Reductive Explanation

A Functional Account

Philosophical discussions of reduction seem at odds or unsettled on a number of questions and concerns:

1. Is reduction a relation between real or between reconstructed theories, and if the latter, how much reconstruction is appropriate (Ruse, 1971; Hull, 1974)? Or is reduction best construed as a relation between theories at all (Maull, 1974; Wimsatt, 1976a)?
2. Is reduction primarily connected with theory succession, with theoretical explanation, or with both (Nickles, 1973; Wimsatt, 1976a)?
3. Is translatability *in principle* sufficient, or must we have the translations in hand, and if the former, how do we judge the possibility of translation when we don't have one?[1]
4. What is the point of defending the formal model of reduction if it doesn't actually happen (Hull, 1974; Ruse, 1971), or if the defense has the consequence that if reductions occur, they are trivial and uninformative (Hull, 1974), or merely incidental consequences of the purposeful activity of the scientist *qua* scientist in devising explanations (Schaffner, 1974b)?
5. At least in biology, most scientists see their work as explaining types of phenomena by discovering mechanisms, rather than explaining theories by deriving them from or reducing them to

other theories, and *this* is seen by them as reduction, or as integrally tied to it.[2]

6. None of the philosophers currently writing on this topic are suggesting inadequacies in the kinds[3] of mechanisms postulated by molecular geneticists for the explanation of more macroscopic genetic phenomena.
7. Nonetheless, Ruse (in 1971, though no longer) and Hull (1974) seem to suggest that there is no reduction (only a replacement), and Schaffner (1974b) suggests that a reduction is occurring, but is a merely incidental consequence of the activity of these scientists.

What possibly can explain this wide disagreement between scientists who appear to take reductive explanation seriously and to regard it as perhaps *the* important consequence of their work, and philosophers who are attempting to faithfully characterize their activity and its rationale? Can reduction be as unimportant (or nonexistent) in science as these philosophers seem to suggest? I think the answer must be "no," and that there are four main factors that are responsible for the present philosophical confusion on this point:

1. Philosophers have taken the "linguistic turn" and talk about relations between linguistic entities, whereas biologists are more frequently unabashed (or sometimes abashed) realists, and talk about mechanisms, causal relations, and phenomena. Though not necessarily vicious, I think that the linguistic move has lead philosophers astray. Here I defend a realistic account of reduction.[4]
2. While virtually everyone agrees that a philosopher, by the nature of his task, must be interested in doing some rational reconstruction, doing so serves different ends in different contexts. A failure to distinguish these ends and how they may be served contributes to the apparent defensibility of the formal model of reduction.
3. No real competitor to the formalistic (or more generally structuralist) account of reduction has been forthcoming. Therefore, there has been a tendency to regard *informal reductions* [5] as either nonreductions or as *deficient* reductions, which can be remedied by becoming formalized. I will outline some aspects of a functional account of reduction that suggests that informal re-

ductions are the proper end of scientific analyses aiming at reductive explanations.

4. An emphasis on structural (deductive, formal, logical) similarities has led to a lumping of cases of theory succession with cases of theoretical explanation, with the result that discussions of reduction, replacement, identification, and explanation (which have radically different significances in the two contexts) have become thoroughly muddled.[6] A functional account of these activities yields important clarifications of their nature.

I wish to say something more about item 2, before turning to my analysis of reduction, which concerns primarily item 3 and 4. The first point enters mainly by implication.

Two Kinds of Rational Reconstruction

There are at least two (and probably more) contexts where talk of rational reconstruction seems appropriate in connection with plausible and useful activities of philosophers of science.

Rational$_1$: An Optimal Strategy

One might want to abstract from the often-irrelevant details and sometimes mistaken moves of the actual practice of science to reconstruct the significant patterns of scientific activity.[7] Insofar as these patterns can be claimed to be a relatively efficient, or even an *optimal,* way of achieving or trying to achieve the ends of such activity, the reconstruction could claim to be a rational reconstruction in the sense of rational decision theory—that it represented the way one ought to do that activity. As such the philosopher of science is a *therapist with respect to scientific strategy.*

Rational$_2$: A Canon of Logical Rigor

A physicist (and nowadays with increasing frequency, a biologist) might ask a mathematician for formal help. He might wish to prove a mathematical conjecture whose truth or falsity he is uncertain of and that has important implications for his work. Or he may have an argument that he can formulate more informally, but desires more rigor either to buttress the argument or to determine more precisely the conditions under which it holds. As such, a mathematician is a *therapist with respect to formal argument,* logic, and critical thinking. These are also

roles that could legitimately and usefully be played by a philosopher of science.

In either case the philosopher of science would be analyzing or criticizing an activity in terms of how well it served the ends of the scientist, and in each case, the activity itself and the analysis of it further these ends.

Note that the functions of the philosopher of science in these two cases are, at least *prima facie*, not equivalent. It is not at all clear that improvements in rigor, per se, are a rational and efficient way to do science—say, for finding explanations—nor even that the ultimate end state of science will be to improve the rigor of theories *that are otherwise adequate* (i.e., after their other problems have been solved). Improvements in rigor are sometimes useful, but not always. Philosophers of science have sometimes talked as if improvement in rigor is a scientific-end-in-itself, but no one here is doing so. I believe that the sort of confirmation and troubleshooting suggested above is the main function of rigorous argument in science, and that rigor is not a scientific-end-in-itself.

One effect of logical empiricism (with its emphasis on the logical structure of laws, theories, explanations, predictions, and experiments) has been to blur—even to obliterate—the distinction between these two senses of *rational reconstruction.* This conflation has had a disastrous effect upon the analysis of reduction, proceeding as it has in terms of the formal model. Schaffner's (1974b) thesis of the "peripherality" of reduction suggests that any successful defense of the formal model would win a pyrrhic victory. In terms of the above distinctions, I would describe this peripherality of the formal model as follows: It is not rational$_1$ to view formal (i.e., rational$_2$) reduction as a scientific-end-in-itself because science then becomes an inefficient and ineffective way of pursuing known scientific ends (such as explanation). And although the formal model of reduction is by definition a rational$_2$ model, it is not even an effective *means* to some end because it is not the answer to a request for formal (i.e., rational$_2$) assistance that anyone has made or would be likely to make! Thus, although early discussions of formal reduction seemed to hold out the hope that it would perform the functions of both kinds of rational criticism, it is my impression that more recent sophisticated discussions (such as Schaffner's) have given up on both claims. But these claims are not peripheral and readily dispensable. They represent one of the major motivations for pursuing either a formalistic or a reductionist strategy in science. If they must be given

up, one's claim to be analyzing reduction as that concept is used in science must be suspect.

Paradoxically, if a non-formal (or perhaps, partially formal) account of reduction is allowed, it can be seen to be a rational activity in both senses: It is an efficient (rational$_1$) way in which to proceed, and it proceeds by using logical instruments for the critical (rational$_2$) evaluation of theoretical and observational claims. Because it is a partially formal model, the use of formal methods (as discussed by Schaffner and Ruse) is to be expected on this model also, and it derives confirmation from the cases they adduce to support the formal model. It does not require total systematization, however, which has *not* been exemplified in any of the cases they discuss and which formal reduction requires (see, e.g., Schaffner, 1976, p. 614).

How do we get such an alternative to the formal model of reduction? Just as a characterization of logical *structure* (a rational$_2$ reconstruction) suggests and is suggested by a formal model of reduction, the view of scientific activity as purposive suggests *a functional* analysis and characterization—a rational$_1$ reconstruction—of reduction. Such an analysis may distinguish between activities having similar structure in some respects,[8] while pointing to and explaining further structural differences that are ignored on the formal approach. Most importantly, a functionalist approach shows why the research aims of the scientist *contribute to* (in the sense of moving in the direction of) fulfilling the aims of the formal model, but are in fact *different from* and even *inconsistent with* actually getting there. Then a stronger version of Schaffner's (1974b) peripherality thesis is justified:

(P1) Not only is progress toward formal reduction incidental, but

(P2) It also seems to be epiphenomenal, since this progress toward formal reduction appears to have no *further* consequences.

