Special-Science Autonomy and the Division of Labor

Michael Strevens

To appear in *The Philosophy of Philip Kitcher*,
edited by Mark Couch and Jessica Pfeifer,
Oxford University Press, Oxford.

Abstract

Philip Kitcher has advocated and advanced an influential antireductionist picture of science on which the higher-level sciences pursue their aims largely independently of the lower-level sciences—a view of the sciences as autonomous. Explanatory autonomy as Kitcher understands it is incompatible with explanatory reductionism, the view that a high-level explanation is inevitably improved by providing a lower-level explanation of its parts. This paper explores an alternative conception of autonomy based on another major theme of Kitcher's philosophy of science: the importance of the division of cognitive labor. That the sciences are in practice autonomous is compatible with reductionism, I argue, if we understand the high-level sciences' systematic explanatory disregard of lower-level details of implementation as practically, rather than intellectually, motivated.

How reductionism's star has fallen. Once it lived a blessed life in a grand logical empiricist mansion high on a philosophical hill; now it wanders the streets below, its face drawn and its clothes ragged, carrying under its arm a tired manifesto that no one any more wants to read. Among the revolutionary leaders responsible for its overthrow—among such firebrands as Jerry Fodor and John Dupré—is the sweet and reasonable voice of Philip Kitcher.[1]

Kitcher's "1953 and All That" advances three arguments against the reducibility of classical genetics to molecular genetics:

1. Classical genetics does not contain the kind of general laws required by Nagel's (1979) canonical account of intertheoretical reduction.

2. The principal vocabulary of classical genetics cannot be translated into the vocabulary of lower-level sciences; nor can the vocabularies be connected in any other suitable way (that is, by "bridge principles").

3. Even if the reduction were possible it would not be enlightening, because once you have the cytological explanation of genetic phenomena, the molecular story adds nothing of further interest.

This paper takes issue with the third of these arguments, contending that a robust explanatory reductionism can coexist with the sort of explanatory autonomy that Kitcher considers to be its manifest refutation. In the special sciences, then, we can have it all: a thoroughgoing explanatory autonomy in the lab and the field along with a severe reductionist philosophy of explanation in the all-seeing armchair.

I will not offer anything so immodest as a comprehensive case in favor of reductionism. I discuss only one form of explanatory reductionism, and I defend it against only one objection, the argument from autonomy, while giving no positive considerations in its favor. Reductionism will not be returned to

1. Fodor (1974); Kitcher (1984); Dupré (1993).

its place of former splendor, then—but I do hope to give it some sustenance, a steady job, and a place to hang its hat.

1. Antireductionism from Autonomy

Kitcher's argument from autonomy in "1953 and All That" turns on the explanation of an enhanced version of Mendel's second law:

> *Independent assortment:* Genes on nonhomologous chromosomes assort independently.

A satisfactory explanation of independent assortment, Kitcher asserts, describes meiosis, the process driving assortment, at the level of cell bodies and their interactions:

> Cytology provides the answer. At meiosis, chromosomes line up with their homologues. It is then possible for homologous chromosomes to exchange some genetic material, producing pairs of recombinant chromosomes. In the meiotic division, one member of each recombinant pair goes to each gamete, and the assignment of one member of one pair to a gamete is probabilistically independent of the assignment of a member of another pair to that gamete. Genes which occur close on the same chromosome are likely to be transmitted together (recombination is not likely to occur between them), but genes on nonhomologous chromosomes will assort independently (p. 347).

This is, of course, a standard textbook explanation of the Mendelian effect.

Surely it can be enhanced, though, by describing the molecular mechanisms that implement the processes in question? By explaining at the chemical level how chromosomes—long strands of DNA—are reassembled and assigned to gametes, will we not see still more deeply why genes assort independently?

Kitcher allows that some further details at the cytological level will deepen our understanding, such as a description of "the formation of the spindle and the migration of chromosomes to the poles of the spindle just before meiotic division", which will allow us to see that "the chromosomes are not selectively oriented toward the poles of the spindle" (p. 348). But there he draws the line. Molecular details—chemical details—of the working of the spindle, even those that bear directly on the equiprobability of orientation, contribute nothing to the explanation.

That is not to say that the molecular details do not add to our total explanatory knowledge. They constitute an "explanatory extension" of classical genetics, Kitcher writes, but he maintains that "it does not follow that the explanations provided by the [classical] theory can be improved" by attaching these extensions (p. 365). In other words, we can have lower-level understanding of (at least parts of) high-level theories, but this lower-level knowledge has no legitimate place in the explanations offered by the high-level theories. For the purpose of understanding independent assortment, then, "it's irrelevant whether the genes are made of nucleic acid or of Swiss cheese" (Kitcher 1999, 200). Indeed, specifying the molecular implementation of meiosis "would *decrease* the explanatory power" of the cytological explanation, because it would "disguise the relevant factor" (Kitcher 1984, 348).

In what follows I put aside the claim about disguise and decrease, which as Kitcher himself concedes in the same passage may be "too subjective", turning as it does on our own limited cognitive powers. What is important is the claim that the molecular details, because they are irrelevant, do not increase the power of the explanation. That is where reductionism comes to grief.

In "The Hegemony of Molecular Biology" Kitcher introduces a further example to make roughly the same point. How to explain the striking fact noted by John Arbuthnot, that in every year between 1628 and 1709 the number of males born in London exceeded the number of females? A schematic explanation puts together the fact that the sex ratio of humans and many other animals is 1 : 1 at sexual maturity with the fact of greater male infant mortality: though more boys are born, more die before puberty, so that the ratio at puberty is in large populations just about exactly 1 : 1.

This explanation benefits considerably from a fleshing out. It is augmented in particular, as Kitcher observes, by the evolutionary explanation of the 1 : 1 sex ratio so influentially advocated by R. A. Fisher (1930).[2] According to Fisher's story, the sex ratio has become fixed at 1 : 1 because the even ratio is a stable and unique equilibrium. And equilibration occurs at 1 : 1 because, in a population with more females than males, individuals with a propensity to produce more males than females will have a higher expected number of grandchildren and vice versa. Why a higher expected number of grandchildren? Your expected number of grandchildren is proportional to your expected number of children and your children's expected number of matings. Since matings require exactly one male and one female, a male's expected number of matings will increase, relative to a female's, as the proportion of males in a population decreases.

