

Chapter 1: Introductory Remarks on Explanation, Idealization, and Ineliminability

Nicholaos Jones
jones.1736@osu.edu

October 2006

For the most part, scientific explanations are answers to why-questions. A scientific explanation of some phenomenon P answers the question ‘Why is it the case that P obtains?’ One might ask why straight sticks appear to be bent when immersed in water, why some metals become magnetic, why the moon eclipses the sun, etc. Science aims to answer these questions. This explanatory task goes beyond the descriptive task of science, because explanations do not merely describe what is the case; they also show why it is the case.

A scientific explanation can be thought of as a set of sentences, one of which – the *explanandum* – stands in a certain relation to the others – the *explanans*. The explanandum is a sentence to the effect that the phenomenon to be explained obtains; there is always a presupposition that the explanandum is true, because there is no explanation of why a false sentence is true nor of why a phenomenon that does not obtain obtains. The explanans is the set of sentences that purport to explain the explanandum. Whether an explanans explains its explanandum depends upon whether there is an appropriate relationship between them; this relationship distinguishes explanations from non-explanations, preventing all but a privileged few answers to a why-question from being explanatory.

Thinking of scientific explanation in this way overlooks some issues about scientific explanation that happen to be tangential to the focus of this dissertation. For instance, there has been some debate about whether sentences, or only the facts that sentences describe, do explanatory work.

No substantial thesis in this dissertation hinges upon which of these views is correct. For this reason, I take the liberty of sometimes treating a member of the explanans as a fact rather than a sentence describing the fact. I also assume the irrelevance of any difference between explaining why a sentence is true and why the phenomenon described by the sentence obtains.

Technical niceties aside, there are two putative scientific explanations on which this dissertation focuses. The first is the statistical mechanical account of the occurrence of phase transitions. This account applies to phenomena such as the melting of ice and the spontaneous magnetization of iron. The second putative explanation is a statistical mechanical account of why non-equilibrium systems, when left to themselves, irreversibly approach a state of equilibrium in a finite amount of time. This account applies to phenomena such as the spread of a puff of cigarette smoke throughout a room. A presumption of this dissertation, defended in Chapter Five, is that these accounts are genuinely explanatory.

Like many scientific explanations, the statistical mechanical accounts of phase transitions and irreversibility are idealized. An *idealized explanation* is an explanation in which at least one member of the explanans is an idealization. There are two characteristic functions of idealizations (see [8], pp. 177-178). First, every idealization replaces a description of a system with a description of an idealized version of that system. (The idealization itself is neither the replaced description nor the replacement description.) Second, every idealization simplifies: the replacement description is, in some sense, simpler than the original description.¹ For example, working with the replacement description might simplify the mathematical analysis of the system. The simplification function distinguishes idealizations from other kinds of replacement, such as the replacement of truth with fiction that occurs in some cases of deception. Note that this characterization of idealizations appeals only to the operational role of certain syntax; it appeals to neither the syntactic form of idealizations nor a semantic interpretation of such syntax.

Of all the idealizations that occur in the explanans for the accounts of phase transitions and irreversibility, an idealization common to both accounts – and the idealization most salient to this dissertation – is the limit in which a system’s number N of particles “tends to infinity”: $N \rightarrow \infty$.

¹The sense in which idealizations simplify is controversial. The discussion to follow relies upon intuitions about when one description is simpler than another. For one suggested analysis, see [21].

If idealizations are statements (declarative sentences), this syntax should be interpreted as being the limit in which a system's number of particles becomes infinitely large, and the application of this limit to an equation that describes a real system with only finitely many particles yields an equation that describes a (non-real) system with infinitely many particles.

Although the accounts of phase transitions and irreversibility are similar to many other scientific explanations in being idealized explanations, one thesis of this dissertation is that the accounts are dissimilar to other such explanations in being ineliminably idealized. (An idealized explanation of some phenomenon is *ineliminably idealized* if, in principle, the only way to explain the phenomenon is to appeal to an idealization in the explanans.)

The statistical mechanical accounts of phase transitions and irreversible behavior are ineliminably idealized.

Most idealized explanations are not ineliminably idealized: with most such explanations, it is possible to alter an idealized explanans for some phenomenon into an explanans that is less idealized, in such a way that the altered explanans also explains the original phenomenon. For instance, it is possible to explain why a certain launched projectile follows a parabolic path under the idealization that the projectile is a perfect sphere; and it is also possible to explain this fact without idealizing the shape of the projectile (although the latter explanation is more complicated). The interesting feature of the statistical mechanical accounts of phase transitions and irreversible behavior is that the idealizing limit in which a system's particle number $N \rightarrow \infty$ is ineliminable to those accounts.

(The arguments in favor of the ineliminability of the $N \rightarrow \infty$ limit in the accounts of phase transitions and irreversibility are based upon technical impossibility results from the current state of science. I do not mean to claim that scientific advances will never show the limit to be eliminable in accounting for these phenomena. Rather, I aim to show that, so far as we know at present, the limit is ineliminable to the accounts of phase transitions and irreversibility. Moreover, even if future progress provides explanations of these phenomena that do not require an appeal to the $N \rightarrow \infty$ limit, that progress would not undermine the thesis that the current accounts of phase transitions and irreversibility do require an appeal to this limit.)

A second thesis of the dissertation is that many extant philosophical accounts of scientific explanation fail to accommodate ineliminably idealized

explanations.

Many extant philosophical accounts of scientific explanation fail to accommodate ineliminably idealized explanations.

The overarching goal of this dissertation is to provide a philosophical account of how these explanations can be explanatory despite being ineliminably idealized. But first, it will be helpful to discern the conditions a philosophical account of explanation must satisfy in order to accommodate idealized explanations in general. For if a philosophical account of scientific explanation disallows the existence of idealized explanations, it also disallows the existence of ineliminably idealized explanations.

