What is Right with the Miracle Argument:
Establishing a Taxonomy of Natural Kinds

Martin Carrier*

IT CERTAINLY strikes us as one of the most remarkable types of scientific achievement when apparently disparate phenomena are unified theoretically. What appeared to be disparate to the untutored eye turns out to arise from the same underlying mechanism and thus to be identical in kind. When a door slams because the windows are open and it is windy outside, this happens due to the same cause and according to the same mechanism that makes a plane lift off the ground. The *prima facie* conclusion is that science succeeds in going beyond the specious distinctions of the senses. It teaches us what things are truly alike.

My aim in this paper is to examine the viability of this popular view. And the result will be that the view is basically correct. More precisely, I will try to show, first, that in some distinguished cases science arguably manages to induce the right classification or taxonomy among the phenomena, and that, second, this is the only access to reality that science is justifiably able to gain. Accordingly, what I am aiming to do is to support a particular and comparatively weak form of scientific realism.

Scientific realism contends that claims about certain unobservable 'items' which emerge from the theoretical or experimental activity of scientists are literally true; these claims faithfully refer to what is, as it were, going on behind the scenes. There is some quarrel, however, about what these 'items' are legitimately supposed to be. The leading brand of this doctrine is *theory-realism*. According to this position, the successful theories of mature science are approximately true. That is, these theories correctly portray the not-directly-observable processes and mechanisms that make the phenomena occur the way they do.

A more attenuated version of scientific realism is *entity-realism*. On the one hand, entity-realism is an immediate consequence of theory-realism. The truth of a theory implies the existence of its theoretical entities. On the other hand, there is also a more autonomous type of entity-realism which is advanced on the basis of experiment-centered arguments. In this version, entity-realism says that the capacity to manipulate certain unobservable entities, and, in particular, to manipulate them in order to experiment on something else, gives

*Center for Philosophy of Science, University of Konstanz, 78434 Konstanz, Germany.

Received 29 October 1992; in revised form 10 February 1993.
strong evidence for the real existence of these entities. Entity-realism is thus non-committal as to the (near) truth of the accepted scientific account of these entities. It is no theory’s entity but the experimenter’s entity to which actual existence is attributed. Belief in these entities does not imply belief in any one of the theories involved.

An even more attenuated variant of realism, and the one I set out to support, is a realism of kinds. The claim is that science, at least on some rare and distinguished occasions, manages truthfully to forge links among the phenomena. Science sometimes succeeds in collecting phenomena into equivalence classes that reflect truly existing similarity relations among these phenomena. Whereas entity-realism implies a kind-realism, the reverse does not hold. It will be seen that commitment to the experimentally established entities entails commitment to the kind-structures they introduce. By contrast, the assumption that science is sometimes able to unveil true relations of similarity does not imply that the mechanisms employed to establish these relations and the entities they invoke reflect anything in nature.

The argument comes in three steps. The first one is introductory. I sketch the notion of ‘natural kinds’ and show, using some examples from the history of science, that the structure of natural kinds is theory-dependent and indeed changes in the course of scientific progress. This finding implies that a naive form of a realism of kinds is untenable empirically. In the second step I consider the experimental route to natural kinds. The *prima facie* advantage of this alternative is that it seems to be exempt from the uncertainties and vicissitudes of theoretical reasoning. It turns out, however, that this position is likewise beset with a historical counter-example. Next comes, finally, the constructive step. I take up some earlier results of mine about the actual impact of the so-called ‘Miracle Argument’ in favor of scientific realism. The argument says that scientific realism is a necessary precondition for any explanation of scientific progress; it is the only choice for divesting progress of its miraculous aspects. I will argue that the Miracle Argument is insufficient for achieving its proper objective, namely, theory-realism, and that its real accomplishment consists in showing that there are instances in which science provides us with a veridical portrait of relations of similarity as they prevail in nature.

1. Changing Patterns of Natural Kinds

Every theory determines a classification among the objects or events it deals with; it induces a taxonomy into its universe of discourse. The laws of the theory are framed using certain descriptive predicates, and the relevant objects obey the laws in virtue of satisfying these predicates. A law thus serves to bind several objects together. It establishes a link of similarity among them by
regarding them as instances of the same law. A so-created equivalence class of like objects is called a natural kind (cf. Fodor 1974, 101–102). Take, for example, the law: All electrons possess the elementary charge $e$. This law generates the natural kind ‘electron’ by quantifying over electrons. Electrons constitute a natural kind because there is a law that applies to them in virtue of their property of being electrons. Different laws may well pick the same natural kind. All further laws about electrons (such as: All electrons have a spin-value of $\frac{1}{2}$) obviously specify the same kind, namely, the class of electrons. This implies immediately that a theoretical change need not involve a change in the concomitant kind-structure. Different bodies of laws may select the same objects as being alike.¹

The concept of natural kinds, as just circumscribed, is entirely non-committal as to the issue of scientific realism. Natural kinds are created by a corpus of laws irrespective of whether these laws are interpreted merely as useful but fictitious unifiers or are thought to refer to a theory-independent reality. Accordingly, natural kinds need not be truly natural, and the attribute ‘natural’ is, strictly speaking, a misnomer. In order to avoid the realist overtones the expression ‘scientific kinds’ has been proposed, but this proposal seems not to have gained acceptance. Thus I stick to the usual term, taking its non-committal sense to be understood. The reality of natural kinds constitutes the topic of the present investigation.