If explaining an explanation's explainers deepens the original explanation—if explanatory relevance is transitive—why not go further? Why not detail the biological mechanics in virtue of which, for example, successful reproduction always and only involves a single male and a single female?

Here Kitcher might have said: As with molecules in the Mendelian case, I draw the line at these gritty goings-on. As long as you see that procreation, if not copulation, is strictly one on one, it is irrelevant whether the protagonists are made of meat or Swiss cheese.

In fact he is somewhat more circumspect, more choosy about the ingredients of the explanatory sandwich. Certain details of implementation are worth investigating, he remarks. Presumably he allows that even quite low-level details might appear in an explanatory extension of Fisher's model. But they

2. On the prehistory of the explanation in the work of Darwin and later writers, see Edwards (1998).

do not, merely because they extend the model, count as explanatorily relevant to the things that the model explains, such as the sex ratio, and if they are sufficiently low level they certainly do not count as relevant.

Indeed, in the sex-ratio case and almost everywhere else, Kitcher appears to hold the following view: moving down the levels of potential explanation from ecology to physiology to cytology to chemistry to fundamental physics, there is some point beyond which further unpacking of mechanisms becomes entirely irrelevant. Thus, for example, deriving the one-on-one nature of procreation from quantum mechanics adds nothing whatsoever to our understanding of Arbuthnot's observation. Thanks to some aspect of the nature of scientific explanation, the facts of implementation cease to explain when the scale reaches so fine a level of description.

What explanatory principle is it that undercuts the transitivity of explanatory relevance, severing the link in certain cases between the explainers of explainers and the explanandum?

In "1953" Kitcher accounts for failures of transitivity by proposing that, when transitivity falls through, it is because categories essential to explaining high-level phenomena cannot be ascribed explanatory relevance by lower-level explanations due to their not constituting natural kinds from the lower-level point of view (Kitcher 1984, 349).

This is a peculiar argument to use against the central target of the paper, reductionism as understood by the logical empiricist tradition: the empiricists had little use for the notion of a natural kind, except perhaps as an honorific bestowed on especially useful categories at the end of inquiry (Quine 1969). Most modern philosophers of science will also, I think, doubt that there is some notion of naturalness that precedes and constrains the facts about explanatory relevance. Certainly, I myself reject this thesis unreservedly.[3]

3. And I reject the underlying theses: that such a notion of naturalness constrains what may count as a law of nature, and that explanation requires laws of this sort. These ideas about the importance of naturalness, also found in Fodor (1974), are I presume supposed to allude to Hempel and Oppenheim's (1948) search for a notion of "lawlikeness" to constrain what may

Something like it has gained prestige in recent years thanks to the work of David Lewis (1983), but Lewis's metaphysics does not supply what Kitcher's argument requires, a notion of high-level naturalness that cross-cuts physical naturalness.

Perhaps it is better to understand Kitcher's appeal to naturalness from the perspective of his own unificationist account of explanation (Kitcher 1981, 1989). The permissible content of an explanation is dictated, the unificationist says, by the argument patterns that appear in the "explanatory store"—that is, in the set of most unifying patterns. Among the constraints imposed by such patterns are limits on the properties that may be mentioned in an explanatory argument. If we understand a property's being natural at a certain level as its being allowed into the characteristic argument patterns of that level, and if a certain principle of conjunction is denied—that concatenating two permissible argument patterns always produces a third permissible argument pattern—then we can make sense, in a unificationist context, of Kitcher's claim that a factor *A* may be a natural explainer of *B*, and *B* a natural explainer of *C*, without *A*'s being a natural explainer of *C*.

For the purposes of this paper, however, I will put aside the question of why Kitcher treats explanatory relevance as intransitive, of why he thinks that the correct story about an explainer's implementation does not always contribute to the explanation. Let me simply take it as given that Kitcher's claims about relevance are correct. It is a fact, I will suppose, that scientists of the high level regard much information about implementation as irrelevant

count as a law of nature and so play a role in deductive-nomological explanation. Hempel and Oppenheim's strategy is to rest lawlikeness on the notion of a "purely qualitative" predicate, which they attempt to precisify in terms of formal logic. Though by their own admission they do not entirely succeed, they would surely have regarded an appeal to a metaphysics of naturalness with unalloyed horror. In any case, there is nothing in their conception of a purely qualitative predicate that would rule out the use, in molecular biology, of the sort of functional definitions that give rise to multiple molecular realizability (which is what Fodor, Kitcher, and others take to preclude the corresponding properties' naturalness). Hempel and Oppenheim's concern is with predicates that are overly specific; they are aiming for, not avoiding, generality.

to their research—the sociological fact of explanatory autonomy. Assuming that those scientists are not wholly mistaken, a philosophical fact follows immediately: details of implementation often *are* explanatorily irrelevant, and below a certain level of description, in many or perhaps even all cases, *every* detail of implementation is irrelevant.

That irrelevance looks to be flatly incompatible with explanatory reductionism. Or as Fodor (1974) memorably writes:

> Reductivism . . . flies in the face of the facts about the scientific institution: the existence of a vast and interleaved conglomerate of special scientific disciplines which often appear to proceed with only the most token acknowledgment of the constraint that their theories must turn out to be physics "in the long run". (112–113)

The next section outlines a notion of explanatory reductionism that proclaims exactly what Kitcher and Fodor deny: the explanatory relevance, always and everywhere, of fundamental physics. I will then show how to make that reductionism fit with the sociological and philosophical fact of explanatory autonomy.

2. Explanatory Reductionism

The physical world, there is ever more reason to think, is the only world we have. Everything is made of physical stuff, and everything that happens, happens because of the way that physical laws push physical stuff around. This is the doctrine of physicalism. It is deniable, but ever ascendant.