1 Explanation and Distorting Idealizations

Philosophical accounts of scientific explanation attempt to provide conditions under which an answer to a why-question is explanatory. That is, they attempt to provide conditions that an explanans must satisfy if it is to explain some explanandum. Historically, these accounts take one of three forms. Nomothetic accounts, best represented by Carl Hempel's deductive-nomological model, hold that to explain a phenomenon is to subsume it under appropriate laws. Causal accounts, best represented by Wesley Salmon, hold that to explain a phenomenon is to show how the causal-nomic structure of the world produces the phenomenon. Unification accounts, best represented by Philip Kitcher, hold that to explain a phenomenon is to unify it with other phenomena. The representative versions of each of these accounts, along with a fairly prevalent semantic interpretation of idealizations, entail that there are no idealized explanations.

1.1 Representative Accounts of Explanation

Carl Hempel's deductive-nomological (DN) model is a representative nomothetic account of explanation. According to Hempel, "the principal requirement for scientific explanation is . . . the inferential subsumption of the explanandum under comprehensive general principles" ([4], p. 445). Hempel provides four adequacy conditions for an explanation. First, the explanans must have empirical content; its elements must be confirmable. Second, the explanans must deductively entail the explanandum. Third, at least one member of the explanans must be a statement of a law of nature, and this

law-statement must be essential to the valid derivation of the explanandum. (The difficulty of adequately characterizing the notion of a law of nature, as well as the debate over whether there are such laws, are acknowledged but passed over as irrelevant to the present discussion: see [19] and [22], respectively.) Finally, every member of the explanans must be true.

Hempel takes this last condition, the factual correctness condition, to be “obvious” ([6], p. 322). He prefers an explanans that is true to an explanans that is highly confirmed, in order to prevent the possibility of an argument that satisfies his first three conditions being an explanation at some point in time but, given scientific advances that highly disconfirm some elements of the explanans, the same argument not being an explanation at some later time. Presumably his conception of explanations as sound arguments explains why he prefers true explanans to false ones.

Wesley Salmon offers a representative version of a causal account of explanation that differs from Hempel’s DN model in not requiring explanations to be arguments. According to Salmon, explanations are assemblages of facts that fit the phenomenon to be explained into its causal nexus – that is, etiological facts about the causal interactions and processes that produce the phenomenon as well as constitutive facts about the composition of the system in which the explanandum occurs. This assemblage might be an argument; and it might not be. There need not be any entailment relation between the sentences that describe the way in which the explanandum is embedded in its causal nexus and the explanandum itself; the only relations required are causal relations between the facts described by the explanans and explanandum. (The details of Salmon’s theory of causation are not important for the present discussion: see [20], pp. 253-257.)

Salmon’s account shares with Hempel’s model a requirement of factual correctness. No causal process or causal interaction is part of a causal explanation of an event unless it is part of the causal nexus for that event. Since non-existent causal processes and non-obtaining causal interactions are not part of the causal nexus for any event to be explained, they are not part of any causal explanation. David Lewis’s account of causal explanation, the major competitor to Salmon’s, also contains something like a requirement of factual correctness. Lewis is unwilling to decide whether accounts that violate this requirement are non-explanatory or just bad explanations: “it is unclear – and we needn’t make it clear – what to say about an unsatisfactory chunk of explanatory information, say one that is

incorrect or too small to suit us. We may call it a bad explanation, or no explanation at all” ([14], p. 218).

Like Hempel’s and Salmon’s accounts, Kitcher’s account of explanation as unification (see [10], [11]) denies that falsehoods are explanatory. According to Kitcher, explanations are valid deductive arguments, and whether an argument is an explanation depends upon whether it instantiates an explanatory argument pattern.

The explanatory argument patterns are the ones that best unify the set of sentences that belong to the belief corpus of scientific practice in the limit of its rational development ([11], p. 498). Kitcher’s account for when an argument pattern best unifies a belief corpus is complex. Although the details are interesting, the important point for this discussion is that his account prohibits falsehoods from being explanatory. An argument is a candidate for being among the best unifiers of a belief corpus only if all of its premises are members of that corpus ([11], p. 434). Since Kitcher identifies truths as those sentences that belong to the belief corpus of science in the limit of its rational development, no falsehood is a premise among the arguments that best unify that belief corpus.

1.2 Idealizations as Distortions

These accounts of scientific explanation are incompatible with the existence of idealized explanations, if the correct semantic interpretation of idealizations is that they are distortions. A *distortion* attributes a feature to a system that the system does not have. Many kinds of sentences, such as lies and ordinary mistakes, are distortions. For example, if Jack tells Jill that he cannot visit her because he has work to do over the weekend, but the reason he cannot visit her is that he has more important plans, then Jack’s statement is a distortion, because it attributes to Jack a reason that he does not have. Ordinary mistakes also qualify as distortions. For instance, Newtonian mechanics allows information transfer to occur at speeds well beyond the speed of light. This is a distortion with respect to how fast information can be transferred, because relativity theory sets the speed of light as the upper bound for such speeds. Again, early theories of combustion hypothesize that some substances contain phlogiston, but this is a distortion with respect to the composition of such substances because there is no phlogiston.

Idealizations often are treated as distortions. Ronald Laymon writes,

“The most natural attitude to take towards idealizations . . . is to assume that their use introduces distortion or bias into the . . . analysis” ([13], p. 354). According to Margaret Morrison, an idealization is “a characterization of a system or entity where its properties are deliberately distorted in a way that makes them incapable of accurately describing the physical world” ([16], p. 38 fn. 1). Likewise, Nancy Cartwright takes an idealization to be a mental rearrangement or replacement of inconvenient features or specific properties of a concrete object with factors “which are easier to think about, or with which it is easier to calculate” ([3], p. 187).

If idealizations are distortions, then an idealization replaces one description of a system with a simpler description that attributes to that system at least one feature that the system does not have. The resultant description, a distorted description, either attributes an incorrect magnitude to some property of the system of interest or qualitatively distorts some property of that system. Hence, if idealizations are distortions, an idealized description of a system is an incorrect description of that system. Note that this interpretation of idealizations does not entail that every distortion is an idealization; only those distortions that perform the appropriate functions qualify as idealizations.