(P3) Finally, if (as I believe) getting there is inconsistent with the real aims of science, this "progress" is bound to remain incomplete.

Successional versus Inter-Level Reductions

The functional viewpoint is perhaps best developed by expanding upon and modifying Schaffner's (1967) model, which has many useful features, though the end result will be quite different (see figures 11.1 and

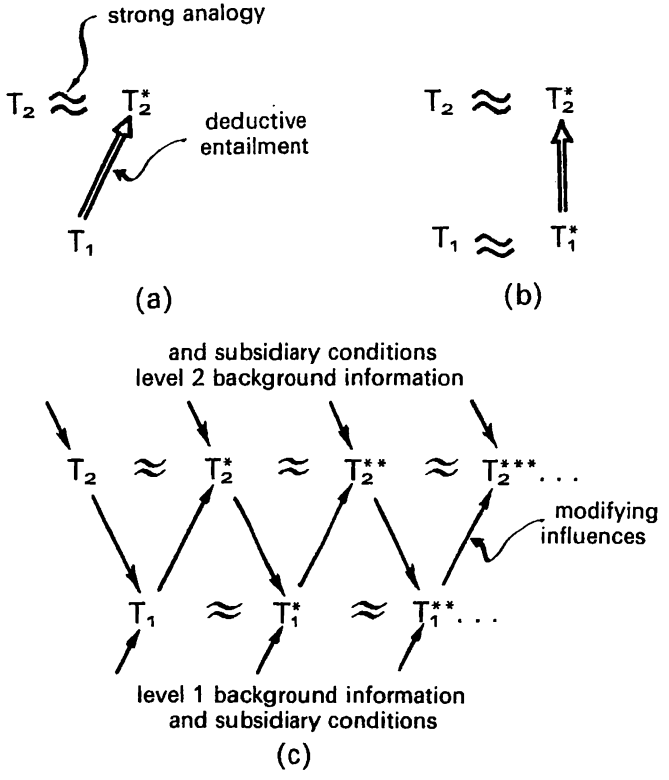


Figure 11.1. (a) Theory reduction, from Schaffner (1967). T_2: reduced theory; T_2^*: corrected reduced theory; T_1: reducing theory. (b) Theory reduction, from Schaffner (1969). T_1^*: modified (corrected?) reducing theory. (c) Co-evolution of theories at different levels, from an early draft of 1976a).

11.3). Most importantly, Schaffner distinguished between and included both a derivability condition between the reducing theory (T_1), and a corrected version of the reduced theory (T_2^*), and a condition of strong analogy between T_2^* and its uncorrected predecessor, T_2. These two relations are prototypic of two distinct relationships, each of which has been called *reduction.*

Schaffner's condition of strong analogy is closely related to Nickles' "reduction$_2$" (Nickles, 1973, pp. 194ff.) and to what I elsewhere (Wimsatt, 1976a) and below call *successional* or *intra-level* reduction. Nickles' account, emphasizing transformational and possibly non-deductive relations between successive competing theories affords an important partial explication of "strong analogy." A functional ac-

count of this activity explains many of the structural features Nickles proposes, and others that he does not mention.

What is not clear on Schaffner's model, but implicit in Nickles' is that reduction$_2$ (which is a kind of "pattern matching" problem and could also be regarded as *demonstrating and analyzing* the "strong analogy" between T_2 and T_2*)[9] is neither automatic nor self-evident. It has a point, involves work, and is performed for reasons separate from the functions of the "other" reductive relation. Nickles suggests that reduction$_2$ performs heuristic and justificatory functions vis-à-vis the uncorrected older T_2.[10]

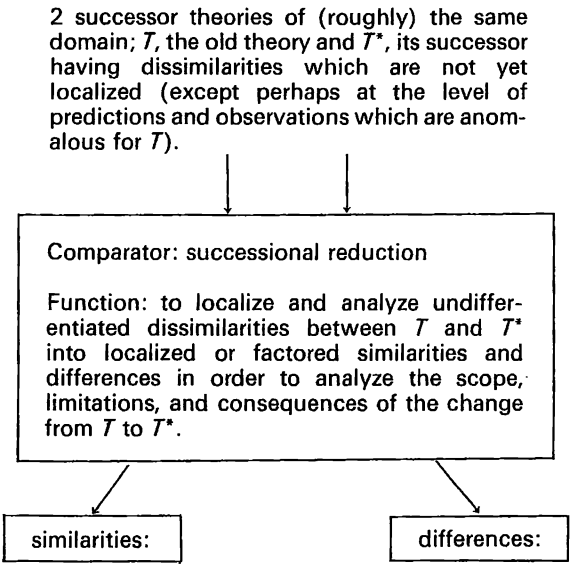
I believe that reduction$_2$ is fundamentally connected with theory succession (of T_2 by T_2*) and performs rather more functions than Nickles makes out. *It is most immediately a transformational operation whose function is to localize and analyze the similarities and differences between T_2 and T_2** that in turn serve a variety of further functions. Most interestingly, because none of these functions are served by making comparisons other than between T_2* and its immediate predecessor, T_2, and in any case, similarities and differences become *less* localizable as changes accumulate, successional reduction would be expected to be *intransitive,* and to behave as a similarity relation.[11] *Thus the intransitivity of successional reduction is an explicable feature, not a given, on the functional account of this activity.*

For further analysis of the specific uses made of these localized similarities and differences between T_2 and T_2* and diagrammed in Figure 11.2 (see part II of Wimsatt, 1976a). However, the following contrasts between "successional" and "explanatory" reductions are noted here.

1. *Successional reduction is* and must be *a relation between theories* (since it is these that exhibit the similarities and differences), unlike *explanatory reduction* which *is not,* in any but degenerately simple cases.

2. *Replacement* occurs only with *the failure* of successional reduction—failure to localize similarities and differences among successive competing theories. Replacement and successional reduction are opposites. But for explanatory reductions, replaceability is closer to and is by many treated as a *synonym* for reducibility. A failure of T_1 to reduce T_2 (perhaps derivatively, by reducing T_2*) would make T_2 and its successors *emergent* and *irreplaceable* relative to T_1. *Replacement obviously has two different meanings here.*

3. *Successional reductions are intransitive.* A number of them "add up" to a replacement. *Explanatory reductions are transitive.* (It is this last fact that raised the hopes among advocates of "unity of science" for



Functions:

(1) give prepackaged confirmation of T^* by showing that it generates T as a special case (Nickles, 1973).

(2) "explain away" old T, or explains why we were tempted to believe it (Sklar, 1967).

(3) delimit acceptable conditions for use of T as a heuristic device, by determining conditions of approximation (Nickles, 1973).

Functions:

(1) explain facts which were anomalous on T, thus confirming T^*.

(2) suggest new predictive tests of T^*.

(3) suggest reanalysis of data apparently supporting T and not T^*.

(4) suggest new directions for elaboration of T^*.

Figure 11.2. Functions of similarities and differences in successional reduction.

Nickles (1973) also suggested that "reductions$_2$" may be done in various ways. This makes sense if the point of the transformation is how best to factor out similarities and differences.

Successional reductions may be possible "locally" (for parts of theories) even when not possible globally (for the whole theory).

Differences in meaning among key terms may be regarded as irrelevant, so long as they are localizable to allow fixing praise and blame on specific components of T and T^* in comparatively evaluating them. See also Glymour (1975). Thus, the "meaning change" objection is avoidable.

great ontological economies through reduction, about which I have more to say below. See also Wimsatt, 1976a.)