From physicalism it follows that any state of affairs or pattern of behavior we find in the world, no matter how high level or abstract, can be derived from fundamental physical facts and laws. It also seems plausible (though it does not strictly follow) that any state of affairs or behavior at a given level can be derived from facts and regularities about entities at "the next level down". Thus, economic regularities can be derived from psychological facts and regularities; psychological regularities from physiological facts and regularities; and so on through cytology, molecular biology, chemistry, to fundamental physics. (The relevant lower-level facts will concern to a great degree the arrangement of and relationships between lower-level entities: we do not get psychology from the study of neurons in isolation, but from the study of the neural structure of the brain as a whole.) This picture, reminiscent of Oppenheim and Putnam's (1958) unity of science manifesto, is a little too simple: sometimes a derivation will have to pull facts from different levels (combining, say, physiological facts about organisms with physical facts about climate), and indeed, the picture of the world as organized into levels corresponding to university departments is surely a caricature. All of this may be allowed without compromising the basic assumption of derivability from the lower level.

Now let me go further. Some of these derivations are also explanations; further, for any higher-level phenomenon there is at least one derivation from the lower level that is explanatory. (There are also, in my view, many such derivations that are not explanatory, for reasons I give in the next section.)

The fact of explanatory autonomy is quite consistent with everything I have said so far. Indeed, Kitcher himself may well agree with most of it; his favored explanation of independent assortment, a regularity identified by classical genetics, takes the form of a derivation from the cytological level, and his favored explanation of the 1 : 1 sex ratio at maturity takes the form of a derivation from facts about individuals' grandchildren's prospects—hence, properties of a population are derived from individuals' propensities to prosper. More generally, Kitcher's notion of an explanatory extension turns on the availability of many such derivations, though perhaps not a derivation for absolutely every high-level state of affairs.[4]

4. Kitcher's skepticism about bridge principles for general categories such as *gene* is, as he himself notes, compatible with the derivability of all observed behaviors of genes, as the behaviors may be derived piecewise, that is independently for particular genes.

A final step, however, will secure a collision with Kitcher in particular and autonomy in general, and will transform the view I am describing into something that is clearly reductionist. It is to endorse the transitivity of explanatory relevance: explain something that explains a phenomenon, I propose, and you have added something relevant to, and so enriched, your original explanation. The cytological explanation of independent assortment is good, but it can be made even better by explaining the cytological explainers in turn—by giving a molecular explanation of the relevant cytology. And better still by providing the chemistry of the relevant molecular facts. Best of all is to derive, from the fundamental physics, the relevant chemical facts. At that point you can descend no further; you have an explanation that is in one sense maximally good, because maximally reductive.[5]

Give the thesis sketched above a name: *explanatory reductionism*. An explanatory reductionist, then, holds that an explanation is always improved by giving a lower-level explanation of its parts—of the initial conditions, regularities, and structural facts that it cites—and that such further explanations are available for every non-fundamental part.

How is this compatible with the sociological fact of autonomy—with the fact that working scientists consider many and in some cases all lower-level details to be irrelevant? That looks like a simple question; the answer, clearly, is that the fact of autonomy refutes explanatory reductionism.

But no: it is a complex question, and compatibility is possible after all. The key is to identify more than one sense of explanatory relevance. Explanatory reductionism is true of one sense; autonomy of the other.

To make this dichotomizing plausible, I will in the next section introduce a reductionist account of explanation—my own kairetic theory—and flesh out the notion of explanatory relevance at its heart. A great deal of lower-level de-

5. If particular matters of fact appear in the explanation, it might be further improved by tracing back in time the genesis of those facts. But this proposal (Strevens 2008, §4.31) has nothing to do with reductionism.

tail turns out to be irrelevant to high-level phenomena. But not all detail—as must be the case, since the kairetic theory is a form of explanatory reductionism. The discussion thus opens the door to, without entirely achieving, the synthesis of reductionism with autonomy. That goal requires the postulation of the second kind of explanatory relevance in section 4.

3. The Kairetic Account of Explanation

The kairetic theory of explanation is a causal theory. It begins with physicalism, not only about things and laws, but about causation itself: the raw material of causality is the fundamental-level relation of causal influence. Newtonian force is the paradigm of such an influence relation. In the Newtonian worldview, force is responsible for all changes of velocity of all objects, microscopic or massive, and so, along with inertial motion, for everything that happens. It is described completely by fundamental physics.[6] The question whether there is causal influence to be found in modern physics is more fraught, but I ask you to put it entirely to one side, and to suppose that physics does indeed give us such a relation, as many philosophers of causation have argued (Reichenbach 1956; Salmon 1984; Dowe 2000; Lewis 2000).[7]

The kairetic account holds that an event or phenomenon is to be explained by showing how it was produced by the aggregate causal influence of other states of affairs and the laws in virtue of which they exerted their influence. In a Newtonian world, for example, an event is explained by exhibiting whatever objects and events pushed around certain constituents of the world in a way that realized the occurrence of the event, along with the laws in virtue of

6. To say that influence is a fundamental-level relation is not to say something about its metaphysical foundation, but only about its relata. It leaves open the possibility that the facts about influence are determined by other facts, even high-level facts such as the direction of entropy increase or the structure of human causal concepts.

7. As this list makes plain, there are many ways to give a metaphysical theory of causal influence. There is no need, in what follows, to choose among them.

which the pushing around occurred, that is, the laws in virtue of which the relevant forces were brought to bear.

It would be a mistake, however, to proffer as the explanation of an event the entirety of the web of antecedent causal influence in which it is embedded—so says the kairetic account. What explains the event is only those aspects of the web that *made a difference* to the event's occurrence. Various familiar philosophical accounts can be given of difference-making. A statistical relevance theory holds that an aspect of the web of influence made a difference to an explanandum event if it raised the probability of the event. A counterfactual theory requires of difference-makers that, had they not been present, the explanandum would not have occurred.

The kairetic account of difference-making provides the following recipe for determining difference-makers. Take a comprehensive causal model of the production of the explanandum, that is, a complete description of the web of influence leading up to, but not including the explanandum. Assume, for simplicity's sake, that the explanandum is deterministically produced and so that the description entails its occurrence. Now: make the description as abstract as possible without either undermining the description's entailment of the explanandum or undermining the description's status as a causal model. What remains in the description after this process of abstraction are specifications of difference-making factors. (Needless to say, a fully adequate presentation of the recipe requires more than a few sentences; for the official version, including the criterion for a description's constituting a causal model, see Strevens (2008).)