The first response to this problem presumably is that we are surely entitled to interpret the taxonomy specified by present-day science realistically. After all, its laws have undergone a large number of severe tests; they are well-confirmed and thus deserve our confidence. And this confidence quite naturally extends to the taxonomy induced by these laws. Most of us would thus be inclined to reason as follows. Since present-day knowledge includes laws about electrons, electrons form a natural kind. For this reason, electrons are in reality of the same kind. Let’s examine the tenability of this assessment.

A condition familiar from the discussion of scientific realism in general is the following retention requirement. For a theory component to be interpreted realistically, it is necessary that it be retained across scientific change. In particular, if earlier successful theories in the mature sciences are approximately true, then the later even more successful theories in the same discipline

¹Note that ‘natural kind’ is used here as a technical term which does not wholly coincide with everyday usage. Biological species, for instance, though they are natural kinds in a rough and ready sense, hardly qualify as natural kinds in the present understanding. Kinds are derived from laws, and it is a matter of dispute in current philosophy of biology whether there are any specifically biological laws. In any event, a regularity of the sort ‘All ravens are black’ certainly does not count as a law of nature and thus does not give rise to the natural kind ‘raven’.
will preserve the earlier laws as limiting cases (cf. Putnam 1978, 20–24; Laudan 1981, 234–240). Applied to natural kinds this means that the kind-structure of earlier well-confirmed theories should be reproduced 'in the limit' by the kind-structure of their respective successor theories.

The intuition behind the retention condition is as follows. Once-successful but now-rejected theories in the mature sciences were in their lifetimes well-confirmed according to our present methodological criteria. If such theories turn out to be wrong on all counts, nothing prevents us from the meta-inductive inference that our present most cherished accounts are likewise doomed to complete failure and that nothing will remain from them in the end. If scientific realism is supposed to be a viable position, this meta-inductive move has to be blocked; and it can only be blocked by assuming that these earlier accounts indeed got something right (cf. Putnam 1978, 25). These correct aspects should be retained by their respective successor theories. Realism about a specific item, be it theory, entity or kind, implies a retentionist claim with respect to this item.

This three-fold retention condition constitutes an epistemic requirement. Scientific realism claims not only that there is something real out there, but also that we gain access to reality through science. And we can only maintain that we have managed to lock on to reality if the features accorded this special status will not be discounted through scientific progress. If science is supposed to get hold of reality, it is necessary that the corresponding insights are here to stay. Accordingly, the retention condition does not amount to the assertion that what is not retained can by no means be real. Rather, the claim is that we have no science-based justification for attributing reality to abandoned features.

Scientific realism about theories thus requires preservation of theoretical laws, scientific realism about entities demands preservation of entities, and, finally, scientific realism about kinds necessitates preservation of kinds. On the other hand, it is certainly not requisite that the respective items be carried over unchanged from theory to theory. A successor theory may well have to say more about the corresponding aspect so that revisions cannot be excluded outright. The point rather is that the successor account must not completely overturn the relevant aspects or its predecessor. This is expressed by merely demanding preservation of the relevant item 'in the limit'. Regarding theory-realism this means that the theoretical laws of the earlier account may be reproduced by a 'corrective reduction'. That is, the derivability of the predecessor laws from the successor theory may be restricted to counterfactual initial and boundary conditions (cf. Carrier and Mittelstrass 1991, 42–50). As regards natural kinds, this means that earlier and later taxonomies have to be compatible in the sense that the later taxonomy can be construed as a more fine-grained version of the earlier one. This demands in turn that an earlier
kind may comprise several later kinds, to be sure, but that there is no partial overlap or cross-over among the kinds involved.²

Now we are in a position to take up the initial question: Is it justified unrestrictedly to interpret the taxonomy of present-day science realistically as it stands? It follows from the discussion that realism implies retention. As a consequence, non-retention favors a non-realist view. Regarding the realist interpretation of kinds, this leads to following empirical requirement. In order that a historical sequence of taxonomic structures be interpreted realistically, the respective kinds must stand in a relation of total inclusion. If, by contrast, cross-over relations among kinds typically occur in the course of theory changes, this tends to discredit a realist view about kinds. I will now argue that the inclusion condition is not satisfied typically.

A claim of this sort is best backed by considering some examples. To begin with, let's cast a brief glance at the transition from classical mechanics to general relativity theory. The basic taxonomic distinction involved in all dynamical theories is the one between force-free and force-induced motion. According to the classical law of inertia, force-free motion is represented by uniform rectilinear motion. Conversely, the motion of a body in physical fields such as the gravitational or electromagnetic field counts as accelerated and thus as force-induced. In contradistinction, it is constitutive of general relativity that it takes gravitation to be part of space–time structure. This means that gravitation-induced motion is construed as inertial motion. More precisely, the motion of a 'test particle' (i.e. a small, non-rotating particle with negligible mass) in the exclusive presence of a gravitational field is considered force-free. This interpretation does not extend to other forces such as the electromagnetic one, however. General relativity maintains that a charged particle in an electromagnetic field moves non-inertially. Accordingly, gravitation-induced motion is now grouped together with classical inertial motion and separated from motion in an electromagnetic field. There is thus non-inclusion, but rather partial overlap, between the respective kind-structures.