The interpretation of idealizations as distortions provides a semantic meaning to mathematical syntax that satisfies the characteristic functions of idealizations. In order to illustrate this interpretation of idealizations, consider two systems in which idealizations are treated as distortions.

As a distortion, the **simple pendulum** is a pendulum subject to no friction or other non-gravitational forces; it has an extension of zero; its string is rigid and has no mass; and so on. The simple pendulum lacks features that real pendula have, and it has features that real pendula lack. For instance, real pendula are subject to a non-zero amount of friction and have a finite, non-zero extension. (See Figure 1.) Accordingly, a description of the simple pendulum is false of every real pendulum. Nonetheless, treating real pendula as simple pendula allows complicated (hard-to-solve) equations of motion for real pendula to be replaced with equations that are simpler (easier to solve).

Idealizations also are applied to gases. As a distortion, an **ideal gas** is a gas in which collisions between particles are elastic, in which the forces between its particles have no magnitude, etc. An ideal gas lacks features that real gases have, and it has features that real gases lack. For instance,

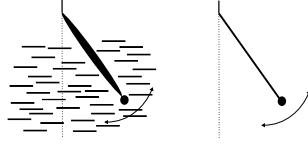


Figure 1: Real Pendulum vs. Simple Pendulum

collisions between particles of real gases are not elastic and intermolecular forces in real gases have finite, non-zero values. Accordingly, a description of an ideal gas is false of every real gas. Nonetheless, treating real gases as ideal gases allows complicated (hard-to-solve) equations of state for real gases to be replaced with equations that are simpler (easier to solve).

1.3 Accommodating Idealized Explanations

If idealizations are distortions then, according to the representative accounts of scientific explanation, there are no idealized explanations. Consider, first, Hempel's DN model. Hempel requires that every member of a putative explanans be true in order for it to be explanatory. If an explanation is idealized, then at least one member of its explanans is an idealization. If idealizations are distortions, this member is false. Hence, putative explanations that involve distorting idealizations are not explanatory according to Hempel's account.

Hempel attempts to avoid this conclusion by appealing to the notion of a proviso (see [5]). A *proviso* is a statement that the conditions under which an idealized law is true obtain. Hempel's strategy with provisos is two-fold. First, for any law that occurs as an element of a putative explanans, if the law is true only of idealized versions of real systems, it is to be replaced by a slightly different law. The consequent of this new law is to be the consequent of the original law. The antecedent is to be the antecedent of the original law conjoined with a proviso, which expresses the conditions under which the original law is true. Secondly, the putative explanans is to be supplemented with this proviso. This strategy prevents idealized laws from being false and allows one to infer that the idealized law is true when the conditions stated in the proviso obtain. The strategy is supposed to avoid violations of the factual correctness requirement.

Consider an example given by Hempel that illustrates the use of provisos. Suppose that β is a metal bar to which iron filings are clinging, and the explanandum is that, when β is broken into two shorter bars and

the shorter bars are suspended close to each other at some distance from the ground, the bars orient themselves so as to fall into a straight line. (See Figure 2.) From the theory of magnetism, it is possible to deduce the

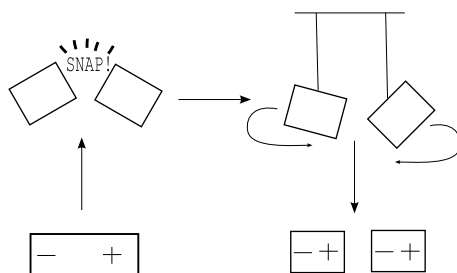


Figure 2: Orientation of Broken Magnet

law that if a magnet is broken into two bars, then both resultant bars are magnets whose poles attract or repel each other. This law can be used to deduce the explanandum. Yet, as Hempel notes, the law about magnets is false: if the bar β is broken at high temperatures, it becomes demagnetized. Hence, the explanation requires a proviso to the effect that the bar is not broken at high temperatures. This proviso is added to the explanans, and the law about magnets is replaced with a law stating that if a magnet is broken into two bars but not broken at high temperatures, then both resultant bars are magnets. Hempel takes this to allow for an explanation of the explanandum, when in fact β is not broken at high temperatures.

Hempel's appeal to provisos appears to allow idealized laws to be explanatory in some cases, because sometimes real systems satisfy the conditions set forth in the proviso. This appears to avoid the conclusion that putative explanations involving idealizations are not explanatory if idealizations are distortions. But the appearance is deceptive. When a real system satisfies the conditions set forth in the proviso, the idealizations that characterize the idealized law governing the system are not distortions; instead, they are true of the system, and the law that governs the system is not an *idealized* law. When a real system does not satisfy the conditions set forth in the proviso, the proviso is false. When a putative explanans contains a false proviso, it fails to be explanatory because the factual correctness requirement is violated. Hence, the conclusion previously drawn remains valid: any explanans that contains an idealization is not explanatory according to Hempel's account, if idealizations are distortions of real systems.

Like Hempel's account, Salmon's causal account does not permit ideal-

ized explanations if idealizations are distortions. Any idealized (distorted) version of a real system lacks some detail about the causal nexus of the real system and, moreover, contains details that are false of the causal nexus for the real system. Hence, every idealized version of a real system fails to exhibit the real system as it is embedded in its causal nexus. Thus, according to Salmon's account, idealized versions of real systems are not causally explanatory of those systems.

In an attempt to avoid this result, one might weaken Salmon's requirement on causal explanations. Instead of requiring an explanation to exhibit a phenomenon in its entire causal nexus, one might require only that an explanation exhibit a phenomenon in some relevant portion of its causal nexus. Paul Humphreys suggests this approach ([7], pp. 287-288). Humphreys distinguishes between complete and true causal explanations. A complete causal explanation of an event cites all and only the causes of the event; a true causal explanation cites causes that are causally relevant to the event. (Humphreys defines a causal factor X as causally relevant to an event e if a change in X invariably results in a change in e (p. 294).) While every complete causal explanation is a true causal explanation, some true causal explanations are partial rather than complete.