4. Talk about elimination might be appropriate for the posited entities of corrected and replaced theories if the new theory is sufficiently different that there is no significant continuity between old and new entities. But such talk is frequently illegitimately extended to contexts of *explanatory reduction.* This is often motivated by talk of ontological or postulational simplicity in the light of supposed translatability and deducibility (discussed further below), but in at least some cases looks suspiciously like treating reduction and replacement as opposites. Thus, in arguing that the formal model of reduction doesn't fit the relation of Mendelian to molecular genetics, Hull and Ruse[12] each suggest that it looks more like a case of replacement. As I suggested in item 2 above, the opposition between reduction and replacement is appropriate for successional reduction, but *not* for interlevel or explanatory reduction. Their claim is thus misplaced if it concerns the relation between T_1 and T_2. Though intelligible if construed as concerning the relation between T_2 and T_2*, I would disagree on the facts of the case, and agree with Schaffner (1976) and Ruse's (1976) more recent view that there is no replacement, but a reduction. To explain why, I must say a great deal more about explanatory reductions to which I now turn.

Levels of Organization and the Co-Evolution and Development of Inter-Level Theories

Rather than talking directly about reductive relations between theories, the approach I have taken (Wimsatt, 1976a) is the realistic one of regarding levels of organization—features of the world—as primary, and defined in such a way that it is natural that theories should be about entities at these levels of organization. The notion of a level implies a partial ordering, such that higher level entities are composed of lower level entities. In a universe where reductionism is a good research strategy, the properties of higher-level entities are predominantly best explained in terms of the properties and interrelations of lower-level entities.

But I argue further that levels of organization are primarily characterized as local maxima of regularity and predictability in the phase space of different modes of organization of matter. Given this, selection forces (and at lower levels, the stability considerations into which these shade) suggest that the majority of readily definable entities will be found in the (phase space) neighborhood of levels of organization, and

that the simplest and most powerful theories will be about entities at these levels.[13]

Nothing in this approach entails that levels defined as local maxima of regularity and predictability must always be well-defined and delineated, or strictly linearly orderable (although they usually are for simpler systems), and certain conditions can be suggested (in *this* world) where these assumptions are false (see Chapter 9 and Wimsatt, 1976a, part III). These are conditions where neat composition relations cannot be specified for all (or perhaps even for any) of the entities in these different *perspectives*. (Level talk *requires* the possibility of specifying composition relations, so I talk about perspectives when this condition is not met.) This failure of orderability leads to the "intertwining" of theories mentioned by Schaffner (1974b) in discussing the operon model (see also Schaffner, 1974a) in support of his thesis of the "peripherality of reduction," and to the much more extreme situation suggested by Maull (1974) in her penetrating analysis of the same case—which she sees as the development of an inter–(multi-)level theory rather than the tying or merging together of preexisting theories.

These sorts of complexities are ignored in discussions of the standard model of reduction, and Hull's (1974) discussions of the difficulties of translation just begin to characterize one of their major effects. Nor is this problem limited to genetics. Fodor's (1974) discussion supports the view that the standard model is of substantially greater scope and provides a careful analysis of problems that arise for the standard ("type reduction") account of reduction in these areas. But the standard model just looks so right that it is hard to see how it *could* be wrong. In this light, claims like those of Hull and Fodor seem almost counterintuitive, and it becomes easy to give them short shrift. There are several sources of bias in favor of the standard model that contribute to this appearance:

1. There is a general tendency to characterize the lower-level theory (T_1) as "more general" and "more explanatory" than the upper-level theories (T_2 and T_2*), trading on our general reductionist prejudices in favor of using compositional information (rather than, e.g., contextual information) in an explanation. This has complex sources that I have discussed elsewhere (Wimsatt, 1976a), and has as one of its effects the tendency to assume that lower-level theories correct upper-level theories, but not conversely.[14]

2. Another important source of bias leading to this error is the distinction between contexts of justification and contexts of discovery, and the attention paid to the former at the expense of the latter. We pri-

marily worry about justifying edifices—theoretical structures that have already undergone substantial revision and selection, and that we have begun to presuppose in a variety of other areas and are thus loath to revise in any substantial way. We discover and propose models tentatively and usually without much commitment. We give them up or modify them easily because little else depends upon it. For reductions (or at least for those that look much like they will come close to satisfying the formal model), the lower-level theory is already well into the edifice stage, and it is thus not surprising that lower-level corrections are less visible, having for the most part already occurred (this is entrenchment, in the sense of Chapter 7).

3. Another bias toward the standard model is introduced via the view that explanations involve giving laws, rather than citing causal factors or giving causal mechanisms. How this is introduced (laws suggest greater systematization than do causal factors) and avoided (by accepted Salmon's account [1971] of statistical explanation) is discussed below.

4. Discussions of translatability tend to revolve around those cases where it looks easiest to give a translation, and it is often easier for properties than for objects (which are characterized by a variety of theoretically relevant properties if they are important objects). It is easier for objects if they are not functionally defined (or are fallaciously *treated* as if they were not) since function makes features of the *context* highly relevant. (As linguists know, a context-dependent translation is an incomplete translation.) Functionally defined processes can be the most difficult, since they will often be associated with a number of objects that will also be involved in *other* functional processes (see Chapter 9), and can be realized in different ways. This is the domain where the functional localization errors induced by the aggregativity biases discussed in Chapter 12 have the largest effects.

Discussions of reduction in genetics have not even approached the translation of some of these terms. Terms from population genetics like "heterosis," "additive (multiplicative, non-additive, non-multiplicative) interactions in fitness," and Lewontin's "coupling coefficient" (1974, p. 294), represent things we look for and find mechanisms for, but general or context-independent translations at a molecular level seem absurd—both impossible and pointless. Context-dependent translations are easy to come by, of course. Discovering the mechanisms in specific cases *gives us* that. But that won't do for the formal model: for those purposes a *context-dependent translation is not a translation.*

What would a new view of inter-level reduction look like?

Schaffner's (1969) modifications to T_1 in order to affect the reduction (Figure 11.1b) is a step toward the picture I would draw: *Theoretical conceptions of entities at different levels coevolve and are mutually elaborated* (particularly at places where they "touch"—where we come closest to having inter-level translations)[15] *under the pressure of one another and "outside" influences* (see Figure 11.1c). In this picture, both successional reductions (or replacements) and explanatory reductions are occurring in an intricately interwoven fashion. Very roughly, all corrections in theory get packed into a "successional" component (because Leibniz's Law applied to inter-level identities ferrets them out of the other component), and all unfalsified explanatory and compositional statements get packed into the "explanatory reduction" component. Theory at different levels progresses by piecemeal modification, in a manner paradigmatically exemplified by Maull's discussion of the operon theory (1974, chapter 2, and 1976) (see Figure 11.3 for the following discussion).

Three things should be noticed about these modifications.

1. Their form may well be deductive or quasi-deductive in character, but if so, the arguments are usually both enthymematic and riddled with *ceteris paribus* assumptions. Typically, it is decided that a T_1-level mechanism cannot accommodate a T_2-level phenomenon without modification to T_1*, in which case inferential failure of T_1 is the source of the change; or from T_1 and appropriate boundary conditions, we infer, predict, or deduce that a phenomenon that is incompatible with T_2, but not with a T_2* and observed results should occur, in which case an inferential success of T_1 and its associated mechanisms is the source of the change.

2. The modification occurs without a total deductive systematization, or often even an informal recodification of the theories. The new theories are characterized in terms of the changes from the preceding theories, but because they were similarly characterized there is hardly ever a thorough systematization.

3. The important difference of this picture from Schaffner's is that it is primarily the *changes* in theories that result from deductive arguments. Seldom if ever is any even sizeable fragment of a theory deduced wholesale from another, and seldom if ever is even a single theory sufficiently systematized to meet the conditions for applying the formal model. Furthermore, it is so clearly unnecessary and irrelevant to the search for explanations.