Another case in point is the taxonomy induced by Newton's corpuscular optical theory and the modern account. 'Newton's rings', i.e. the colors exhibited by transparent thin plates, are attributed by the current account to the occurrence of interference between the light waves reflected at the upper and lower surfaces of the plate. Newton, by contrast, explained the effect by appeal to his particle model. When light particles enter the plate they produce a longitudinal shock wave in the plate material. This shock wave moves faster through the plate than the light particles and puts the opposite side of the plate in oscillation. If this oscillatory motion happens to be counterdirected to the light particles' motion upon their arrival, the surface throws the particles back;

²This requirement is elaborated in Buchwald 1992, 40–41.
otherwise it lets them pass through. The critical factor is thus the phase of the surface oscillation when the surface is hit by the light particles. And the salient point of Newton's construction is that this phase depends on the passage time of the light through the plate and thus on the distance the light traverses in the plate. Since, furthermore, different colors are refracted differently by the same material, light particles representing different colors have to cover different distances until they reach the opposite surface of the plate. They are thus reflected differently. Consequently, if white light falls at a given angle of incidence on a transparent plate, only one color is reflected and the remainder is let through (cf. Newton 1730, 206–214, 280, 370–371).

The point is that Newton followed the very same approach in order to accommodate the permanent colors of bodies. He assumed the particles of bodies to be transparent and of variable size. Then he identified a body's color with the light reflected by its particles according to the process described. The reflected color is thus dependent on the size of the respective particles (cf. Newton 1730, 248–256).

Accordingly, the theory forges a link between Newton's rings and the colors of bodies. The two phenomena arise from the same mechanism and thus form a natural kind.

In contradistinction, these two phenomena are interpreted today as growing out of quite different mechanisms. The colors of bodies come about through a process of partial absorption and re-emission of the incident light by the atoms or molecules involved. Newton's rings, on the other hand, are the result of interference. So, what once belonged into the same natural kind is now thought to be of entirely distinct nature. A fundamental taxonomic breach splits the former kind into heterogeneous components. This constitutes another instance of a non-inclusive taxonomic development.

Just one more example. The caloric theory of heat assumed that heat is constituted by a material substance, namely, the matter of heat or caloric. Liquefaction and vaporization were interpreted as combinations with caloric. Liquids and gases are caloric compounds. This implies that vapors are in this respect identical to other compounds. This means that, say, oxygen gas belongs into the same natural kind as, say, a solid metallic oxide. Needless to say, this linkage has been cut by modern theory.

In order not to multiply examples gratuitously let me stop here. I take it to be the upshot of the discussion that the history of science is rife with instances of non-inclusive kind splitting or kind cross-over. This implies that the
What is Right with the Miracle Argument

retentionist claim with respect to kinds is untenable historically, with the consequence that a kind-realism fails—at least in the sweeping and unqualified version given in the first part of this section.

2. The Experimental Route to Kinds
   I: Manipulation Here Creates Kind-Structure Here

The theory-dependence of natural kinds along with the occurrence of profound theoretical changes across history appears to vitiate a theory-based realist interpretation of kinds. An alternative option might be to dispense with theories altogether and to ground the kind-structure in experimental distinctions and experimenters' abilities. It was proposed by Jed Buchwald that it is experimental set-ups and experimental devices that sort effects into kinds. Buchwald connects the evolving kind-structure of optical phenomena with the available laboratory equipment. The doubly refracting crystal established the distinction between polarized and unpolarized light. Additional invocation of the Fresnel rhomb produced the sub-kinds linearly and circularly (or elliptically) polarized light. Buchwald's claim is that 'the apparatus proper often constitutes an embodiment of the relevant kind-structure'.

The idea underlying the experiment-centered approach to kinds seems to be as follows. The ability to manipulate an object or effect gives evidence of relations of analogy and disanalogy. If we intervene in an apparently homogeneous phenomenon and manage to elicit heterogeneous, qualitatively distinct responses, then the components constitute different kinds. Briefly, what behaves differently in an experiment is different in kind. We still don't know what it is that manifests itself differently, but we know it's different.

In the present context the question is whether the experiment-based kind-structure is trustworthy. It is to be admitted at once that the occurrence of empirical differences is indeed necessary for sorting effects into distinct categories. Things that behave alike under all circumstances will hardly qualify as different in kind. The problem rather lies with the sufficiency part of the claim. Things that behave differently in an experiment may yet be alike.