The distinction between complete and true explanations is designed to allow for causal explanations that cite some, but not all, of the causes of an event. Yet the distinction does not circumvent the conclusion that idealized versions of real systems are not causally explanatory of those systems if idealizations are distortions. As a distortion, an idealization of a causal interaction replaces the actual interaction with either a non-actual interaction or no interaction at all. In either case, the resultant idealized causal nexus is not a true-but-incomplete version of an actual causal nexus. Rather, it is an incorrect version of the actual causal nexus, because it is a distortion of that nexus. A similar point holds for idealized causal processes. Given the requirement that every member of an explanans must be factually correct, every idealized (distorted) version of an actual causal nexus is factually incorrect. Accordingly, even if causal explanations can be partial, idealized (distorted) versions of actual systems are not causally explanatory of those systems.

Finally, consider Kitcher's account of explanation as unification. Suppose that idealizations are distortions. Then there is an idealized explanation, according to Kitcher's account, just in case the belief corpus of

science in the limit of its rational development contains an argument that instantiates an explanatory argument pattern and this argument contains, as one of its premises, an idealization. Since, by hypothesis, idealizations are distortions, they are false. But an argument instantiates an explanatory argument pattern only if all of its premises belong to the belief corpus of scientific practice in the limit of its rational development. Since Kitcher takes the members of this corpus to be true, no idealization is a premise in any explanatory argument pattern. Hence, Kitcher's account disallows idealized explanations if idealizations are distortions.

One might attempt to avoid this conclusion by invoking Kitcher's distinction between correct explanations and acceptable ones. Correct explanations are genuinely explanatory while acceptable explanations are explanatory so far as the present scientific community can tell. Even though Kitcher's account disallows correct idealized explanations, it seems to permit acceptable idealized explanations, because the premises of an explanation that is acceptable relative to the present scientific community need only be endorsed as true at present – and this is compatible with those premises not being endorsed as true in the limit of the rational development of scientific practice.

Although the distinction between correct and acceptable explanations allows idealized explanations in principle (albeit acceptable rather than correct ones), for the most part it fails to allow them in practice. Any idealized explanation acceptable to the present scientific community is an argument in its belief corpus instantiating an acceptable explanatory argument pattern and containing an idealization as one of its premises. Since most idealizations that appear in putative explanations are known to be idealizations, most idealizations are not only false but also known to be false, if idealizations are distortions. And since an argument is in the belief corpus of a scientific community only if all of its premises are endorsed as true by that community, most idealizations are not part of the belief corpus of the present scientific community. Hence, for the most part, Kitcher's account disallows acceptable idealized explanations. (See Chapter Two for a modification of Kitcher's account that accommodates (correct) idealized explanations by allowing falsehoods to be explanatory.)

The representative accounts of scientific explanation entail that no explanations contain any distorting idealizations. This result does not depend upon the details that differentiate these accounts from one another. Rather,

the result follows because the accounts accept an adequacy condition on explanation, according to which any scientific account that involves appeal to a falsehood is not explanatory. An idealized account is a putative explanation that involves appeal to at least one idealization. If idealizations are distortions, such accounts involve appeal to at least one falsehood. Hence, given this adequacy condition on explanation, no such account is explanatory.

Any account of explanation that treats idealizations as distortions and accepts a factual correctness adequacy condition on explanation entails that there are no idealized explanations. This result is worth emphasizing, in part because it seems not to have been noticed. The result does not show, however, that there are no idealized explanations. It merely provides a constraint for any philosophical account of explanation that accommodates idealized explanation. *Any philosophical account of scientific explanation that accommodates idealized explanation must either abandon the interpretation of idealizations as distortions or allow some falsehoods to be explanatory.* This result provides a framework for discussing how the statistical mechanical accounts of phase transitions and irreversibility can be explanatory despite being (ineliminably) idealized.

2 Clarifications

The preceding discussion might raise the following worries: (1) that the project of this dissertation is hopeless because there are no idealized explanations; (2) that idealizations should be characterized according to syntactic rather than operational criteria; (3) that it is not possible to interpret idealizations as anything other than distortions. Before considering these objections, I shall enter one remark about the kind of explanandum that is the concern of this dissertation.

There are two types of phenomenon that one might want to explain. (For a similar distinction, see [1], [2].) A *type-1 phenomenon* is an individual event or class of events. A *type-2 phenomenon* is a multiply realized pattern among some class of events. Typical type-1 phenomena include the irreversible behavior of a particular gas, or the decrease of a particular pendulum's period upon a decrease in the distance between its pivot and center of mass, or the buckling of a particular strut under a sufficiently heavy load. Typical type-2 phenomena include the shared pattern of spacing between

rainbow fringes, or the general buckling of struts under sufficiently heavy loads.

A statistical mechanical account of phase transitions is an account of why a phase transition occurs in some particular system, rather than an account of why diverse systems share certain commonalities when undergoing phase transitions. And a statistical mechanical account of irreversible behavior is an account of why a particular system exhibits irreversible behavior, rather than an account of why diverse kinds of systems share certain commonalities when behaving irreversibly. Since the accounts of phase transitions and irreversibility concern type-1 phenomena, this dissertation focuses exclusively on explanations of this type of phenomena.

2.1 The Possibility of Idealized Explanation

One might object to the project undertaken here on the grounds that no scientific explanations are idealized. That is, one might be content to accept the conclusion that there are no idealized explanations, insisting that idealizations are distortions and only truths can be explanatory. There are two reasons to resist this attitude.