To explain, for example, the Broad Street cholera epidemic of 1854, begin with a complete description of the epidemic: the leaching of the cholera bacteria into the water supplying the Broad Street pump, the carrying of the water to various residences, its ingestion by various people, the course of the disease in each victim, its transmission perhaps to others. Notionally, the process is described at a maximal level of detail: the position of every bacterium and the disposition of every drop of water. Now take away what you can without invalidating the description's entailment of the explanandum, that is, of the fact of the epidemic. The precise positions of individual microorganisms do not matter at all; that information can be deleted, leaving only a specification of the approximate density of the organisms in the water retrieved from the pump. The time of day (I assume) that the water is pumped is also irrelevant; what matters is only that a certain amount was consumed. Likewise, the course of death need not be charted in excruciating detail; the rough facts about the degree of dehydration and its inevitable physiological effects is enough to entail an upward step in the statistics of mortality. Throughout this process of information removal, nothing is added; rather, more detailed specifications of the causal web are replaced with strictly less detailed specifications. What you are left with is a description of the same causal web with which you began, but an extremely abstract description: bacteria in considerable quantities leaked into the water supply; the water was consumed by a significant number of people; given the prevailing conditions, also specified at a high level of generality, they went on to contract cholera.

More or less the same recipe applies to the explanation of regularities. Why does Mars' orbit around the sun conform approximately to Kepler's laws? Begin with a complete causal model for Mars' orbit—a complete specification of the causal influences on the planet's trajectory over the course of a Martian year, along with the relevant physical laws. Your model will predict every minor twist and turn in Mars' movement. But with or without these perturbations—with or without the other stars, the planets, the interstitial rubble—the model predicts Keplerian behavior. The kairetic criterion therefore orders the perturbers' removal, or more exactly, it tells us to replace the painstaking specification of the distribution of mass with something as abstract as possible having the same net implications for the explanandum—in this case a specification that the total gravitational force due to objects other than the sun did not exceed some (small) upper bound. What remain are the difference-makers: the physics of gravitation; the relative size, position and velocity of Mars and the sun; and the aforementioned upper limit.

Note that to determine explanatory relevance it is not necessary to construct complete causal models. We can see that certain things are not going to count as difference-makers without going through the rigmarole, so it is possible to have knowledge of difference-makers while having only a very rough knowledge of the underlying causal web. The kairetic recipe is the ultimate criterion for the correctness of claims about difference-making, but here as so often in life we do not deploy the ultimate criterion in day-to-day cognition.

For the purposes of this paper, two things matter about the kairetic account. First, for many high-level or coarse-grained explananda (epidemics, approximate orbital trajectories), it declares vast amounts of physical detail to be explanatorily irrelevant, even when that detail has some causal connection to the course of events to be explained.

Second, the kairetic account is nevertheless a species of explanatory reductionism. This is because its specifications of difference-makers, thus its explanations, are descriptions of the web of causal influence, hence of properties of the fundamental laws of physics and the physical states of affairs, whose aggregate causal impact determined that the explanandum holds or occurred. A complete kairetic explanation of Keplerian behavior does not have to specify in any level of detail the distribution of mass in the solar system, but it does have to describe what that mass is made of and how, at bottom, gravitational attraction works: curved space-time, geodesics, stress-energy tensor and all.[8] Conversely, an explanation that does not have something to say about the fundamental laws or configurations of things has not said everything that an

8. The question of exactly what details can be removed from a description of the gravity mechanism is determined by the kairetic account's "cohesion" constraint. Rather than attempting to describe the workings of the constraint in this paper, I simply spell out its reductionist consequences—which reductionism is what matters for the purposes of the discussion.

explanation should say. To add what is missing will improve the explanation.

Putting it another way, an explanation that bottoms out at, say, the cytological level implies that there is something about the web of fundamental-level influence that makes certain cytological claims true, but does not specify what that something is. According to the kairetic account, it is obliged to do so. The most natural and straightforward way to discharge the obligation is to give an explanatory derivation of the relevant cytological facts and generalizations from the physical facts and generalizations.

There is much that such a derivation will not say about the physical realization of the cytological facts: it specifies only those things that make a difference to the cytological facts' holding. For example, it will omit pointless enumerations of the precise vibrational energies of various components of, and for that matter the precise positions of, various important molecules. But it will specify approximate positions, and lay out in physical terms the relevant properties of DNA, of telomeres, of the spindle structure in meiosis, describing the fundamental physical basis of the molecular interactions that make the whole machine work the way it does. It will, that is, give precisely the details, and only the details, you would expect if you were to ask a molecular biologist to explain the relevant features of the cytology, and they were in turn to ask a physicist to explain the relevant features of the molecular biology.

A reductionist theory of explanation can, as the kairetic account shows, prescribe very abstract explanations, but they must be at the same time physical explanations. There, it seems, is the rub: the fact of explanatory autonomy is simply the fact of the irrelevance of these physical explanations. Their high degree of abstraction is no palliative. The kairetic account says that the physical structure of the spindle must be specified, albeit not in great detail. Autonomy says it does not. Thus the kairetic account, like all reductionist accounts of explanation, is false—so you might conclude.

What next? One strategy is to give something like the kairetic account a stronger criterion for difference-making—strong enough to imply the absolute irrelevance of all facts below a certain level (Franklin-Hall forthcoming). Another is to find a different kind of explanatory irrelevance, orthogonal to the irrelevance diagnosed by the kairetic criterion, and to show that autonomy is all about that other kind of relevance. That is the route I will take.

4. From Contextual Irrelevance to Autonomy

If a detail makes no difference to a phenomenon, say that it is *objectively irrelevant* to the explanation of that phenomenon. It is compulsory, when explaining, to ignore objectively irrelevant factors. There is another reason for ignoring details that is not compulsory in quite the same way, but that allows for the existence of explanatory models that say nothing about lower-level implementation; I call it *contextual irrelevance*.

The notion of contextual irrelevance is not proprietary to the difference-making approach to explanation: any philosopher of explanation, whether reductionist or antireductionist, whether allied to the difference-making, the unificationist, or some other view, should recognize the existence and importance of irrelevance of this sort. As I will eventually show, however, the notion of difference-making does have something important to contribute to our understanding of the systematicity of contextual irrelevance.