This peculiarity has three possible sources. First, theoretical unification. Different sorts of electromagnetic radiation—such as infrared, visible light, and ultraviolet—show different behavior. But these sorts of radiation are still not sufficiently different to be counted as distinct natural kinds. They rather constitute different instantiations of the same kind, namely, electromagnetic radiation. The reason is that they merely differ in the values of the pertinent

Buchwald 1992, 57; for the entire argument cf. ibid. 46-58.
parameters, namely, frequency and wavelength, whereas they all obey one and the same corpus of laws.

Second, experiment-induced distinctions may be spurious and insignificant because the experiments may later be discovered to be spoiled by unnoticed factors and uncontrolled side-effects. Consider eighteenth-century affinity theory as an example. The theory was created by Newton and dominated important parts of the chemistry of the period. Affinities were conceived as attractive short-range forces between particles, and chemical reaction and chemical bonding were attributed to their influence. Affinity forces were thought to be substance-specific, i.e. they should be constant for a given pair of substances, and to possess a point of saturation. Saturation means that a particle of a substance $A$ can attract only a limited number of particles of a substance $B$. These assumptions suggest the use of substitution reactions for ordering affinity strengths empirically. When a substance $C$ replaces a substance $A$ in a compound $AB$, i.e. if the reaction $AB + C \rightarrow AC + B$ occurs, then the affinity from $A$ to $C$ is stronger than the one from $A$ to $B$. If, in addition, $D$ is able to substitute $C$ in the compound $AC$, then this testifies to the fact that $D$ is attracted more strongly by $A$ than $C$ is (cf. Newton 1730, 376–383). Substitution reactions were employed to establish so-called ‘affinity series’ experimentally. At the top the substance in question was placed, and then followed its reaction partners in the order of decreasing affinity strength. The results sketched thus give rise to the following midget affinity series. $A$: $D, C, B$.

The affinities operative in the presence of three substances were called ‘simple affinities’. In addition to them, so-called ‘double affinities’ were assumed. Double affinities were supposed to govern reciprocal substitutions of the general form: $AB + CD \rightarrow AC + BD$. They were considered completely different from simple affinities on the ground that no consistent set of relative affinity strengths could be found that was suitable to accommodate both types of substitutions.

What we have here is a fundamental distinction between two different kinds of chemical reactions. Simple substitutions are determined by the action of simple affinities, reciprocal substitutions are guided by double affinities. The crucial point in the present context is that this distinction between natural kinds was introduced on purely experimental grounds. Reactions of both types appeared to instantiate different behavior; they could not be accommodated by an overall scheme of affinity strengths. That is, one and the same compound was found to behave qualitatively differently according to whether a simple or a double substitution was performed.

This experiment-based distinction later collapsed completely. As Claude Berthollet demonstrated around 1800, the concept of affinity was defective in that the influences of temperature and of the quantities of the relevant
substances had been left out of consideration. Due to uncontrolled fluctuations of these factors the experiments on affinity series were unreliable; they were not indicative of any significant trait. Along with the constant affinity forces, the distinction in kind between simple and double affinities was abandoned. Berthollet endeavored to explain both types by one unified scheme.4

The moral to be extracted from this episode is that experimentally introduced distinctions can later be recognized as faulty and insignificant. In the case sketched, a kind-distinction was revoked as a result of theoretical progress. What behaved differently experimentally turned out by theoretical reasoning to be of the same kind in the end.

Third, it is sometimes impossible to tell from the facts alone whether various experimental outcomes are really heterogeneous or whether they merely represent different instantiations of the same kind. After all, if an experiment yields gradually varying magnitudes of the same quantity, it does not give rise to a kind-distinction at all. In order that the experimental approach be sufficient to establish natural kinds, we must be able to recognize empirically whether the ensuing experimental results are qualitatively different or whether they merely represent changing magnitudes of one and the same quantity. An example from roughly the same period in the history of chemistry makes it clear, however, that this task is sometimes hard to accomplish.

Consider the distinction between compounds and mixtures that Joseph Louis Proust introduced around 1800. Proust tried to establish empirically the 'law of constant proportions' which involved precisely this distinction. The evidential basis was that some reactions ended up with compounds of definite proportions and some did not. For Proust the former constituted actual compounds and the latter mere mixtures or mixtures of compounds. Proust's proposal was criticized by Berthollet. Berthollet's theory implied that variable proportions are the rule and that constant proportions arise from the influence of particular additional factors. In general, compounds are characterized by variable proportions (cf. Carrier 1986, 374–377).

Both rivaling approaches thus interpreted the same empirical findings differently. In particular, they both led to a different kind-structure. Proust's difference in kind between compounds and mixtures was denied by Berthollet and reinterpreted as gradual change due to gradually varying influences. Berthollet held that the occurrence of definite proportions under some circumstances arose from the presence of ephemeral factors. In fact, there is only one single kind that embraces Proustian compounds and mixtures alike.

The point is that experience alone does not show who is right; it does not show whether the observed differences indicate differences in kind or only differences in magnitude. It is true that Proust finally won the day. But this

victory was crucially due to the acceptance of John Dalton's atomic theory. Only in this framework could a theoretical distinction between compounds and mixtures be drawn. The experimental evidence provided too shaky a basis for a non-arbitrary distinction between them.