First, if no explanations are idealized, then our best sciences provide very few explanations. Most scientific accounts are idealized in some way or other. For instance, many journal articles in physics undertake to advance our scientific knowledge of some narrow domain of natural phenomena, in virtue of explaining those phenomena; and the results in these articles nearly always appeal to idealizations. Any philosophical account of scientific explanation that forbids idealized explanations thereby fails to make sense of much of scientific practice. It is counter-intuitive to hold that our best sciences furnish very few explanations. Nor is it satisfactory to treat these results as mere stop-gaps on the way to results with explanatory power, for this fails to explain why the results are taken to be a contribution to our shared scientific knowledge, and thereby fails to explain why they are taken to have epistemic, rather than merely heuristic, value. If science provides us with any explanations at all, there must be something about idealized descriptions that has explanatory power for physical systems. The philosophical task is to understand how scientific explanations can be explanatory despite being idealized, not to deny that explanations can be idealized.

Second, there are scientific accounts of phenomena that appear to be

explanatory despite being idealized. Consider the rough, qualitative proportionality between a pendulum's period and the distance from its pivot to center of mass. If this distance were to increase, the period would increase; and if the distance were to decrease, the period would decrease. (This proportionality holds for most typical pendula, but there are exceptions.) Galileo took advantage of this proportionality in designing the *pulsilogium*, the first instrument to objectively measure pulse speed (see [15], pp. 88 - 90). The *pulsilogium* is a pendulum attached to a movable peg that runs the length of a scaled board. The peg is connected to the string of the pendulum: moving the peg alters the length of the pendulum. The pendulum hangs perpendicular to the scale. (See Figure 3: the top picture shows the *pulsilogium* from above, with the hole indicating where the pendulum hangs down through the board and the star indicating the movable peg; the bottom picture shows the *pulsilogium* from the side.)

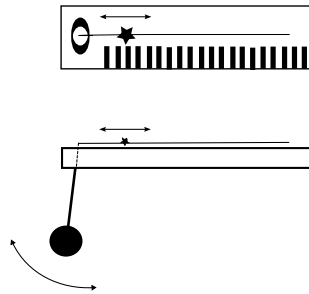


Figure 3: *Pulsilogium*

Given the rough proportionality between a pendulum's period and the distance between its pivot and center of mass, one can expect to alter the period of the pendulum by sliding the peg across the scale, thereby altering the length of the pendulum string. Different periods correspond to different string lengths, which correspond to different markings on the scaled board. Thus, one can measure a patient's pulse speed by sliding the movable peg until the pendulum oscillates in rhythm with the patient's pulse. This is a remarkable phenomenon, and one might very well wonder why the period of a pendulum is roughly proportional (in a qualitative sense) to the distance between its pivot and center of mass.

One account of this proportionality appeals to the simple pendulum. The simple pendulum is an idealized version of a real pendulum. A real pendulum has several characteristic properties. All of its components have a non-zero mass and extension, and imperfect rigidity. There is a point at which the pendulum connects to some support structure; this point is the

pivot about which the pendulum oscillates; and the support structure itself might oscillate. There might be friction at the pivot point, which tends to impede the oscillations of the pendulum; there might be a mechanism that drives the pendulum, tending to enhance its oscillations. The pendulum also swings in some medium, such as air, oil, or water; this medium further impedes the pendulum's oscillations.

The simple pendulum is quite unlike any real pendulum. The simple pendulum is an extensionless point at the end of a massless rigid string. The point contains the pendulum's entire mass. The point at which the simple pendulum connects to its support structure is fixed, and there is no friction as the pendulum pivots about this point. There is also no resistance from the surrounding medium; it is as if the simple pendulum oscillates in a vacuum. Nor is there any mechanism that enhances the simple pendulum's oscillations; only gravity affects its motion. The simple pendulum is an idealized version of real pendula, because it idealizes many properties of real pendula. A more complete list of properties of real pendula that the simple pendulum idealizes includes: the shape, rigidity, and mass distribution of real pendula, as well as variations therein; non-gravitational forces such as air resistance and friction between any parts of real pendulum systems; driving forces; variations in the temperature, mass or length of real pendula; variations in the magnitude of gravity with altitude and variations in the direction of gravity with location on an irregularly shaped earth; and movement of the support structure.

In virtue of these idealizations, the period of the simple pendulum is roughly proportional to the length of its string. The simple pendulum can be used to predict that if the distance between a pendulum's pivot and center of mass were to increase (decrease), then the period of the pendulum would increase (decrease). The simple pendulum has all the properties that are relevant to this proportionality between pendulum period and pendulum length. The equation that governs the motion of the simple pendulum is also law-like, since it is derivable from basic principles of Newtonian mechanics (along with appeal to several idealizations, of course); and the argument pattern used to derive the equation for the period of the simple pendulum is a common argument pattern within Newtonian mechanics, suggesting that the pattern is, in some sense, a unifying pattern. Reasons like these support the claim that the explanans that explains the rough, qualitative proportionality between a simple pendulum's period and the

length of its string also explains this proportionality as it exists in real pendula.

A second phenomenon amenable to idealized explanation is the familiar fact that a faint fog forms around the opening of carbonated drinks when they are first opened. This is most apparent in champagne, but it also appears in soda. (See Figure 4.) This phenomenon can be explained, in part, by appealing to the ideal gas law. The ideal gas law governs

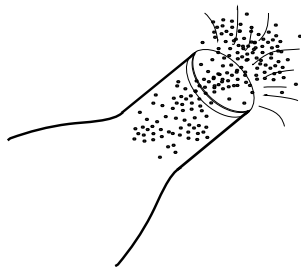


Figure 4: “Soda Fog”

ideal gases, gases that are idealized versions of real gases. Unlike real gases, an ideal gas is one in which every collision between its components is perfectly elastic and in which there are no attractive forces between any of the components. Using kinetic theory, it is possible to derive the well-known law that governs ideal gases: $pV = nRT$, where p is the absolute pressure of the gas, V is its volume, T is its absolute temperature, n is the number of moles of the gas, and R is the universal gas constant. (One mole of gas contains 6.02×10^{23} particles of the gas.)