4.1 Contextual Irrelevance Suppose that a team of archeologists in the distant future excavate a Ford factory and attempt to reverse-engineer the internal combustion engine—to construct an explanation of how the thing works. They might reasonably pursue their task by dividing the engine into (apparent) functional units: the piston assembly, the starter motor, the cooling system, and so on. A team would be assigned to each unit with the task of determining how that unit contributed to the functioning of the whole. The advantage of such a division of labor is, of course, that one part of the engine can be analyzed without any detailed understanding of the other parts.

Although the other parts cannot be ignored altogether, they can be treated as "black boxes" with specified inputs and outputs but no internal details. The team working on the piston assembly needs to know that there is a cooling system, and something about the parameters of that system's operation—how it reacts to increasing engine temperature, how fast it cools, and so on—but that is all.[9]

The full explanation of the engine's workings is attained by bringing together the teams and replacing each black box with the proper internal causal model of its workings (omitting objective non-difference-makers). Until that time, no individual fully understands the engine. This gap in each investigator's knowledge is all to the good: it would be a mistake, in a practical sense, for the piston people to start thinking about cooling, as it would unravel the efficiencies gained by the division of epistemic labor. Thus, the prime directive or principal norm governing each team's subproject is as follows: the details of any other component's workings are to be considered irrelevant to your explanation; in deciding what does and does not matter for explanatory purposes, pay them no heed.

This is what I mean by contextual irrelevance in explanation. The workings of the cooling system are objectively relevant to explaining the engine's behavior but they are contextually irrelevant—irrelevant, that is, according to the procedural rules laid down for the the archeologists' investigation of the piston assembly. Contextual relevance is quite different from objective relevance. Whereas the facts about objective relevance are decided by the world, the facts about contextual relevance are decided by us, the organizers of inquiry into the world. Whereas including objectively irrelevant factors in an explanation is an intellectual error, including contextually irrelevant factors is a social or practical error. Nevertheless, the success of an explanatory

9. Does a black box specify causal relationships (without specifying a mechanism), counterfactual relationships, actual-world regularities, or something else? In this paper I leave the question open; since black boxes function only as placeholders for explanatory parts and not as explainers themselves, it does not much matter.

investigation may hinge just as much on the norms of contextual relevance as on the norms of objective relevance. Investigators must take both equally seriously. (I should add that the norms for both kinds of relevance govern what goes into an explanation, but they do not govern what the explanatory investigators may contemplate. A successful explainer will of necessity have to spend rather a lot of time thinking about non-difference-makers, just so as to be able to recognize them as such.)

4.2 The Division of Scientific Labor and Autonomy That significant philosophical consequences follow from science's division of cognitive labor among competing approaches to a single problem or domain is one of Philip Kitcher's enduring contributions to our field (Kitcher 1990, 1993). Science also divides its labor among complementary fields of research, in two ways.

First, like the archeologists with their engine, it implements what might be called *functional compartmentalization*, as when in the study of the physiology of the human body the immune-system specialists black-box the parts of the brain controlling the respiratory system, or when in evolutionary biology some biologists black-box the mechanisms of inheritance while others black-box the phenotypic causes of differential reproduction (Potochnik 2010).

Second, science implements what might be called *functional stratification*, the individuation of the explanatory pursuit by levels: the physicists study the fundamental particles, the chemists the molecules made up by those particles, the cytologists the cells made up by those molecules, and so on. Typically, investigation at one level black-boxes everything that goes on at lower levels—not ignoring lower-level goings-on altogether, but introducing them by way of functional specifications, as when chemists (though not quantum chemists) talk in terms of bonds or nuclei, without asking how a chemical bond is implemented or a nucleus constructed, or population ecologists talk in terms of predators and prey, without attempting to represent the mechanics of predation.

The notion of a black box is the same whether you are doing functional compartmentalization of an engine or functional stratification of an ecosystem. But what you do with the black boxes is somewhat different: in compartmentalization you plug black boxes into your system's inputs and outputs; in stratification you build your model of the system with black boxes.

I will focus, naturally, on stratification. One motivation for functional stratification is objective irrelevance: if the details of a lower-level process do not make a difference to a high-level explanandum, they should be omitted from the explanation. But not all black-boxing of the lower level can be accounted for in this way. Consider, for example, a population ecologist's explanation why a boom in the predator population can lead, down the line, to a boom in the population of prey: the predator boom causes the prey population to crash; the paucity of prey then causes the predator population to crash; the prey population, having a higher rate of reproduction or shorter generation length, subsequently recovers more quickly, and without a substantial predator presence, grows much faster than it normally would. Such an explanation would typically account for the population change in each of these phases using a standard Lotka-Volterra model, which represents the rate of predation as increasing with the number of predators but supplies no further information about the way in which predators hunt for prey (Roughgarden 1979). Such a model black-boxes predation: it specifies the effect of predator numbers on prey numbers, but says nothing about the mechanism responsible for the effect. The mechanism is, however, not irrelevant to the explanation in the objective sense: if you are trying to understand the relation between population booms and busts, and it turns out that the relation between predator population and predation rate plays an important role (as it clearly does in the case at hand), it is far from irrelevant to ask: "Why is the rate of predation proportional to the number of predators?" (Lotka 1956; Holling 1959). If the predation mechanism does not appear in the explanation, then, it must be because the population ecologists have been excused, by some principle of division of labor, from having to supply it.

To turn this into a general proposition: the principle by which the economists black-box psychology, the psychologists black-box neuroanatomy, the evolutionary biologists and population ecologists black-box various aspects of physiology, and so on, is the principle of contextual irrelevance. It divides the explanatory enterprise into many different parts, in this case layers or strata, for the sake of more efficiently producing the components of complete explanations.

The efficiency exists for two reasons. First, puny human intellects think better when relieved of the burden of bearing in mind the big picture. Second, many explanatory models overlap at higher levels of organization. Predator-prey ecosystems that look quite different at the organismic level—pelagic fish versus woodland animals, say—may and often do share certain mathematical properties when viewed in the abstract; further, these high-level similarities have important ecological consequences. It would be a waste of many scientist-hours to have the fish ecologists and the mammal ecologists independently derive the consequences; better to have a single group working on the high-level properties while the fish and mammal specialists confine their attention to those lower-level phenomena where scales and fur come apart.