These examples give sufficient evidence for the conclusion that experiments are insufficient to distinguish among natural kinds. First such experiment-based distinctions may be taken back subsequently as a result of progress in theoretical unification. Second, their introduction may have been flawed right from the beginning due to the unrecognized influence of additional factors. Third, the distinction among natural kinds exclusively on the basis of observable differences may be entirely arbitrary due to the inability always to distinguish unambiguously between differences in quality and changes in quantity.

3. The Experimental Route to Kinds
   II: Manipulation Here Creates Kind-Structure There

   Whereas the version of the experimental approach discussed above endeavored to establish a kind-structure with respect to the phenomenon or entity experimented upon, another variant of the same approach leads to the introduction of a kind-structure with respect to the entity used for experimenting on other phenomena. At least this is the by-product of an enterprise whose explicit objective is to support a realist interpretation of entities. In this section I briefly sketch the entity-realism advanced by Nancy Cartwright and Ian Hacking, elaborate its implications as to a realist interpretation of kinds, and, finally, propose a counter-example.

   As just indicated, Cartwright and Hacking attempt to establish realism about entities by focusing on our experimental abilities. That is, their entity-realism is not a consequence of theory-realism but is supposed to stand on its own feet. Theory-realism is inadequate, they argue, since there is no unanimously accepted body of theory about anything that is part of present-day research. Different scientists may hold different and sometimes incompatible models about the same entity: and quite legitimately so, for these different models suit different descriptive purposes best. Every discipline embraces a multitude of distinct accounts of the same theoretical entity. As a result, entity-realism cannot be based on the assumed truth of the relevant theories.

   By contrast, entities are rightly attributed reality when we succeed in using them for investigating something else. The successful manipulation of an entity for the sake of intervening in other processes of a more hypothetical nature is the best possible evidence for the existence of this entity. Fictitious entities have no causal powers. When we know how to use an (initially theoretical)
entity so as to create new phenomena, then this entity is rightly thought to be real (cf. Cartwright 1983, 87–99; Hacking 1983, 262–265).

The rationale for the existence of theoretical entities thus is that they can be employed to exercise causal influence. Relying on them we can deliberately produce effects. An entity-realism of this sort implies a kind-realism if the entities accepted are the ones that science specifies. For science frequently traces different effects back to the action of one and the same entity. These effects are thus equal in kind in that they are brought about by similar mechanisms or in an otherwise similar way. In this vein, Cartwright holds electrons to be instrumental in the Millikan-experiment and likewise to be the cause of cloud-chamber tracks (cf. Cartwright 1983, 92–93, 99). Causal judgments of this sort generate a link between the two phenomena and thus induce a natural-kind structure.5

If the argument passed muster, it would indeed be suited to justify a realist view about kinds. In order to examine if it does, I resort to the touchstone of retention: for any item to be interpreted realistically, it is necessary that it is retained across scientific change (see Section 1). As a matter of fact, Cartwright herself makes a claim of roughly this sort. An entity that is experimentally warranted in the way sketched is ‘seldom discarded in the progress of science’ (Cartwright 1983, 98). I now argue that phlogiston passed the experimental reality test. Nevertheless, it was later abandoned.

The phlogiston theory stands in the tradition of the ‘chemistry of principles’. This tradition assumed certain abstract principles or elements that were considered to be bearers of general properties such as hardness, combustibility and volatility. Empirical substances gain their properties by incorporating the property-bearing principles; the principles thus explain the properties of substances found in the laboratory. The theory did not attach one principle to each property; it rather introduced just a few such principles. The challenge was to explain the multitude of empirical properties by recourse to combinations of these few general principles.

In the present context it is the principle of combustibility termed ‘phlogiston’ by Georg Ernst Stahl, the creator of the phlogiston theory, that deserves particular interest. In the pre-Stahlian tradition, the principle of combustibility was supposed to comprise a number of related sub-principles. It was Stahl’s chief objective to dispose of the multitude of distinct but related entities. He tried to demonstrate that there is only one such principle, namely, phlogiston. In particular, all combustion processes and all calcinations (i.e. oxidations of

5This feature becomes even more explicit in Cartwright’s later work on capacities. Causal claims are made about properties, and an individual object is causally effective because it possesses this property. Aspirins relieve headache because of being aspirins; cf. Cartwright 1989, 141. This specification obviously strongly resembles the definition of natural kinds given above — with the sole difference that she does not refer to laws but to ‘causal capacities’. 
metals in modern terms) are to be interpreted in the same fashion as the release of phlogiston.

This novel claim was backed by the following experiment. If the unified account of combustion and calcination is true, it should be possible to produce a metal from its calx (i.e. the corresponding metallic oxide) by supplying phlogiston that originates from non-metallic combustible substances. Stahl succeeded in verifying this prediction by producing metallic lead out of lead calx (PbO) by heating it with glowing charcoal. Obviously, the calx has accepted the phlogiston released by the charcoal and accordingly turned into the metal.