The ideal gas law relates the pressure, volume, and temperature of a gas, and this relation is a central component for an explanation of the phenomenon of “soda fog”. An unopened container of carbonated beverage contains a carbonated fluid topped by a gas of water vapor and carbon dioxide (or whatever gas is used for carbonation). The pressure of this gas is greater than the atmospheric pressure, so that the pressure of the gas decreases towards the atmospheric pressure when the container is opened. Treating this gas as an ideal gas, the ideal gas law shows that, as the pressure of the gas decreases, its volume increases; that is, when the container is opened, the gas starts to expand beyond the opening of the container.

One consequence of the ideal gas law is that an ideal gas does positive work (in the technical sense of ‘work’) as it expands. The only source of this energy is the internal energy of the gas. Hence, according to the first law of thermodynamics, the internal energy of the gas must decrease as the

gas expands. This change in internal energy is proportional to a change in temperature, for ideal gases. Thus, as the gas expands beyond the opening of the container, it undergoes a temperature decrease. This temperature decrease results in condensation of some water vapor in the gas, and this condensed vapor is the “soda fog”.

Despite the idealizations that characterize an ideal gas, the ideal gas law can be used to predict that a faint fog forms around the opening of carbonated drinks when they are first opened. An ideal gas has all the properties that are relevant to the formation of this “soda fog”. The ideal gas law is law-like, since it is derivable from basic principles of kinetic theory (by appeal to several idealizations). The argument pattern used to derive the ideal gas law is a common argument pattern within kinetic theory, suggesting that the pattern is, in some sense, a unifying pattern. Reasons like these support the claim that the explanans for the explanation of the formation of “soda fog” in ideal gases also explains the formation of “soda fog” in real gases.

2.2 Syntactic Characterizations of Idealizations

Regardless of whether there are idealized explanations, one might object that idealizations have been improperly characterized. The characterization of idealizations adopted in this chapter only appeals to the operational role of idealizations – they replace one description of a system with a simpler description. There are other characterizations of idealizations that appeal to strictly syntactic marks of idealizations without referring to the operational role of that syntax. For instance, Leszek Nowak takes the distinctive mark of idealizations to be their mathematical form, and he takes this form to be one in which the magnitude of some property of a system is set equal to zero ([17]). So, according to Nowak, if x represents a property of a system, such as its mass, radius, charge, etc., an idealization of x is a sentence to the effect that $x = 0$. Karl Popper proffers a similar characterization, prompting him to call idealization the “zero method” ([18], p. 141 fn. 2).

Nowak and Popper’s approach to characterizing idealizations has the drawback that putative idealizations lacking the appropriate mathematical form are disqualified from being idealizations, despite the fact that they are mathematically equivalent to other syntax with the appropriate mathematical form. For example, a putative idealization is $g(h) = C$, where

$g(h)$ represents gravity as a function of height above the Earth's surface and C is a non-zero constant (typically 9.81 m/s^2). If idealizations are distortions, this says that the force of gravity on an object is the same at all heights. This is a putative idealization, because the gravitational force on an object varies slightly as a function of the height of that object above the Earth's surface. Since this idealization does not have the form $g(h) = 0$, it does not qualify as an idealization according to the Nowak-Popper criterion. However, the syntax $g(h) = C$ is mathematically equivalent to the syntax $dg/dh = 0$, where dg/dh represents the rate of change of the gravitational force on an object with respect to its height above the Earth; given the rules of the calculus, $g(h) = C$ and $dg/dh = 0$ are derivable from each other (differentiate the former with respect to h or integrate the latter with respect to h). This latter equation qualifies as an idealization according to the Nowak-Popper criterion, even though it is mathematically equivalent to the equation $g(h) = C$, which does not qualify as an idealization according to the same criterion.

This oddity of the Nowak-Popper criterion can be avoided by using non-syntactic, functional marks to characterize idealizations. The equations $g(h) = C$ and $dg/dh = 0$ perform the same operational role and thereby produce replacement descriptions that are equally simple (in whatever sense of simplicity is appropriate). Hence, according to the operational characterization of idealizations, $g(h) = C$ is an idealization just in case $dg/dh = 0$ is. Of course, this is not to say that the Nowak-Popper criterion cannot be modified to yield the result that $g(h) = C$ also counts as an idealization. For instance, one might modify the criterion so that the distinctive mathematical form of an idealization is $x = C$, where C is a constant that is sometimes zero.

It is likely that more extensive modifications than this will be required for an adequate syntactic characterization. For instance, the criterion will need to accommodate cases in which the magnitude of some quantity is made to approach a value of infinity. An example of this kind of case is the idealization according to which the Earth's radius R is made to be infinitely large. This can be construed as idealizing the earth to be flat rather than curved, since in the limit $R \rightarrow \infty$, the curvature of the Earth vanishes.

Supposing that such a modified syntactic criterion is available, there remains the task of distinguishing between syntax that meets the modified criterion and yields a simpler description than the original description,

from syntax that meets the modified criterion and does not yield a simpler description. For instance, the statement that the charge on a neutron is zero satisfies Nowak’s criterion for being an idealization, but this statement is used to obtain a *correct* description of neutrons rather than a *simpler* description. *Prima facie*, there is no way to mark such a distinction by appeal to syntactic criteria alone. Moreover, many different kinds of syntax can satisfy the same function, of replacing one description with a simpler description. So rather than develop a syntactic characterization of idealizations only to append to it a non-syntactic criterion, it is more straightforward to neglect the syntactic marks of idealizations altogether. For this reason, an operational, non-syntactic characterization is preferable to the Nowak-Popper characterization and is, accordingly, the characterization of idealizations adopted for the dissertation.

2.3 Alternative Interpretations of Idealizations

Regardless of whether there are idealized explanations, and even if idealizations should not be characterized with purely syntactic criteria, one might object that idealizations are distortions by necessity. That is, one might insist that part of what it is for something to be an idealization is for it to be a distortion (see [17], pp. 31ff; [9], p. 175). This insistence is resisted for the following reason.