The explanatory pieces so efficiently produced are made for a higher purpose, to be sewn together into complete explanations, that is, explanations in which all objective difference-makers, whether described at the level of physics, cytology, psychology, or whatever, are brought together into a single explanatory model. It is one of the great glories of modern science that we have, in many cases, the necessary materials for something approaching this full understanding—that we have, in other words, the ability to sketch answers to the chain of explanatory questions leading all the way down to the fundamental level. But of course, no one person is custodian of this understanding. It is spread across the sciences, with (in the predator/prey case) the population ecologists leading off, the cognitive ethologists then taking up the baton, and so on all the way to the particle physicists.

It would be madness to organize explanatory inquiry in any other way. The creation of explanations is therefore a fragmented process; fragmentation is, however, merely a means to a unitary end that can be achieved only by science as a whole.

The conception of explanation I have proposed marries reductionism and autonomy. On the one hand, autonomy: the work of explanatory inquiry is divided among many domains, each of which is not merely permitted but required to black-box the explanatory models generated by other domains. Within a domain, an explanatory model is complete if it omits only details that are either objectively or contextually irrelevant.

On the other hand, reductionism: explanatory models within a domain are a means to a greater end, namely, explanations that replace all black boxes with substantive models accounting for the explanandum in physical terms, that is, relating what goes on to the fundamental-level causal web, as the kairetic account requires. Such an explanation is complete if it omits only details that are objectively irrelevant.

Two complementary senses of explanatory completeness figure in this scheme of things: contextual completeness, that is, completeness relative to the standards by which explanatory labor is divided among domains, and what you might call absolute completeness. The ultimate end of explanatory inquiry is the reductionist goal of completeness in the absolute sense; however, the means to this end, given the division of explanatory labor, is contextual completeness, and so it is completeness in the contextual sense that dominates scientific practice day to day—for which reason you find in every domain black-boxing explanations that are rightly called "complete". That is no reason, I hope I have demonstrated, to embrace antireductionism; practical considerations explain black-boxing equally well on either a reductionist or an antireductionist approach to explanation.

4.3 The Systematicity of Contextual Irrelevance There is a systematicity to the facts about objective relevance, that is, to the facts about difference-making. It is not that stardust makes a difference to the orbits of some planets but not to others, or that the vibrational modes of a telomere make a difference to meiosis in some types of cell but not in others: the non-difference-making, the irrelevance, is across the board. High-level explainers may consequently ignore certain kinds of detail as a matter of general policy; they need not treat each case on a custom basis.

Approaching the question of orbits, the celestial mechanic has good reason in advance to abstract away from planetary constitution and the existence of interplanetary pebbles and dust. For explanatory purposes, they need not see their system as made up of molecules at all; it is enough to see it as made up of planets—discrete spherical objects of great mass and indeterminate composition. The facts about difference-making, then, provide the explainer, by way of abstraction, with a high-level taxonomy of the system in question from which certain aspects of the underlying fundamental physics have disappeared altogether.

Likewise, the cytologist is apt to think in terms of telomeres rather than in terms of DNA sequences, let alone in terms of the individual atoms that make up such sequences, each with its own particular position, velocity, modes of vibration and so on, and the population ecologist will naturally think in terms of populations or standard types rather than actual organisms in all their individuality, let alone in terms of the organisms' molecular makeup.

Contextual explanatory relevance is also systematic: there are long-lasting, across-the-board rules determining what is and is not contextually relevant within an explanatory domain.

That may sound dubious. What is contextually relevant is determined by the way in which the explanatory enterprise is parceled out among the members of the scientific community. These allocations are nothing more than epistemic heuristics; why expect such matters of practical policy to be uniform over time, from place to place, or across a discipline?

If research strategies were simply a matter of fashion, systematicity would be surprising. But given their goal of investigative efficiency, they are strongly constrained by the world out there: there are certain divisions of explanatory labor that are very efficient, and many that are grossly inefficient.

Consider again the investigation of the internal combustion engine. One way to divide the work of understanding the engine is for one team to take the left half, one team the right half. The left-hand team might find themselves, then, trying to understand the workings of the left-hand side of a piston while black-boxing the piston's right-hand side. This is not impossible in principle, but the black box in question, which must specify all behavior of the right-hand side that makes a (relevant) difference to the behavior of the left-hand side, will have to contain an extremely detailed description of the right-hand side—so detailed that it will hardly be a black box at all.

More generally, black-boxing is useful only when the system to be compartmentalized or stratified is somewhat modular, in the sense that it can be divided into units or strata, each of which makes a difference to the others in ways that can be specified compactly and tractably.

The sense of difference-making relevant to determining the proper level of detail in a black box's functional specification is not quite identical to the objective difference-making discussed in the previous section, but turns on the same core idea and is equally objective. Here, in outline, is a kairetic characterization of the notion (though there are no doubt other viable approaches to characterizing functional difference-making).[10]

Take some proposed division of your system into putative black boxes, either at the same level of description (compartmentalization), or at different levels (stratification), or both. The aim is to determine, for any such division,

10. Perhaps the most important difference, on the kairetic approach, between the kind of difference-making that applies to causal factors and the kind of difference-making that applies to functional specifications is that the latter allows for difference-making properties that are radically multiply realizable.

which aspects of the proposed black boxes are difference-makers for other boxes and which are not.

To this end, associate with each proposed black box a complete functional specification of the box's inputs and outputs—not just the apparently relevant inputs and outputs, but every way in which the box interacts with its surroundings. This specification will, therefore, be a comprehensive mapping from environment to behavior. The functional specifications, together with whatever aspects of your explanatory model are not black-boxed, if any, will entail the explanandum (as in the previous section, I assume determinism).

Now, delete everything from a box's functional specification that can be removed without invalidating the entailment of the explanandum, using the same sense of removal or abstraction that the kairetic account applies to determine objective explanatory relevance. What remains are the difference-making elements of the specification.