After this initial empirical conformation of his unified approach, Stahl moved on to bring to bear the causal capacities of his newly established entity. He employed phlogiston so as to bring about one more formerly unknown phenomenon, namely, the synthesis of sulfur from sulfuric acid through the action of phlogiston. There were empirical indications for the view that sulfuric acid was in fact dephlogisticated sulfur, and Stahl managed to confirm this view by his then celebrated sulfur synthesis, published in 1697.

The first step of the experiment consisted in combining sulfuric acid with potash (K$_2$CO$_3$) in order to reduce its volatility. The result is ‘fixed alkali’ (K$_2$SO$_4$), interpreted as a compound of sulfuric acid and potash. When this compound is brought to a glow with charcoal — whose combustibility indicates that it contains a large quantum of phlogiston — ‘liver of sulfur’ is formed (which is actually a mixture of various potassium-sulfur compounds such as K$_2$S$_3$ or K$_2$S$_2$O$_3$). Finally, Stahl precipitated sulfur from the reaction product and thus confirmed that it really contained sulfur (cf. Kopp 1873, 46–47; Partington 1962, 673–674).

What happened in this experiment is roughly and qualitatively represented by the following scheme.

\[
\text{K}_2\text{SO}_4 + \text{C} \rightarrow \text{K}_2\text{S}_3, \text{K}_2\text{S}_2\text{O}_3 + \text{CO}
\]

The experiment was supposed to embody a phlogiston transfer. The phlogiston released from the charcoal was taken up by the fixed alkali and transformed the sulfuric acid contained in it into sulfur. According to present lights, the core reaction that occurred is \(\text{SO}_2 + 2\text{C} \rightarrow \text{S} + 2\text{CO}\).

The point of this experiment is the following. What Stahl did was to create a new phenomenon relying on the previously established causal properties of the entity in question. He manipulated phlogiston in order to experiment on another, more hypothetical phenomenon, namely, the composition of sulfur. Stahl employed phlogiston so as to intervene in and actively change other
processes. I conclude that phlogiston qualifies as real according to the Cartwright–Hacking criterion. Nonetheless, it was later given up.

The upshot is that this reality-criterion does not single out the right entities as real. As a consequence, the kind-structure induced by such experimentally supported entities cannot be trusted in either. These kind-structures may be real, to be sure, but the Cartwright–Hacking criterion does not license the inference to their reality. We are not justified in attributing reality to kind-structures supported in the experimental fashion.

It is true that Stahl’s sulfur synthesis does not strictly entail this conclusion. After all, Cartwright only claims that entities backed in this way are ‘seldom discarded’ (see above), and one single counter-instance does not trespass this limit. On the other hand, the history of science shows that entities are less frequently abandoned than theories. This stability speaks in favor of entity realism, to be sure, but at the same time it reduces the testability of entity realism by reducing the number of possible counter-examples. And, if among the fairly limited class of discounted entities there is (at least) one that satisfied the Cartwright–Hacking criterion, then this one actual counter-instance is sufficient to justify doubts as to the viability of the experimental approach to entities and, derivatively, to natural kinds.6

4. The Distinguished-Theory Approach to Kinds

I take it to be the upshot of the examples discussed above that neither the theory-based sweeping kind-realism sketched in Section 1 nor the two variants of the experiment-centered approach to natural kinds examined in Sections 2 and 3 succeed in reliably establishing a realism about natural kinds. Are we stuck with this null result? Mercifully not, I believe. Since the experimental approach to kinds has not stood up to empirical scrutiny, we are back with theories. But in light of the results of Section 1, we cannot rest our case on just any theory. Still, we may appeal to a particular subset of distinguished theories: theories, that is, distinguished by their outstanding empirical success. The basis for this distinction is the so-called Miracle Argument.

A sharpened version of the Miracle Argument says that there are two types of empirical success which can only be explained by recourse to realism, namely, novel predictions and consilience of inductions. If a theory succeeds in correctly predicting a novel regularity, formerly unknown and not to be expected against the background knowledge, this capacity is inexplicable (or

6Meehl is certainly right in pointing out that a large number of systematically evaluated case-studies provides a better basis for assessing the merits of metatheoretical claims than ‘informal, impressionistic (and often biased) reliance on selected case studies’ (Meehl 1992, 272). Still, if the set of possible counter-instances — i.e. abandoned entities — is comparatively small, one actual counter-instance serves as a large enough sample.
miraculous for that matter) unless we assume that the theory has got something right. Analogously, if a theory manages to unify regularities that appeared to refer to completely disparate phenomena before, and if it did so naturally, as it were, i.e. without introducing modifications and adjustments for the sake of producing the unification, this ability is simply mysterious unless we presume that the theory has grasped something correctly.\(^7\)

An example of a theoretically anticipated regularity is Einstein’s prediction of the correct magnitude of gravitation-induced light-bending within the framework of general relativity theory. The theory managed to foresee what no experimenter had yet seen, and this remarkable achievement strongly suggests that it is veridical in some respect. An example of consilience of inductions is the connection between black-body radiation and the photoelectric effect as forged by Einstein’s light quantum hypothesis. Einstein realized that the two phenomena are likewise governed by the quantum relation \(E=nh\), and the value of Planck’s constant \(h\) turned out to be identical in the two cases. It would indeed constitute a strange coincidence if two otherwise unrelated phenomena involved a numerical agreement of this sort. It is much more plausible to assume that the theoretical account that gives rise to this agreement truthfully reflects something real.