Schaffner's own accounts (1974a, 1974b) and that of Maull (1974) are beautiful confirmation of this highly efficient but formally highly

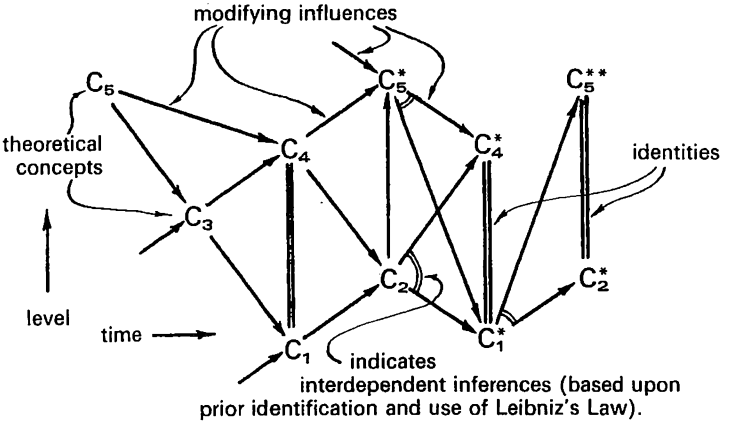


Figure 11.3. An extension of the model of Maull (1974) involving the use of identities as proposed in Wimsatt (1976a) in the co-evolution of concepts in the development of an inter-level theory of the operation of a causal mechanism. Strong analogy between concepts and their descendants ($C^n_m{}^*$, $C^n_{m^{+1}}{}^*$) is assumed generally (but not necessarily universally) to hold, but is not represented here, in order to simplify the diagram.

(a) Inference structure of the development of the theory. (b) Resultant causal structure of the mechanism according to the theory; development of an inter-level theory.

confusing strategy of theory evolution. These suggest that the vertical arrows *not* be interpreted as total entailments between theories (or reductions, where upwards arrows are concerned), but as single rough deductions or inferences from attempts to match the structure of causal mechanisms as described at different levels resulting in changes in *parts* of theories. There is, to be sure, use of deductive argument, and lower-level explanation of upper-level phenomena. The examples of Ruse (1976; hemoglobin and sickle cell anemia), Maull (1974), and Schaffner (1974a, 1974b) are marvelous. But as Hull (1974) points out, they do *not* touch the issue of whether a total deductive systematization is occurring since such cases would also be expected on the view of reduction advanced here. So, then, why should one bother to attempt to characterize reduction along the lines of the formal model? There just seems to be too big a gap between principle and practice for the principle to be very interesting.

Aside from philosophical predilections of an eliminative sort, there seem to be two reasons for holding onto the formal model of reduction:

1. The belief that as the fit gets better between upper-and lower-level theories, their relationship asymptotically approaches the conditions of the formal model of reduction.
2. The belief that even if the fit never asymptotes, or if it does, doesn't converge on the formal model, the latter represents an aim of scientists.

While Schaffner (1974b) has questioned whether trying to accomplish the reductionist program per se is a good scientific strategy, I suspect that he (and perhaps many scientists) believe that it is at least a secret hope or end. I want to examine the grounds for this latter belief, and suggest an alternative interpretation that is more consistent with scientists' actual behavior. This interpretation also raises serious questions about the first assumption.

Finally, the formal model would not be nearly as tempting if there was not, for each philosopher talking about "translating away" upper-level vocabulary, a scientist talking about "analyzing away" upper-level entities. It thus looks as if a claim about words can be "cashed in" for a claim about entities; a claim that many scientists appear to accept. This claim will be analyzed from another perspective in the next chapter—what does it mean to say that a property of the whole is "nothing more than" properties of the parts?

So the formal model appears to have direct support in the talk of many scientists of the "nothing more than" persuasion. But of what are they persuaded? Are the translations or analyses like those promised by Schaffner *immediately* forthcoming? Usually not. No one actually ground them all out, but that's said to be just a practical difficulty. It is *in principle* possible. But in principle claims have been failing, only to be replaced by new ones, since the time of Democritus. Given their history, such in principle claims could not plausibly be treated as self-warranting. But then what else warrants them? How can we evaluate these in principle claims, to distinguish good ones from bad ones? Or perhaps these in principle claims are not the claims they seem to be, to knowledge the claimant cannot have. I suggest rather that they are important tools in the task of looking for explanations. Before discussing this, I must consider the views about *explanation.*

Two Views of Explanation: Major Factors and Mechanisms versus Laws and Deductive Completeness

I accept Salmon's (1971) account of explanation as a successful search for "statistically relevant" partitions of the reference class of the event being explained, with two provisos. First, I will make some modifications (explained below and in the appendix) to bring it into line with a view of science as an activity conducted according to cost-benefit considerations. Second, I assume that in finding statistically relevant partitions, we are doing so with the aim of partitioning the reference class into *kinds of mechanisms,* or kinds of cases involving a given mechanism (I am thus giving a realist interpretation to his model). In a *reductive* explanation, these mechanisms or factors are at a lower level of organization than that of the phenomenon being explained.

One of the intriguing features of Salmon's account is his move from constructing (statistical) *laws* to a search for statistically relevant *factors.* Laws suggest the need for a complete account of the conditions under which they apply and are correct, and the connection of explanation with laws thus naturally suggests the sort of exhaustive search for factors and conditions that would go along with a complete translation of terms or a complete deductive reduction. By contrast, a search for factors (especially a search for the *major* factors—enter cost-benefit considerations!) ties in naturally with a view of explanation as a search for the mechanisms that produce a given phenomenon, and as an account of how they do it. This search stops short of an exhaustive deductive account by sticking much

of the initial and boundary conditions and many background assumptions into a *ceteris paribus* qualifier on the explanation *because they are too unimportant or insufficiently general to be accounted as part of the "mechanism."*

The deductivist or formal account *can* give superficial recognition to such differences of importance by different labeling (laws, boundary conditions, initial conditions, etc.) of different parts of the deductive basis. However, in looking first for a valid deduction, the formal account treats all such information as if it were fundamentally alike because it is all equally necessary for the deduction to go through. It thus rides roughshod over realistic intuitions as to differences in the roles and importance of these different kinds of information. Hull (1974) is sensitive to this in arguing that a single molecular mechanism can lead to different Mendelian traits, for which he has been criticized by Ruse (1976) and Schaffner (1976). Neither Hull nor I nor the scientists who would agree with us are anti-reductionists or anti-determinists. We are simply responding to widespread and reproducible intuitions as to when a change in the total state-description is counted as a change in the mechanism, and when it is not.

This judgment and its reproducibility are explicable on a combination of realistic, evolutionary, and cost-benefit considerations about the nature of scientific theorizing: A mechanism is a "kind," and cost-benefit considerations on the complexity of the theory introduce a "crossover point" beyond which a phenomenon or state is too infrequent or unimportant in a theory to be reified as a kind. There will thus be cases involving the same mechanism with different outcomes that will be attributed to differences in the (more variable and less central) initial or boundary conditions, or to violation of the nebulous *ceteris paribus* clause.

The deductivist also makes and must make such judgments of relative importance, but the baggage of having to construct a valid deduction and of having to treat the correspondences between lower and upper levels as "translations" leads to dangerous misdescriptions of what is going on in several respects.

1. It is only too easy to assume that variations in the boundary conditions are predictively negligible because they are treated as of negligible or lesser general explanatory importance. A failure to include them as part of the mechanism as Hull (1974) has done indicates the latter, but in no way implies either that the same mechanism always produces the same output, or that this failure indicates that the same

total state of the system is on different occasions yielding different outcomes. These are mistaken interpretations that become tempting when Hull's discussion of mechanisms is read as if it were about state-descriptions, and when the only differences of importance are assumed to be differences of deducibility or predictability.