An efficient division into black boxes is one that allows the removal of as much detail as possible, so enabling the researchers working on each subunit in a compartmentalization, or each level in a stratification, to proceed knowing relatively little about the other researchers' work, because they have in the short black box specifications everything they need to complete their own allotted task.[11] Science seeks an efficient division of explanatory labor, hence an efficient black-boxing scheme, for purely practical reasons. But the facts about a black-boxing scheme's efficiency are not pragmatic or observer-relative; they depend ultimately (according to the kairetic account) on the entailment relation, which resists all attempts at persuasion and intimidation, and is oblivious to intellectual fashion.

11. As always with matters determined by difference-making, the complexity of the explanatorily relevant functional specification is relative to the explanandum. If you are trying to explain why the engine weighs what it does, having one team determine the weight of the right half and the other team determine the weight of the left half will not create ludicrously complex black boxes, since for the purpose of understanding the right half's weight nothing at all need be known about the behavior of the left half—the one half makes no difference to what the other half weighs.

In short, contextual irrelevance depends on the efficient allocation of cognitive labor, which depends on the opportunities for compact black-boxing on offer, which depend in turn on a kind of "functional difference-making" very closely related to the objective causal difference-making discussed in the previous section. The facts about causal difference-making are, I suggested above, systematic; for the same reasons, I propose, the facts about functional difference-making are also systematic. As before I am generalizing speculatively from a small if wide-ranging set of examples: population ecology will, for most purposes and in most populations, be able to get by with relatively brief functional specifications of its organisms; economics of its rational actors; chemistry of its bonds.

The division of explanatory labor will inherit such systematicity. The same details of physical implementation will be ignored for a wide range of explanatory tasks in a given high-level domain; consequently, the domain will acquire a certain explanatory taxonomy and a certain distinctive explanatory style. In this way, the unabashedly reductionist kairetic theory of explanation explains why the sciences form a mosaic of explanatory sub-cultures, autonomous units each only peripherally and sporadically concerned with the work going on next door.

5. Two Antireductionist Arguments

The reductionist can, I have shown, explain the fact of autonomy. But is their explanation as good as the antireductionist's? Two antireductionist arguments may suggest not.

5.1 The Argument from Parsimony I use two notions of relevance, objective and contextual, to make sense of autonomy, whereas the antireductionist perhaps needs only one. Should an autonomy theorist inclined to parsimony opt for the antireductionist story with its unified conception of relevance?

Hardly. First and most briefly, an antireductionist account of autonomy is likely to posit far more in the way of local rules, cultures, ontologies, and epistemologies than a reductionist account, which sees scientific practice as (at bottom) unified.[12] If parsimony points somewhere, it is toward unity and reduction.

Second, the notion of contextual relevance has many uses within the philosophy of science: it occurs whenever a potentially relevant factor is put into what I call the *explanatory framework* (Strevens 2008, §5.3). Such factors are typically objectively relevant; their being placed in the framework, however, renders them contextually irrelevant, hiding them by fiat from the explanatory spotlight. There are numerous practical reasons to put something into the framework and thereby to put it, explanatorily speaking, off limits. The division of labor is one; the pragmatics of conversation is another; personal interests are another still. As long as explanatory discourse is responsive to these forces, there will be frameworking.

Frameworking and contextual relevance can be used in addition to found the distinction between causes and background conditions; to make sense of apparently invidious distinctions among the explanatory relevance of various absences; to interpret talk of prevention; and to account for the apparent failures of the transitivity of difference-making. (See Strevens (2008), respectively sections 6.1, 6.3, 6.4, and 6.5.) Contextual relevance is, therefore, very likely to remain with us whatever the fate of reductionism. To use it to explain autonomy requires nothing that we were not carrying already.

Third, there is something of a test to diagnose contextual as opposed to objective relevance, and it shows that there is plenty of both. The test begins with the idea that, because what is contextually relevant varies with the explanatory framework, assertions of irrelevance will appear to be more secure when the framework is clearly fixed, best of all by explicit specification

12. Exception: an account on which the explanatory boundaries between the disciplines are dictated by a single interdisciplinary criterion, as in Franklin-Hall (forthcoming).

using the words *given that*, *assuming that* or similar. Consider, for example, the following two claims:

> The molecular implementation of meiosis is irrelevant to understanding the independent assortment of genes on non-homologous chromosomes.
>
> Given that the molecular implementation of meiosis ensures that individual chromosomes are not selectively oriented toward one pole or other of the spindle, the workings of the implementation are irrelevant to understanding independent assortment.

The status of the first claim is at best somewhat murky; the second, by contrast, seems quite reasonable. Its rightness is palpable because the "given that" locution specifies that the objectively relevant parts of the molecular implementation are being frameworked, hence excluded from the explanatory picture. The same locution has no effect when dealing with objective irrelevance:

> Rubble in the Kuiper belt is irrelevant to Mars's orbit's conforming approximately to Kepler's laws.
>
> Given that Kuiper belt rubble exerts only a minuscule gravitational force on Mars, it is irrelevant to Mars's orbit's conforming approximately to Kepler's laws.

"Given that" does not improve the irrelevance claim at all; indeed, it gives the impression of sloppy writing. "Because" is far better. Put "because" in the claim about independent assortment, however, and you get:

> Because the molecular implementation of meiosis ensures that individual chromosomes are not selectively oriented etc., the workings of the implementation are irrelevant to understanding independent assortment.

That seems malformed, indeed perverse, whether or not you are an explanatory reductionist.

There are two kinds of irrelevance claim, then: the kind that sounds good with a "given that" rationale and bad with a "because" rationale and the kind for which the situation is reversed (unless "given that" is heard as meaning "because"). Now a sweeping claim: Take all of the antireductionists' claims of irrelevance for lower-level details that the kairetic theory counts as difference-makers. Apply the "given that/because" test. Their rationales will sound better with "given that". Do the same with lower-level details that the kairetic theory counts as non-difference-makers, hence as objectively irrelevant, and the rationales will sound better with "because". That is just what this paper's distinction between objective and contextual irrelevance predicts: in the former case, the lower-level facts do something explanatorily important, but their contribution is frameworked; in the latter case, the lower-level facts are explanatorily otiose. The antireductionist who collapses the two kinds of relevance can make no sense of it.