Many equations of mathematical physics appear to represent relations among properties of unobservable entities. (My use of ‘represent’ and its cognates is intended to be neutral regarding the success or correctness of the representation.) For instance, Coulomb’s law appears to represent the force F between two charged particles as a relation between the charges of each particle, q_1 and q_2 , and the distance r between the particles (k is a constant):

$$F = k \frac{|q_1||q_2|}{r^2}$$

But force and charge and the particles themselves are, in some sense, unobservable. Maxwell’s equations of electromagnetism appear to represent relations between electric and magnetic fields. These fields are also, in some sense, unobservable. Most famously of all, Schrödinger’s equation in quantum mechanics $-\hbar^2 \psi / dx^2 + k^2 \psi = 0$ (for a particle moving in one dimension and subject to no force) – appears to represent the dynamical behavior of an unobservable wave function ψ . Within twentieth century

philosophy of science, there is a history of worrying about whether these pieces of mathematical syntax in fact represent the way the world is. For instance, typical instrumentalist interpretations of quantum mechanics treat Schrödinger’s equation as a mere calculation device that does not represent any “wave function”. And more global versions of instrumentalism treat all such equations – indeed, all claims about unobservable entities – as mere “inference tickets” that permit inferences from old observational claims to new observational predictions.

The philosophical community’s treatment of these worries about whether such syntax makes representational claims as legitimate, and the status of instrumentalism as a genuine rival to other interpretations of science, suggests the truth of the following *Instrumentalist Conditional*:

It is possible that a piece of mathematical syntax does not represent the way the world is if it is possible to interpret that syntax as not representing the way the world is.

The Instrumentalist Conditional, supplemented with a story telling us how claims about unobservable entities can be treated as “inference tickets”, entails the possibility that an instrumentalist interpretation of such claims is correct. There is no reason to suppose that the Instrumentalist Conditional applies only to equations of mathematical physics (or other claims) that appear to represent relations among properties of unobservable entities. The Instrumentalist Conditional might also apply to syntax that satisfies the functional criteria for being an idealization. If it does, and if it is possible to interpret that syntax as not representing the way the world is, then it is possible that such syntax does not represent the way the world is.

Given the absence of reasons to the contrary, it is reasonable to suppose that the Instrumentalist Conditional applies to syntax that satisfies the operational criteria for being an idealization. Moreover, it is possible to interpret this syntax as not representing the way the world is. For instance, it is possible to interpret this syntax as syntax that allows us to ignore certain features of the world (in the pursuit of certain descriptive, predictive, or explanatory aims, within certain margins of tolerable error), rather than as syntax that is a misrepresentative statement about those features. Chapter Four further elaborates and substantiates this claim. Hence, it is possible for syntax to satisfy the functional criteria for being

an idealization without representing the way the world is. A fortiori, it is possible that such syntax does not *incorrectly* represent the way the world is. Therefore, since it is possible for putative idealizations (pieces of syntax) to satisfy the functional roles of idealizations and for none of these putative idealizations to be distortions, it is not necessary that all (or even any) idealizations are distortions. And if some putative idealizations satisfy the functional criteria for being idealizations but happen not to be distortions, they should not be interpreted as distortions even though they are idealizations.²

3 Chapter Synopses

Having introduced the project undertaken in this dissertation, I shall provide a rough outline of the dissertation itself.

Chapter Two surveys philosophical accounts of idealized explanation that allow falsehoods to be explanatory. In particular, the chapter considers the accounts given by Ronald Laymon, Alexander Rueger and David Sharp, Philip Kitcher, and R.I.G. Hughes. The challenge for these accounts is to show how an explanans can be explanatory despite containing a falsehood, and thereby show how an idealized explanans can be explanatory despite containing a falsehood. Chapter Two presents the main details of these accounts. The chapter is largely expository, serving as a prelude to the critical discussion in the next chapter. (The chapter is also, to the best of my knowledge, the first discussion to set these accounts side by side.)

Chapter Three confronts the accounts of idealized explanation from Chapter Two with the statistical mechanical accounts of phase transitions and irreversibility. The chapter provides relevant details of these scientific accounts, as well as arguments that both accounts are ineliminably idealized – thereby substantiating one of the main theses of this dissertation. The chapter also argues that the philosophical accounts of idealized explanation surveyed in Chapter Two do not show how these scientific accounts are explanatory. Given the presumption that the scientific accounts are explanatory, it follows that the philosophical accounts from Chapter Two

²In giving this argument, I am not endorsing an instrumentalist account of laws. My goal is to give an instrumentalist-like account of idealizations that is compatible with a non-instrumentalist account of laws and other non-idealized elements that occur in various explanans – and to do this without undermining the explanatory power of accounts that invoke idealizations.

are inadequate. (This presumption is defended in Chapter Five.)

There are two reasons that would explain the inadequacy of the philosophical accounts from Chapter Two. Either those accounts do not correctly identify the conditions under which falsehoods can be explanatory, or no philosophical account that takes idealizations to be distortions can accommodate ineliminably idealized explanations. Chapter Three presents a paradox of ineliminable idealization, which shows that no philosophical account of idealized explanation that takes idealizations to be falsehoods can show, even in principle, how ineliminably idealized explanations are explanatory. That is, the paradox shows the existence of ineliminably idealized explanations to be incompatible with the treatment of idealizations as distortions. Given the presumption that the ineliminably idealized accounts of phase transitions and irreversible behavior are explanatory, the key conclusion of Chapter Three is that the interpretation of certain idealizations as distortions is mistaken.

Conservatively speaking, the paradox of ineliminable idealization only shows that some idealizations are not distortions, namely, the ineliminable ones that occur in the explanations of phase transitions and irreversible behavior. In the absence of an independent, principled reason to interpret some idealizations as distortions but not others, it is ad hoc to limit the conclusion of the paradox to the claim that only some idealizations are not distortions. A uniform interpretation of idealizations is preferable to a disjoint interpretation, if a uniform interpretation is possible. The aim of Chapter Four is to provide such an interpretation.