The mainstream realist position is theory-based entity-realism. This position involves commitment to the real existence of the entities specified by successful scientific theories, and this existence claim is backed by the supposed (approximate) truth of the relevant theories. It is this supposition that the Miracle Argument is intended to support. The intuition underlying the argument is as follows. There is a particular type of empirical success — that may be termed strong success — in which the domain of application of a theory extends itself naturally and without any deliberate adaptation for this purpose to phenomena the theory was not designed to accommodate. The occurrence of strong success is a remarkable and surprising event and is thus in need of an explanation. Realism provides such an explanation by arguing from the (approximate) truth of the theory to the existence of the relevant entities and their modes of interaction. A theory is strongly successful because it is essentially correct. This explanation is admittedly not complete since the empirical success of a basically correct theory may be thwarted by inadequate auxiliary assumptions. Although realism is thus not sufficient for the explanation of strong success it is nonetheless necessary. Without it we are at a loss to

account for strong success in any event. Realism thus explains how strong success is possible.

The history of science teaches, however, that the Miracle Argument is in fact unsuitable for supporting the approximate truth of strongly successful theories. This can be gleaned from the sketch given above of Stahl's phlogistic accomplishment. Stahl correctly predicted two novel effects. He produced a metal from its calx through the assistance of a non-metallic substance, and he managed to synthesize sulfur from its purported components (see Section 3). But according to present lights, the theory he employed for this purpose is wrong, and the central entity he invoked is non-existent. So we have come across an instance of strong success that can obviously not be explained along the lines of the Miracle Argument. Two additional counter-instances are provided by Joseph Priestley's prediction of the reductive properties of hydrogen based on a later version of the phlogiston theory and by Dalton's and Joseph Gay-Lussac's prediction of the quality of the thermal expansion rates of all gases based on the caloric theory of heat.\(^8\) In all these instances theories that are wide of the mark descriptively and referentially were nevertheless strongly successful. Accordingly, strongly successful theories may fail in the retention test. The conclusion is that the Miracle Argument does not support theory-based entity-realism.

But this is not the end of the story. My claim is that the Miracle Argument is basically all right; it was only applied to the wrong subject matter. What the argument in fact supports is a realism of kinds.\(^9\) The Miracle Argument is right in supposing that there has to be an explanation of strong success. It is wrong in attributing the reason to the truth of theories and to the existence of entities. What explains strong success is that the corresponding theories induced the right relations of similarity among the phenomena in question.

Let's take another look at the phlogiston example. Stahl's experiments worked because combustion and calcination are actually alike; they both constitute oxidation processes. He was only wrong in taking what is actually an oxygen transfer from sulfuric acid or lead oxide to charcoal to be a phlogiston transfer in the opposite direction. Although Stahl's account is now dismissed in its entirety, the classification it induced among the phenomena involved is retained until the present day. Oxidation of metals and non-metals are still considered as being equal in kind. The same goes analogously with respect to the two further examples mentioned. Priestley's novel prediction was borne out because — speaking in modern terms — it correctly regarded oxidation and reduction as the same process going off in opposite directions.

\(^{1}\)Cf. Carrier 1991, 29–31 for an analysis of these cases.

\(^{2}\)The view that (in distinguished cases) a theoretically generated classification of phenomena reflects a real order was first advanced by Duhem; cf. Duhem 1906, 24–28.
Dalton’s and Gay-Lussac’s prognosis was successful because it rightly linked the physical constitution of gases with their equal expansion rates (cf. Carrier 1991, 32–33). It is the retention of the kind-structure that provides the sought-for explanation of strong success.

It is to be noted that kind-retention is, in general, restricted to the phenomena connected by strong success. It does not automatically extend to other classifications specified by the same theory. Large parts of the kind-structure introduced by the phlogiston theory were overturned by the subsequent progress of science. In the framework of this theory, sulfur and metals belonged in the same natural kind as wax and oil. All these substances were purported to be alike in that they contain a large proportion of phlogiston. From the contemporary perspective, by contrast, the two groups constitute distinct kinds. The first group includes chemical elements and the second organic compounds.

The overall situation is thus as follows. We are presented with an epistemic argument to the effect that strongly successful theories should correctly reflect an aspect of reality. In order to single out this aspect we apply the retention test. Every theoretical feature that is supposed to reflect something real has to be retained across scientific change. It turns out, then, that theories and entities fail in this test whereas kinds pass it. The conclusion is that kind-structures backed by strong success are real.