2. Schaffner's claim (1976, pp. 624–625) that Hull's discussion of mechanisms misconstrues the logic of the formal model is double-edged. He would in effect substitute talk about state descriptions. But if the scientists are interested in *mechanisms* and Hull's point is defensible in terms of the way we investigate and reason about mechanisms (as I think are so), of what relevance is Schaffner's probably correct claim that the formal model is defensible if we translate from talk about mechanisms to talk about state descriptions? If scientists aren't interested in state descriptions, Schaffner has apparently defended the formal correctness of his model at the cost of showing its irrelevance to how scientists talk and reason about reduction. Schaffner's claim about the peripherality of reduction begins to look more and more as if it applies more modestly and correctly to *the formal model of reduction.*

3. An equally dangerous move accompanies Schaffner's account of the relation between micro- and macro-descriptions as "translation." Schaffner (1976, p. 630n 25) *assumes* the constancy of the environment and unstated initial and boundary conditions over a range of different cases in constructing his "translation" for the dominance relation. This is done "for reasons of simplicity and logical clarity" (ibid.). But while this is an appropriate defense of simplifying assumptions in a model or idealization, it is not an appropriate move in defense of a "translation" that is to be used in the way that his are. Thus *one thing his assumption does is to mask the real context-dependence of his "translation" by artificially assuming that the context in constant!* But if one is trying to establish that context-independent translations can be given (a necessary move if one is to use these translations as general premises in a deduction over a range of cases in which the context changes), this move is to beg the question; it is to hide deductive incompleteness by trading it for translational incorrectness or equivocation. Schaffner *cannot* do so (see Schaffner, 1976, pp. 622–623).

Schaffner would not assume this constancy if it were admitted or discovered that there were an important variable (or "part of the mechanism") contained in that set of things assumed constant. He would then attempt to delineate that variable, and include it in the translation. Thus the boundary between what is in the translation and what is as-

sumed constant is fixed by the same judgments of importance used in delineating "mechanism" from "background" on the model that I (and I believe Hull) would defend. But what is not in the translation (or mechanism) is not thereby constant. It is quite variable, in fact, and *its very variability is one of the reasons for not including a detailed specification for it in the general theoretical account.* Its variability makes it unimportant for theory construction, and often for selection as well,[16] though it can often produce divergent predictive results and frustrate attempts at translation.

Although Salmon is probably not considered to be a scientific realist, his account of scientific explanation is a natural ally of realistic accounts of science because of its natural structural affinities for such explanations in terms of major factors and mechanisms, in general, and lower-level mechanisms in the case of reductive explanations (see Shimony, 1971; Boyd, 1973, 1980; Campbell, 1974a, 1974b; Wimsatt, 1976a).[17]

Levels of Organization and Explanatory Costs and Benefits

Suppose that the primary aim of science and of inter-level reduction is explanation. We wish to be able to explain every phenomenon under every informative description by showing, first if possible, how it is a product of causal interactions at its own level, but barring that, how it is a product of causal interactions at lower levels (a micro-level or reductive explanation), or least probably and desirably in our reductionist conceptual scheme (but absolutely unavoidably in a world of evolution driven by selection processes), how it is a product of causal interactions at higher levels (most commonly, a functional explanation).

This order of priorities in the search for an explanation follows naturally from the account of levels as local maxima of regularity and predictability, together with acceptance of a weakly but generically reductionist worldview, and the assumption that the *search* for explanatory factors is also conducted according to some sort of efficiency optimizing or cost-benefit considerations. The rationale for this is discussed more fully in Wimsatt (1976a) and is roughly as follows:

1. The characterization of levels of organization as local maxima of regularity and predictability implies that most entities will most probably interact most strongly with (and most phenomena will be most probably explained in terms of) other entities and phenomena at the same level.

2. A reductionist conceptual scheme (or world) is at least one in which when explanations are not forthcoming in terms of other same level entities and phenomena, one is more likely to look for (or find) an explanation in terms of lower-level phenomena and entities than in terms of higher-level phenomena and entities.
3. If a search for explanatory factors is conducted along some such principle as "Look in the most likely place first, and then in other places in the order of their likelihoods of yielding an explanation," then the above order of priorities is established.[18]

Salmon's account (1971) of explanation will be generally presupposed here, but with a cost-benefit clause added to it: not only are "statistically irrelevant" partitions products of a choice of explanatorily irrelevant variables (as he points out), but "statistically negligible" partitions are similarly products of explanatory negligible variables. This change is consonant with the remarks of the preceding section on recognizing the different roles and importance of mechanisms, boundary conditions, and the like in an explanation, but also has important further ramifications. Most crucially the intuitive sense of what it is for one variable to "screen off" another changes (as described in the appendix).

The idea that there can be explanatorily negligible partitions of the reference class of the event or phenomenon being explained suggests an asymmetry of explanatory strategy for cases that do and cases that do not meet macroscopic regularities or laws. When a macro-regularity has relatively few exceptions, redescribing a phenomenon that *meets* the macro-regularity in terms of an *exact* micro-regularity provides no (or negligibly) further explanation. All (or most) of the explanatory power of the lower-level description is "screened off" (Salmon, 1971, p. 55, but see the appendix below) by the success of the macro-regularity. The situation is different, however, for cases that are anomalies for or exceptions to the upper-level regularities. Since an anomaly does not meet the macro-regularity, the macro-regularity *cannot* screen off the micro-level variables. If the class of macro-level cases within which exceptions occur is significantly non-homogeneous when described in micro-level terms, *then* going to a lower-level description can be significantly explanatory, in that it may be possible to find a micro-level description partitioning the cases into exceptional and non-exceptional ones at the macro-level. We would then have a micro-explanation for the deviant phenomenon.

Thus, for example, the ideal gas law (or its corrected phenomenolog-

ical successor), as a relationship between macroscopic causal factors, is explanation enough for occasions when gases obey it. Going to the micro-level in such a case is not (or negligibly) more explanatory. Of course, if all of the molecules go to one corner of the container, the micro-level must be invoked since the macro-level law does *not* apply, and in that case partitions in terms of micro-variables will be statistically relevant.

So one reason to look for information at lower levels is to explain exceptional cases at the upper level. The other main reason is to explain upper-level regularities. But part of explaining exceptional cases involves explaining why they are exceptional in a way that is consistent with the patterns found in the motley of cases explained by the upper-level law (*qua* set of interrelated causal factors). This usually involves explaining exceptional and motley cases in terms of a single class of mechanisms or micro-variables. This requires that the relevant kinds of micro-descriptions necessary to explain the exceptional cases *also* be usable in generating the upper law as a "special case" or "limiting" or "approximate" description. It thus leads to an explanation of a revised version of the upper-level law.[19]

But what is a law, and why bother to explain one if, as I have argued, mechanisms and major factors bear the primary role in explanations of events that laws have been thought to do? The answer that suggests itself in the cases I have looked at where laws are being explained in terms of lower-level factors and mechanisms is that *laws are regularities involving distributions of cases characterized at the macro level.* They are explained as the product of the interaction of the mechanisms and major factors invoked at the micro-level with the micro-level distributions of initial and boundary conditions. They are not *mere* regularities (or "accidental generalizations" as Nagel [1961] characterizes the infirm statement of law-like form) because they are exhibited as the product of *causal* interactions of micro-level mechanisms, factors, and initial and boundary conditions. Such law-statements thus support the appropriate counterfactual and subjunctive conditionals. Indeed, when a macro-regularity is explained in this manner, an understanding of the micro-level mechanisms and conditions that generate the macro-level distribution and how they do so give a much richer structure of counterfactuals expressible in terms of micro-descriptions than before.

I am not sure whether this characterization of a law is generalizable. It might seem limited to cases where the phenomena of a law admit of meaningful redescription at a lower level. But, at least in those cases

where this characterization applies, and this would appear to cover all cases of (inter-level) reductive explanation, a law should be explicable in the same general way as an event. The only difference would be that instead of talking about individual constellations of mechanisms, factors, and conditions, we are talking about assumed *distributions* of the above.