Rejecting the interpretation of idealizations as distortions, Chapter Four presents a semantic interpretation of idealizations according to which idealizations are abstractions. According to this interpretation, idealizations are not statements (declarative sentences) and, accordingly, do not incorrectly describe the way the world is; rather, they are “inference tickets” that allow us to ignore certain features of the world without thereby misrepresenting it. If idealizations are distortions, then an idealized explanans contains at least one falsehood; whereas if idealizations are abstractions, then an idealized explanans is incomplete (but does not necessarily contain a falsehood).

Chapter Four also partially develops a philosophical account of idealized explanation that is appropriate to an interpretation of idealizations as abstractions. The account respects the intuition that falsehoods are not

explanatory by treating idealized explanation as a kind of incomplete explanation. (However, I do not argue that falsehoods are not explanatory; I endorse that constraint for the sake of argument.) After presenting a partial account of idealized explanation, the chapter proceeds to show how, on the account proposed, the statistical mechanical accounts of phase transitions and irreversible behavior are explanatory. The content of Chapter Four satisfies the overarching goal of this dissertation.

Chapter Five defends the presumption that the statistical mechanical accounts of phase transitions and irreversible behavior are explanatory. Given the philosophical account of idealized explanation developed in Chapter Four, this presumption entails the striking conclusion that sometimes the correct description of a system requires the omission of details about the system. Chapter Five addresses two kinds of argument against the presumption that some explanations are ineliminably idealized. In the course of defending the presumption, I also make a case for considering phase transitions and irreversible behavior to be emergent properties of real systems.

Chapter Six provides an independent motivation for abandoning the interpretation of idealizations as distortions in favor of the interpretation of idealizations as abstractions. The motivation comes from cases in which there are two idealized hypotheses about a system, both of which are explanatory but cannot be accepted simultaneously as characterizations of the system, owing to their incompatibility with each other. Ordinarily inference to the best explanation provides a method for deciding which of the hypotheses should be taken as characterizing the system, thereby connecting explanation and ontology. Yet inference to the best explanation is not a cogent form of inference if idealizations are falsehoods. For, under such an interpretation, idealized hypotheses are false, but the conclusion of an inference to the best explanation is that the hypothesis with the most explanatory power is probably true.

Chapter Six critically discusses extant alternatives to inference to the best explanation as a method of deciding which idealized hypothesis, from a set of competing idealized hypotheses, should be taken to characterize the system of interest. There are two such accounts, one given by Lawrence Sklar, the other by Paul Teller. The chapter argues that neither of these accounts adequately characterizes the connection between idealized explanation and ontology. The account further argues that interpreting ideal-

idealizations as abstractions provides an adequate characterization of this connection, in accordance with the account of idealized explanation given in Chapter Four.

Chapter Seven is the last substantive chapter of the dissertation. The chapter provides a second independent motivation for abandoning the interpretation of idealizations as distortions, based upon a problem due to Michael Shaffer, which challenges Bayesian confirmation theorists to show how at least some idealized hypotheses have at least some degree of confirmation. Shaffer argues that, in order to accomplish this task, one must either abandon Bayesianism or develop a coherent proposal for how to assign prior probabilities to counterfactual conditionals. This chapter develops a Bayesian reply to Shaffer's challenge that avoids the issue of how to assign prior probabilities to counterfactuals. The reply treats idealized hypotheses as abstract descriptions and idealizations as abstractions. It allows Bayesians to assign non-zero degrees of confirmation to idealized hypotheses and to capture the intuition that less idealized hypotheses tend to be better confirmed than their more idealized counterparts.

References

- [1] Robert W. Batterman. Explanatory Instability. *Nous*, 26(3):325–348, 1992.
- [2] Robert W. Batterman. *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford University Press, New York, 2002.
- [3] Nancy Cartwright. *Nature's Capacities and their Measurement*. Clarendon Press, Oxford, 1989.
- [4] Carl G. Hempel. *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. Free Press, New York, 1965.
- [5] Carl G. Hempel. Provisos: A Problem Concerning the Inferential Function of Scientific Theories. pages 19–36. 1988.
- [6] Carl G. Hempel and Paul Oppenheim. The Logic of Explanation. pages 319–352. 1953.
- [7] Paul W. Humphreys. Scientific Explanation: The Causes, Some of the Causes, and Nothing But the Causes. In Kitcher and Salmon [12], pages 283–305.
- [8] Andreas Hüttemann. Idealizations in Physics. pages 177–192. 2002.
- [9] Martin R. Jones. Idealization and Abstraction: A Framework. pages 173–217. 2005.
- [10] Philip Kitcher. Explanatory Unification. *Philosophy of Science*, 48:507–531, 1981.
- [11] Philip Kitcher. Explanatory Unification and the Causal Structure of the World. In Kitcher and Salmon [12], pages 410–505.
- [12] Philip Kitcher and Wesley C. Salmon, editors. *Minnesota Studies in the Philosophy of Science, Volume XIII: Scientific Explanation*. University of Minnesota Press, Minneapolis, 1989.
- [13] Ronald Laymon. Experiment and the Legitimacy of Idealization. *Philosophical Studies*, 77:353–375, 1995.
- [14] David Lewis. Causal Explanation. pages 214–240. 1986b.
- [15] Michael R. Matthews. *Time for Science Education: How Teaching the History and Philosophy of Pendulum Motion can Contribute to Science Literacy*. Kluwer Academic / Plenum Publishers, New York, 2000.
- [16] Margaret Morrison. Models as Autonomous Agents. pages 38–65. 1999.

- [17] Leszek Nowak. *The Structure of Idealization: Towards a Systematic Interpretation of the Marxian Idea of Science*. Reidel, Dordrecht, 1980.
- [18] Karl R. Popper. *The Poverty of Historicism*. Routledge and Kegan Paul, London, 1957.
- [19] Stathis Psillos. *Causation and Explanation*. McGill-Queen's University Press, Ithaca, 2002.
- [20] Wesley C. Salmon. *Causality and Explanation*. Oxford University Press, New York, 1998.
- [21] Elliot Sober. *Simplicity*. Oxford University Press, Oxford, 1975.
- [22] Bas C. van Fraassen. *Laws and Symmetry*. Oxford University Press, New York, 1989.