5. Establishing a Natural Taxonomy of Kinds

It might be objected that the vast majority of the examples presented stem from outdated theories that may appear antediluvian to the present-day reader. Some may even doubt whether the phlogiston theory ever qualified as truly scientific. These misgivings are unjustified, however. All the theories I referred to were well-confirmed in their respective lifetimes according to our present methodological criteria and had received high marks from the corresponding scientific communities. They all formed part of ‘mature science’. After all, the phlogiston theory scored a strong success at its inception. Moreover, if we want to examine the appropriateness of retentionist claims we simply have to consider possible counter-instances, and such instances, in the nature of the case, are only provided by once successful but now rejected theories. Significant evidence as to retained features can only arise by examining what is left when everything else is gone. Accordingly, theories that appear strange today should not be disallowed as respectable philosophical examples for this reason alone.

It is worth noting, furthermore, that the argument for the reality of distinguished kind-structures is not beset with the following circularity. The
judgment about what is real and what is not is based on present-day knowledge which, however, according to the very approach advocated here, is not justifiably reliable in telling what there is. In fact, the argument involves no such circularity; its logic is as follows. Adopting scientific realism implies commitment to the realist interpretation of at least one aspect involved in contemporary scientific theories. This suggests that preceding theories that were once confirmed in roughly the same fashion as their successors are now should likewise be trustworthy ontologically with regard to the aspect in question. From this follows the retention condition: for anything to be counted as real, it is necessary that it is retained across scientific change. This condition is accepted by realists and anti-realists alike. In fact, the anti-realist argument from scientific change proceeds from precisely this condition: because nothing significant is retained, the realist claims are mistaken.10

The second step of the argument brings to bear the judgment that strong success is an extraordinary and astonishing phenomenon and deserves an explanation. The Miracle Argument provides us with a possible reason; namely, it attributes strong success to the fact that the relevant parts of the corresponding theory truthfully reflect something real. The Miracle Argument is, however, quite unspecific as to what precisely constitutes this veridical aspect. There are three possible candidates, namely, theoretical mechanisms, theory-introduced entities, and theory-induced kinds. In order to determine which of them fills the bill the retention condition is invoked. Application of this condition to the history of science singles out kinds as the only theoretical features that may legitimately be interpreted realistically.

In short, satisfaction of the retention condition constitutes a necessary precondition for the realist interpretation of a theoretical feature. It is seen, then, that it is at most kinds that can be so interpreted. If, in addition, the retained theoretical feature serves as a basis for an explanation of strong success, the Miracle Argument licenses its realist interpretation. Retention together with the capacity to furnish such an explanation is sufficient for attributing reality to it. If a theoretical feature passes both the miracle test and the retention test we are entitled to accept it as part of reality. My contention is that (some) natural kinds indeed pass both tests. From this follows a prediction for the future course of science: phenomena once connected by strong success will continue to be of the same kind in all subsequent theories on the subject.

It may appear problematical to ground such a far-reaching contention on the discussion of no more than three cases. But in fact the evidential basis is not that narrow. It also includes cases in which the occurrence of strong success was due to theoretical aspects we still take to be correct. Consider the

10This constitutes the central contention of Laudan 1981.
prediction of the phases of Venus as implied by Copernicus's theory. In this example of strong success, the relevant aspects of the theory, the entities involved, and the induced kinds are equally retained up to now. We still hold the view that Venus and the Earth revolve around the Sun, we still regard Venus, Earth, and Sun as acceptable entities, and we still stick to the kind-structure induced by the theory; namely, that—in contradistinction to the preceding account—Venus and Earth equally belong into the category 'planet' and the Sun does not. Cases of this type indiscriminately support all three positions at stake. The examples presented in Section 4, by contrast, constitute differential support for kind-realism. They can only be accommodated by kind-realism. And it would be over-demanding to require that the class of differentially supporting instances be large. In that event the alternative explanations would be obviously false and would never have gained the wide acceptance they enjoy.

The realism of kinds advocated here involves a peculiarly non-Aristotelian view of kinds. In the Aristotelian tradition, kinds are individuated through their essential properties. The class of human beings is rightly determined on the basis of the characterization 'rational animal', but not by means of the attribute 'featherless biped'. No such distinction between essential characteristics and accidental features is implied by the present account. Quite the contrary. My claim is that only the induced taxonomy, but not the theoretical means used for establishing it, is (in distinguished cases) justifiably to be considered real. It is true, our prime epistemic access to kinds is through theories; kinds are individuated by means of theories. Still, it is only the results, and not the means used for their production, that are arguably reliable ontologically. It is the relations of similarity among phenomena that (sometimes) deserve our confidence. No such case can be made for the theoretical mechanisms employed for specifying these relations and for the entities involved in them. It is clear that the phenomena collected into equivalence classes are equivalent in some respect. But precisely what this respect is cannot reliably be specified. Only the relation of similarity is legitimately to be interpreted realistically.

Acknowledgements—I am grateful to Peter McLaughlin, Alexander Rueger and two anonymous referees for their helpful comments.

References

Ben-Menahem, Y. (1990), 'The Inference to the Best Explanation', Erkenntnis 33, 319–344.
Carrier, M. (1991), 'What is Wrong with the Miracle Argument?', Studies in History and Philosophy of Science 22, 23–36.
Fodor, J. A. (1974), 'Special Sciences (or: The Disunity of Science as a Working Hypothesis)', Synthese 28, 97–115