The reduction of thermodynamics to statistical mechanics would provide useful examples of explanations of this sort (see, e.g., the much discussed explanations of the second law of thermodynamics). But so also would the history of the assumption of the "purity of the gametes in the heterozygote" that Hull (1974, 1976) makes much of in arguing that molecular genetics replaces, rather than reduces, Mendelian genetics. I believe that Hull is incorrect in his conclusion, and that an illustration of how this "law" is explained reductively helps us to see how much real continuity there is between Mendelian and molecular genetics.

An Example: The Assumption of "the Purity of the Gametes" in the Heterozygote

This assumption began life as Mendel's (1866/1956) "law of segregation"—to explain the fact that some apparently lost characters ("recessives") reappeared apparently unchanged in successive generations. Mendel's explanation was that in the company of certain alleles ("dominants") the factors did not express themselves as characters, *but that they were transmitted to offspring unchanged (by their allelic factors or anything else)* to express themselves in future genotypes in which they were homozygous or dominant.[20]

In the Mendelism that Castle attacked (Castle and Phillips, 1914), with his belief that the allelic genes "contaminated" one another in the heterozygous state, it was accepted that genes affecting a given character came in pairs (were alleles), but Mendel's other law—of "independent assortment" (that non-allelic genes assorted independently of one another in the offspring)—was being challenged in the early years of the twentieth century, both experimentally and theoretically, by Bateson and others, including Morgan and his students.

The "linear linkage" model of the Morgan school explained some of Castle's results (gradual changes in coat color conformation in rats) by the gradual accumulation through selection of so-called modifier genes at *other* loci (presumably linked on the same chromosome) that modified the *effect* of the genes identified as producing coat color, *without modifying the allelic genes themselves.* There was thus no need to sup-

pose (in this case) that allelic genes "contaminated" one another in the heterozygous state. Castle's supporting claim that these modifications were irreversible was successfully contested experimentally (see Carlson, 1966; Wimsatt, 1992).

The Morgan model supposed that the genes were linearly arranged on chromosomes, with allelic genes on corresponding places on the homologous paired chromosomes. According to this model, homologous chromosomes would, at a certain part of the cell cycle, wind around one another forming "chiasmata," break, and exchange segments. This was called *crossing-over* and *recombination.* A central feature of the model was that genes on the same chromosome would tend to assort together, constituting linkage groups. This was in contradiction to Mendel's law of independent assortment. A prediction of the linear model and the mechanisms of recombination was that the probability of recombination between two points along the chromosome was a monotonic increasing function of the distance between points (being approximately linear for small distances and approaching 50 percent [or random assortment] for large distances).[21] These also were experimentally confirmed. Furthermore, *in the absence of any "atomistic" assumptions* (placing a lower bound on minimum distance between recombinations), *this model would predict a finite frequency for crossing-over within genes of any finite size.*

A gene has a size, and members of the Morgan school recognized this, though different ways of estimating it produced different results (Carlson, 1966, pp. 83, 85, 158ff. reviews this and the other issues of this paragraph; see also Wimsatt, 1992). Although it was usually assumed that the genes behaved like "beads-on-a-string" (or independent atoms) as far as recombination was concerned, Muller, a Morgan student, questioned whether these "atoms" were the same for recombinational and for mutational events. Other observed phenomena (like "position effect") also raised questions about the beads-on-a-string model. It also was generally supposed that genes had an underlying molecular nature, though it was unknown what this was, and how it produced the properties manifested by genes, so the idea that genes had a molecular infrastructure was not new. Indeed, the atomicity of the genes was clearly believed, to the extent that it was only with respect to the genetic or biological properties of the genes.

The details of how the molecular account of the gene explain "position effect" and the possibility of differences between recombinational, functional, and mutational criteria for individuating genes are well

known (see, e.g., Hull, 1974, or any modern genetics text) and uncontroversial here. All of these have the effect of compromising the view of genes as monolithic, monadic atoms with respect to some of their biological properties. If there are any "atomic" units of DNA, it is the individual base pair—again not because smaller changes are impossible, but because if they occur, they are not counted as *genetic* changes. While this would show that there were no "atomic genes" of the size Morgan and his school had assumed, and that their different criteria of individuation picked out *different* larger compound assemblages of bases as genes, it is not necessarily a disproof of their genetic "atomism." It could just as well be taken as a demonstration that their atoms were smaller than they had thought (see note 24) and (being of at that time unknown constitution) had some unexpected properties that explained others that the genes had been thought to have.

How does the assumption of the "purity of the genes in the heterozygote" fare? This becomes a question of the possibility of intra-genic recombination—but not a simple question: we must ask not only what happens, but also, what an experiment detects. We can now explain in terms of the design of the recombination experiment why, even if they should occur readily, it was very difficult to find intra-genic cross-overs and recombinations. We can do this in terms of the molecularly characterized gene, *but there is no need to do so.* Morgan could have done so himself, as it is an obvious consequence of the classical model of the genome.

1. On this model, there were a large number of genes on each chromosome. Muller estimated in 1919 that there were at least 500 genes on the X chromosome in *Drosophila,* and we now know that to have been at least a four-fold underestimate.[22]
2. It was taken as a given then as now that any individual gene has a very high stability, which would have applied either to intra-genic recombination or to any other mutational event.
3. The design of a recombination experiment involved looking at a small number of marker genes spaced along the chromosome in order to see how frequently they (or more accurately, the traits that signal their presence) stay together in offspring. The usual number of marker genes was two, though Sturtevant occasionally used three and four to detect multiple crossing-over.[23] Supposing even that one could detect any intra-genic recombination occurring in any of the marker genes (see item 4), the very small fraction of the genome being used as marker genes renders it very

probable that recombinational events will not occur in any of the markers, but will occur elsewhere along the chromosome, separating whole the marker genes on either side of the break.

4. We now know that intra-genic recombination would produce a non-functioning gene. This would have been scored by the Morgan school as a "loss" or "mutation" of a gene, rather than as an intra-genic recombination, so they probably did *not* detect any such events that did occur. (Only with later work on intra-cistronic complementation were the classical techniques sufficiently refined to detect such intra-genic events. But it is worth emphasizing that the problem was a technical one, and not a conceptual one for the classical approach.)

The net effect of this is twofold:

1. The classical model itself predicts that if genes are as small and as numerous as they had to be (and they were smaller and more numerous), intra-genic recombination would be hard or impossible to detect; even if virtually all recombinational events were intra-genic.
2. What was *seen* in recombination experiments was whole (marker) genes separating from one another untouched.

The first fact might have produced caution. It did not. The second observation led to an extrapolated assumption that recombination occurred *between genes, generally,* rather than just *between the observed genes.* But the first fact means that the new molecular picture is *not* that different from the old model. By analogy with the old model:

1. Crossing-over should be a monotonic increasing function of the length of the DNA involved.
2. The probability of crossing-over should be very near 0 for lengths of DNA of the order of functional genes—e.g., cistrons.
3. Individual base pairs, *at least,* still have the "atomistic" status of the bead-like genes of the old model, since crossing-over cannot meaningfully be said to occur within a base.
4. The linear arrangement of the genes on chromosomes (preserved in the linearity of the primary structure of the DNA molecule) is unchanged in the modern account, and plays a central role in accounting for the high stability of the genes, the high reliability of

> the segregation mechanisms (without which genetics would be impossible), and the low frequency of "contamination" in the heterozygote.

But intra-genic recombination is assumed to be possible on the molecular account, and not on the beads-on-a-string model. Does this make the molecular theory a "neo-contaminationist" theory rather than a neo-classical one?[24]

Castle (1919a–1919c) had no well-worked-out mechanism, only a set of experiments that purported to show that classical (pre-Morganian) Mendelism did not work. There was little in Castle's work from which "neo-contaminationists" could claim descent. The purported phenomena of Castle's experiments for "contamination" turned out to be non-existent or to admit of Morganian explanations. His explanations had no important connections with the explanations a molecular neo-contaminationist would give for his neo-contamination phenomena, but Morgan's did. Thus, without a theory, a mechanism, or a set of phenomena persisting through time to call their own, there is no "Castlian genetics," and there are no molecular neo-contaminationists.