5.2 The Argument from Incredulity The antireductionist argument from incredulity goes like this: I want to explain the Broad Street cholera epidemic. I see how some of the details are important. But explanatory reductionism implies that a description of the process at the level of quantum mechanics—particle by particle, potential by potential—will somehow help me to understand what happened. How can that possibly be? It seems crazy…

Such arguments make two mistakes (intentionally, perhaps). First, by emphasizing minutiae—"particle by particle, potential by potential"—they presume that a physical description is necessarily a fully concrete description, that to go physical is to relinquish abstraction. But that is not the case. Even an incontestably physical notion such as center of mass is highly abstract and therefore extraordinarily multiply realizable: a galaxy, a gorilla, and a gallium atom may have exactly the same center of mass. The kairetic account pushes for abstraction wherever it is possible, while insisting that the tie to physical implementation is preserved. It is quite possible to have plenty of both. "Physical" does not entail "detailed", and most kairetically complete explanations of high-level phenomena, though physical, are not detailed at all.

Resist, I am tempted to say, every antireductionist argument that avidly deploys the word "details". The strategy is to put before your mind as paradigms of physicality objectively irrelevant minutiae, and to insinuate that the reductionist is committed to giving such trifles an explanatory role. Not so. (Kitcher uses the word throughout his writing on the topic: "In neither case are the molecular details relevant. Indeed, adding those details would only disguise the relevant factor" (Kitcher 1984, 348). The last use in that paper (p. 370) is a resounding "gory details".)

Second, the argument from incredulity trades on the sheer psychological impossibility of our entertaining even a highly abstract quantum-mechanical description of a high-level phenomenon, running together a failure to comprehend the description with a failure of the description to explain. (Kitcher uses this strategy in the passage just quoted, but then acknowledges the reductionist riposte and puts it aside (Kitcher 1984, 348).) Attempting to grasp in its entirety a complete explanation for something like independent assortment in sexually reproducing life on earth—the explanation assembled in the scientific congress at the end of time from the contributions of geneticists, cytologists, molecular biologists, chemists, and physicists—in an attempt to eyeball its explanatory value is simply not something we are capable of doing.

What we can do without cerebral overload is to ask whether it is worth our explanatory while to trace particular lines of implementation down to their physical foundation. Do we understand independent assortment better when we grasp the molecular-level reasons for chromosomes' lack of selective bias toward the spindle's poles? Do we understand it better still when we grasp the physical reasons for the symmetries of intermolecular forces that underlie the lack of bias? In this almost purely defensive essay I have not argued the case for an affirmative, reductionist answer to these questions, but the issue is surely a live one; the affirmative answer is both reasonable and plausible.

I will finish with another version of the antireductionist argument from incredulity, a one-liner by Jerry Fodor (1974):

> What is interesting about monetary exchanges is surely not their commonalities under *physical* description. (103–104)

A sensible reductionism does not (as Fodor implies) deny the importance of a high-level, functional characterization of monetary exchanges. But it asserts the interest *in addition* of the reasons why such exchanges instantiate the patterns that they do—of the psychological principles that drive spending and saving, borrowing and lending, gambling and charitable giving, and then in turn of the neurological reasons that the psychology takes the form it does. Most of the "gory details" of the neurology will be objectively irrelevant, but some abstract neural facts will genuinely contribute to our understanding of an economy's monetary ebb and flow—that is the reductionist's explanatory bet.

Acknowledgments

For their philosophical insights and critique, thanks to Laura Franklin-Hall, Dmitri Gallow, Angela Potochnik, Alex Rosenberg, Kyle Stanford, Brad Weslake and the audiences at the Lorentz Center 2010 workshop *Understanding and the Aims of Science* and the University of Cologne 2010 workshop *Types of Explanation in the Special Sciences*.

References

Dowe, P. (2000). *Physical Causation*. Cambridge University Press, Cambridge.

Dupré, J. (1993). *The Disorder of Things*. Harvard University Press, Cambridge, MA.

Edwards, A. W. F. (1998). Natural selection and the sex ratio: Fisher's sources. *American Naturalist* 151:564–569.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.

Fodor, J. A. (1974). Special sciences. *Synthese* 28:97–115.

Franklin-Hall, L. R. (Forthcoming). The causal economy account of scientific explanation. In C. K. Waters and J. Woodward (eds.), *Causation and Explanation in Biology*, volume 20 of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, Minneapolis.

Hempel, C. G. and P. Oppenheim. (1948). Studies in the logic of explanation. *Philosophy of Science* 15:135–175.

Holling, C. S. (1959). The components of predation as revealed by a study of small-mammal predation of the European pine sawfly. *Canadian Entomologist* 91:293–320.

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

———. (1984). 1953 and all that: A tale of two sciences. *Philosophical Review* 93:335–373.

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

———. (1990). The division of cognitive labor. *Journal of Philosophy* 87:5–21.

———. (1993). *The Advancement of Science*. Oxford University Press, Oxford.

———. (1999). The hegemony of molecular biology. *Biology and Philosophy* 14:195–210.

Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy* 61:343–377.

———. (2000). Causation as influence. *Journal of Philosophy* 97:182–197.

Lotka, A. J. (1956). *Elements of Mathematical Biology*. Second edition. Dover Publications, New York. First edition published in 1925 as *Elements of Physical Biology*.

Nagel, E. (1979). *The Structure of Science*. Hackett, Indianapolis, IN.

Oppenheim, P. and H. Putnam. (1958). Unity of science as a working hypothesis. In H. Feigl, M. Scriven, and G. Maxwell (eds.), *Concepts, Theories, and the Mind-Body Problem*, volume 2 of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, Minneapolis.

Potochnik, A. (2010). Explanatory independence and epistemic interdependence: A case study of the optimality approach. *British Journal for the Philosophy of Science* 61:213–233.

Quine, W. V. O. (1969). Natural kinds. In *Ontological Relativity and Other Essays*. Columbia University Press, New York.

Reichenbach, H. (1956). *The Direction of Time*. University of California Press, Berkeley, CA.

Roughgarden, J. (1979). *Theory of Population Genetics and Evolutionary Ecology*. Macmillan, New York.

Salmon, W. C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press, Princeton, NJ.

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