.

Why Explanations Lie

Do idealizing explanations lie? If taken literally, yes. If understood correctly, no: what appear to be lies are true claims about the identity of the non-difference-makers. Explanations lie, then, in order to communicate a certain kind of truth, a truth about what does not matter to the causal production of the explanandum.

References

Batterman, R. W. (2002). *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence.* Oxford University Press, Oxford.

Brush, S. (2003). *The Kinetic Theory of Gases: An Anthology of Classic Papers with Historical Commentary.* Imperial College Press, London.

Cartwright, N. (1983). The simulacrum account of explanation. In *How the Laws of Physics Lie.* Oxford University Press, Oxford.

———. (1999). *The Dappled World: A Study of the Boundaries of Science.* Cambridge University Press, Cambridge.

Clausius, R. (1857). Ueber die Art der Bewegung, welch wir Wärme nennen. *Annalen der Physik* 100:353–380. English translation reprinted in Brush (2003).

Ebbing, D. D. (1987). *General Chemistry.* Second edition. Houghton Mifflin, Boston.

Fetzer, J. (1981). *Scientific Knowledge: Causation, Explanation, and Corroboration.* D. Reidel, Dordrecht.

Friedman, M. (1974). Explanation and scientific understanding. *Journal of Philosophy* 71:5–19.

Gemes, K. (1997). A new theory of content II: Model theory and some alternatives. *Journal of Philosophical Logic* 26:449–476.

Glymour, C. (1980). *Theory and Evidence.* Princeton University Press, Princeton, NJ.

Hempel, C. G. (1965). Aspects of scientific explanation. In *Aspects of Scientific Explanation*, chap. 12, pp. 331–496. Free Press, New York.

Kitcher, P. (1981). Explanatory unification. *Philosophy of Science* 48:507–531.

———. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher and W. C. Salmon (eds.), *Scientific Explanation*, volume 13 of *Minnesota Studies in the Philosophy of Science*, pp. 410–505. University of Minnesota Press, Minneapolis.

Levine, I. N. (2002). *Physical Chemistry*. Fifth edition. McGraw-Hill, New York.

Lewis, D. (1986). Causal explanation. In *Philosophical Papers*, volume 2, pp. 214–240. Oxford University Press, Oxford.

McQuarrie, D. A. and J. D. Simon. (1997). *Physical Chemistry: A Molecular Approach*. University Science Books, Sausalito, CA.

Nowak, L. (1992). The idealizational approach to science: A survey. In J. Brzezinski and L. Nowak (eds.), *Idealization III: Approximation and Truth*, volume 25 of *Poznań Studies in the Philosophy of the Sciences and the Humanities*, pp. 9–63. Rodopi, Amsterdam and Atlanta, GA.

Railton, P. (1981). Probability, explanation, and information. *Synthese* 48:233–256.

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

———. (2008). *Depth: An Account of Scientific Explanation*. Harvard University Press, Cambridge, MA.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, Oxford.