The kinds of connections between the two accounts clearly support the claim that the mechanism of the Morganian and molecular theories (especially when looked at with the time and size scale appropriate to the Morganian account—a move appropriate to showing that T_2 and T_2* are strongly analogous) are indeed strongly analogous. I thus agree with Schaffner and Ruse on this issue.

Indeed, there has been so little change, and what has changed has done so with such continuity that it is tempting not to describe this as a case of successional reduction at all. It is very tempting to say that Morgan's gene *is* the molecular gene, at a different level of description, and conversely. But to make this identification in the same breath with a claim of strong analogy is to invite confusion of identity by descent of concepts in successive theories (which is a similarity relation) with referential identity of different level descriptions of the same object (which is an identity relation). The former notion requires no further attention now, but the latter concept and its role in reductive explanations and analyses is radically different on this account from that suggested by the formal model. Furthermore, the much better fit of this account, of the role and uses of identity hypotheses with actual scientific practice, is one of the strongest arguments for this account and against that of the formal model. (Discussion of changing concepts of the gene has con-

tinued through multiple discoveries since. For a more recent review see Beurton, Falk, and Rheinberger, 2000.)

Identificatory Hypotheses as Tools in the Search for Explanations

In its earlier formulations, the classical model of reduction had nothing to say about the role of identifications in reduction. Thus, Nagel (1961) suggested that bridge laws or correspondence rules might be grounded in definitions, conventions, or empirically discovered correlations or hypothesized identifications, as if one was as good as another. The widespread instrumentalism and mistrust of identifications as metaphysical, and as going beyond the evidence, has perhaps led many writers away from asking why scientists might prefer to make one claim rather than another. In the one area where this has been hotly debated (and where postulating identities or postulating correspondences is seen as making a metaphysical difference that bears immediately on matters of importance), philosophers of mind appear to almost universally believe that identity claims are a solely metaphysical and evidentially unsupportable extension beyond the evidence of observable correspondences (see Kim, 1966, for a representative and influential view). Not until the 1970s (see, e.g., Causey, 1972) did philosophers of science find a necessary role for identities in reduction. I wish to suggest an unexplored and absolutely central role for hypothesized identifications as tools in the search for explanations which, among other things, explains a number of features concerning their use that have been considered to be unjustified, unjustifiable, or otherwise anomalous (some aspects of this analysis are discussed more fully in Wimsatt, 1976a, 2006a).

I will assume that we are faced with some upper-level explanatory problem: some phenomenon for which we have no micro-level explanation, or perhaps something that lower-level accounts would lead us to expect at the upper level, but which has not been observed. Such an explanatory failure suggests inaccurate compositional information, or none. How do we discover the source of these inaccuracies, of the locus of our incomplete information? An identity claim, with its subsequent application of Leibniz's Law, provides the most rigorous detector of possible error or of a failure of fit of applicable descriptions at different levels: *Two things are identical if and only if any property of either is a property of the other.* If there are properties apparently had by one but not by the other, then either the identity claim is false (as many are) or else *there are as yet undiscovered translations between descriptions at*

the different levels that show that the relevant properties are indeed shared.

Thus, in principle translatability (or analyzability) is a corollary to and the cutting edge of an identity claim. The identity claim is in turn a tool to ferret out the source of explanatory failures which, by its transitivity, allows one to delve an arbitrary number of levels lower if need be to pinpoint the mismatch, or by its scope, to any properties—however diffuse and relational—to detect a relevant but ignored interaction. (For this reason, I do not share the view of some writers that Leibniz's Law should be weakened in all sorts of ways for intensional contexts, and the like.)

Several interesting features follow from this account:

1. It would be expected that identity claims and claims of translatability should be honored more in the breach than in the observance. They function primarily as templates, which help us to locate and to focus upon *relevant* differences—differences that can help us to solve explanatory problems—in order to remove these differences and thereby to make more accurate identity claims. Thus the warrant for claims of in principle translatability, which was questioned earlier, is the same as that for making the identity claim from which it flows.

2. The warrant for this claim is in part the warrant for using a good tool appropriately: that its employment at this time and in this place may help us to discover a description or suggest a redescription that will allow us to explain some heretofore unexplained phenomenon. There is *no* warrant for using the claim if it is *known* to be false. The strength of the claim, which makes it such a sensitive template, renders it easily falsified, and like any strong claim, its negation carries no or little significant information. Thus, if one of the standard defeating conditions for identification, such as causal relation or failure of spatiotemporal coincidence is known to obtain, the claim is dropped, though perhaps in favor of a correspondence claim (Wimsatt, 1976a, part II).

3. This kind of warrant can, however, apply early in the stages of an investigation, and explains behavior that seems irrational and unjustifiable on a more inductivist account of the making of identity claims. Identity claims are often made on the basis of correspondences between or explanations of only two or three properties, often together with some subsidiary background information of a non-correlational nature. This was in fact true for the early identifications, by Boveri (1902) and by Sutton (1903), of Mendel's "factors" with the chromosomes. To the inductivist, this would look like a wildly irresponsible claim: a projec-

tion from two or three properties of a pair of entities to *all* properties of those entities. Moreover, to add insult to injury, the burden of proof after the making of such a claim is not upon its maker (as one would expect on an inductivist account), but upon those who *doubt* the claim to come up with a counter-instance. Only then is the maker obligated to respond to the putative counter-instance, either by elaborating and defending the claim, or by giving it up, as the case seems to demand. Sutton and Boveri proposed a number of new correspondences on the basis of their identifications, and these were later observed, though subsequent conceptual modifications and clarifications led to an elaboration of the identification claims by Morgan and his students, and the generation of many new predicted correspondences (Darden, 1991; Wimsatt, 1976a, 2006a). The early stages at which identities are proposed; the fact that they seem to provide the basis for, rather than be made on the basis of claims of correspondence; and the location of the burden of proof after the making of an identity claim all support this account of the role of identity claims against the inductivist, who should expect the opposite in each case.

4. The fragility and falsifiability of identity claims are hidden by the "open texture" of our concepts (Waismann, 1951), and in more severe cases, by the same tendency to claim identity by descent of our concepts that makes successional reduction possible. With successional reduction, the similarities *and* differences in the successive theories are analyzed critically and used. Only afterwards is the similarity implied by the possibility of performing a successional reduction invoked to maximize the apparent continuity in this identity-by-descent of theoretical concepts. Similarly, with inter-level identifications, the similarities are used critically to ferret out the differences, and only afterwards are the newly assimilated differences reified after the fact into the original identification. The fact that it has become more specific, more detailed, and sometimes has undergone outright changes is hidden from us, so that we see only the continuity of "identity by descent" in our concept of the specific identifications we have made.

5. This analysis suggests that scientists should prefer identity claims to claims of correspondence when there is no specific reason (such as the violation of one of the identity conditions mentioned in item 2 above) to prefer correspondence. They should do so because they prefer the stronger tool, and not for reasons of "ontological simplicity" (or whatever) as suggested by Kim (1966). From a specific identification, after all, one can generate all necessary correspondences, including new

ones that might arise as new properties and relationships are discovered at one level or another. But from the set of correspondences one might derive from an identification given what is known at a given time, one could *not* (without covert reintroduction of the identification) know how to generate new correspondences to fit the new information as it comes in. Identifications are an effective guide to theory elaboration. Correspondences are not. Thus one can understand not only why identity claims might be made early in the course of an investigation, but also why the metaphysically more conservative strategy of making correspondence claims instead will not work. In a static view of science, identity claims and corresponding claims of correspondence only may be empirically indistinguishable. But in a dynamic view of science, only identity claims can effectively move science forward (this is substantially elaborated in Wimsatt, 2006a).

The analysis of reduction and of correlative activities proposed here has differed from most extant analyses in two important respects. First, it has been primarily functional, with the aim of deriving and explaining salient structural features (including some not explained by the standard model) in terms of their functioning in efficiently promoting the aims of science; most notably, explanation. Second, it has aimed at a dynamical account of science, in which optimally efficient change and elaboration are the primary process, and in which stasis is either an artificial construct, a temporary blockage that must be explained, or an end state that we are not likely to reach in the foreseeable future. I believe further that it supports realistic conceptions of the nature of theoretical entities, and of the functions and roles of scientific theory, and does so while being truer to the ways in which scientists *actually* behave than the extant analyses of these activities deriving from the structuralist, static, and often instrumentalist logical empiricist tradition. Finally, it fits into a broader generically evolutionary account of man and his activities, and encourages me to believe that biology may soon be a source for paradigms and analyses that will inform philosophy and philosophy of science generally, rather than being little more than the backwards field for the brushfire skirmish in which philosophical imperialists moving out from the "hard" sciences stop to try their weapons. The latter time is now fast receding into the past, but it is not yet so far that we cannot remember it.

Appendix: Modifications Appropriate to a Cost-Benefit Version of Salmon's Account of Explanation

Salmon (1971, p. 55) defines what it is for one variable to "screen off" another as follows:

> D screens off C from B in reference class A if and only if:
> (i) $P(B/A.C.D) = P(B/A.D)$ [C adds nothing to D.]
> (ii) $P(B/A.C.D) \neq P(B/A.C)$ [D adds something to C.]

Thus, on this interpretation, microstate description D in statistical thermodynamics *screens off* the macro-state description C from B (a macro-state in accordance with a phenomenological macro-law) in A (a macroscopically characterized assumed-ideal gas). This is so because of those fluctuations from the equilibrium state predictable from D, but not predictable from C, which generates the inequality in (ii).

Note how this definition handles an upper-level anomaly (say, a macroscopically unpredictable fluctuation). Since it would be true that:

(1) $P(B^*/A.C.D) = P(B^*/A.D)$
(2) $P(B^*/A.C.D) \neq P(B^*/A.C)$

where all is as before except that B* is a macro-state violating phenomenological macro-laws, it is clear that according to the above definition, D screens off B* from C in A.

It is the consequence and intent of Salmon's definition that any strict improvement in information requires saying that the variables generating the improvement screen off any other set of variables that they represent this sort of improvement upon. *This is so no matter how small the improvement and how great the cost resulting from adopting the new set of variables.* It is another consequence of accepting a view of scientific method appropriate to LaPlacean demons.

Scientific practice and good sense suggest the value of a different notion of screening off, which, because of its obvious connections with cost-benefit analysis, might be called the "effectively screens off" relation:

C *effectively screens off* D from B in reference class A if (and perhaps not only if):

(a) $P(B/A.C.D) = P(B/A.D)$
(b) $P(B/A.C.D) \approx P(B/A.C)$
[D improves the characterization only a little.]

(c) $C(D) >> C(C)$

[D is enormously more expensive information to get than C.]

(c') D is a *compositional redescription* of C.

Some comments are in order about conditions (c) and (c'), which are probably alternatives, or nearly so. The second condition comes closer to capturing the intended application of the effective screening off relationship in the present context, since I am here considering inter-level explanatory reductions, where the lower level is a compositional redescription of the upper level. Furthermore, at least empirically, the truth of (c') appears to guarantee the truth of (c), at least for those kinds of cases we are likely to regard as interesting compositional redescriptions, and thus for all of those cases where we are likely to find any room for debate in the matter of inter-level reduction. Indeed, I am inclined to feel that the proposed "upper level" is not at a distinct level unless at least most of the compositional redescriptions of upper-level phenomena in terms of lower-level entities meet condition (c), which would, in turn, guarantee that any inter-level reduction would be nontrivial.

Condition (c) gives explicitly the cost part of the cost-benefit condition, whereas the approximate equality in (b) guarantees that the benefits, if any, of using redescription D are small. Obviously, the deviation from strict equality in (b) and the cost-ratio in (c) required for the effective screening off relation to hold are interdependent, and are in turn both dependent upon outside factors that determine the importance of additional information and level of acceptable costs. These may vary with the purposes for which the theory is being used, and with any other factors (such as the current explosion in the development of computers and computational facilities) that may radically affect these costs or importances.

The situation where the approximate equality in (b) is in fact an inequality is by far the most interesting one, for *under these circumstances,* D *screens off* C (according to Salmon's definition) *but* C *effectively screens off* D (on my characterization). Thus, in this case, the two criteria would pick out different factors to include in an explanation of phenomenon B.

Condition (a) was also included for the same reason: it is the same as condition (i) in Salmon's definition of the screening-off relation, and thus points directly to a class of cases in which *X* screens off *Y* but *Y* effectively screens off *X*. Condition (a) would presumably be met in any

case in which a successful and total theory reduction (along deductivist lines outlined by Nagel and Schaffner) holds between two theories, such that D is a description imbedded in the reducing theory and C is a description imbedded in the reduced theory. (I would guess that this should be provable as a theorem in the probability calculus from the characteristics of their model of reduction.)

I am not sure, however, how or even whether this result would be provable for reduction as I have characterized that relation. I rather suspect that it is not. Furthermore, in cases where no reduction or only a partial reduction has been accomplished, it would at least be true that condition (a) would not be known to be met for at least some descriptions C in the upper-level theory (and further, that on a subjectivist notion of probability, condition (a) would almost certainly *not* be met for these cases).

In fact, I see no reason why condition (a) should not be dropped for the effective screening-off relation, since conditions (b) and (c)—or (c')—seem to include all that is necessary; namely, the cost-benefit conditions. I have included it for the time being because it heightens the contrast between the screening-off and effective screening off relations, and because I think that substantial further work is necessary to see what if any other modifications and applications seem desirable in developing a cost-benefit model of explanation. The need for at least one further clarification should be immediately obvious: since Salmon (1971, p. 105) points out that his screening-off rule follows from his characterization of explanation, if I believe that the effective screening off relation says something fundamental about the notion of explanation (as I do), it is necessary for me to produce an appropriately modified concept of explanation. This is better left to some future date.

An important consequence of adopting the effective screening off relation rather than the screening off relation was assumed in the text: although upper-level descriptions meeting upper-level laws would effectively screen off lower-level redescriptions, upper-level anomalies—upper-level descriptions that failed to meet upper-level laws—would fail to effectively screen off lower-level redescriptions. This introduced an important asymmetry between cases that met upper-level laws (and which thus were acceptably explained at the upper level) and cases that were upper-level anomalies (and which thus had to be explained at the lower level). On Salmon's screening off relation, there is no asymmetry, since both cases that meet and cases that fail to meet upper-level laws

are explained at the lower level, because lower-level variables screen off upper-level variables in either case.

This asymmetry arises in the following way for the effective screening off relation. Suppose as before that B* represents an upper-level description that is anomalous for upper-level theory. Presumably then:

(a) $P(B^*/A.C.D) = P(B^*/A.D)$

(b) $P(B^*/A.C.D) \neq P(B^*/A.C)$

The failure of condition (b) occurs because if B* is an anomaly, then P(B*/A.C) must either equal zero, or be very low, and much lower, for example, than the probability of states that are held to be explained by the upper-level theory under similar circumstances. On the other hand, if B* is explicable by an account in terms of lower-level variables, there must exist an appropriate description of B* such that P(B*/A.C) *is* appreciably greater than zero—and in general of the order that similar phenomena held to be explicable on the lower-level theory would exhibit. Thus, the failure of condition (b) means that the benefits of redescribing B* at a lower level are not negligible, and in general justify the greater costs implied by conditions (c) or (c').