

Section II: Reductionism(s) in practice

This section is the core of the book. These four papers take a single pair of connected themes—emergence and reductionism—and shows how a once simple topic (*theory* reduction) reveals new dimensions as we turn from justification to discovery, and abandon *in principle* claims to study the heuristics of real-world practice. Here the gap between philosophers' spare deductivist picture of theory reduction and the enormous proliferation of reductionist practices in real science is most striking. It is not just that there are problems with the classical analyses: whole new problem-areas are revealed. I discuss levels of organization in essay 7. These pivotally important macroscopic compound entities are major determinants—probably *the* major determinants—of the geography of science. They are presupposed but not examined in dozens of philosophical accounts given of inter-level reduction. Related approaches yield strategies for naturalizing “perspectives”—the “what it's like to be” of bats, philosophers, and other beasts—the new focal points for modern anti-reductionist philosophers of mind. We also must understand the perspectives of disciplines. “Levels” and “perspectives” provide crucial context for evaluating accounts—of reduction and subjectivity, the reductionistic problem-solving techniques of essay 3, and modeling tactics of essay 4. We must “calibrate” these tools, identifying and analyzing their strengths and limitations. Essay 9 introduces a new analysis of an old concept—that of aggregativity—which abolishes traditional conflicts between reduction and emergence, and identifies a new (or recovers an old) one. The heuristic uses of criteria for aggregativity give entirely new handles for detecting functional localization fallacies and abuses of “nothing but-isms” and “greedy reductionism”.

A closer look shows that there are three notions of reduction—not one. “Successional” and “inter-level” reductions are distinguished in essay 8. Essay 9 yields two sub-types of the latter: “mechanistic” and “aggregative” reductions. “Reductionistic” strategies should be separately considered for each. Some tools are the same (e.g., approximations have important uses in all three) but some are clearly divergent. Functional localization fallacies or their generalized analogues are critically important for inter-level reduction or issues of aggregativity, but irrelevant to successional reduction. A new essay (#9) on emergence and its opposite, aggregativity, relates those concept to heuristics for finding good decompositions of systems into parts, and to natural kinds. Criteria for aggregativity or non-emergence play roles in discovering parts which compose levels of organization and are articulated to provide mechanistic explanations for upper-level phenomena. Philosophy intersects with work on problem-solving in cognitive science in different ways for each of these problems, with benefits for both. I have spent more time on this range of topics than any other, and feel surest of the robustness of the results¹.

The first two papers (essays 6 and 7) analyze the character and architecture of complex systems. They draw on very general physical and evolutionary constraints, show how the structure of our theories and conceptual frameworks are influenced by them, and what kinds of conceptual structures to expect under different circumstances. These papers have a special status, being about the theory we use in defining the pieces in our “piecewise approximations”. They are broadly realist and

¹Less than half of my writings on reduction and reductionism are reprinted here. Omitted are 1971a, 1976a, 1979, 1980a, 1980b, 1981b, 1986, and 1992. 1992 is a historical case study of the development of mechanistic explanations of linkage phenomena in classical genetics, elaborating these views in an extended example. 1979 is an extensive review, with examples and programmatic notes not found elsewhere. Bechtel and Richardson 1992, Schank 1991, Glennan 1992, Waters 1994, and Sarkar 1998 provide complementary perspectives.

naturalistic, using empirical claims about the systems we find in nature as starting points in characterizing our investigative modes, but also taking heuristic, problem-solving, and cognitivist perspectives on how we structure our pictures of the world. I ask what are easy and what are hard problems in the analysis of these systems, how they arise, and how to recognize them. They provide a new *kind* of framework for considering problems of reductionist methodologies. The last two essays, 8 and 9, address more philosophically familiar topics—reduction and emergence—but posed in this new framework. Both argue for a fundamentally problem-solving and heuristic approach to their subjects, with strikingly different results. Extensive introductions to these four papers provide meta-commentary relating the different perspectives on reduction and emergence.

Complexity and Organization (1974):

Essay 6 considers a task usually taken for granted in most accounts—description. When encountering compositional systems, philosophers like to talk about theory reduction. But we don't always have theories to relate, particularly for complex systems. Without pre-existing formal theories as givens, the task of relating them is undefined. We need to know more generally how we order and relate different descriptions of the behavior of a system to construct explanatory mechanistic accounts of its performance. Or even worse: in our search for laws and mechanisms, it is too easy to assume that we already have the relevant descriptions of behavior, and need only to relate them. But biological systems are difficult to describe and analyze in a *systematically* reductionistic manner—even for one with reductionist sympathies. I try to explain why this is so.

I offer first a “phenomenological” descriptive account of our methodological situation in dealing with complexity, but it is explanatory too. I show how with a thorough-going materialism, we can still encounter situations which strain ordinary intuitions derived from looking at simple systems. In biology, (and elsewhere!) we often have multiple cross-cutting theoretical “perspectives” on a system (e.g., the anatomy, physiology, development, and genetics of an organism). These decompose it into parts in different ways. Each can claim to have some but not all of the necessary information for a total description or causal account of the system (or organism). These multiple boundaries and their relations can be a rich source of ways for detecting order². If the parts from these perspectives are composite entities, they should be *in principle* describable in terms of a common set of still lower level parts. [Insert here your favorite list of atoms, ions, and organic, biochemical, and macromolecules—the tinkertoys for all higher level biological things.] But the objects of these

²Particularly rich examples of multiple overlapping boundaries individuated using different criteria (“descriptive complexity” in this paper) are found in Ian McHarg's famous (1971) *Design with Nature*, a description of various city planning projects which had to deal with an enormous heterogeneity of information which McHarg integrated (in a pre-computer age) using multiple overlays (one for each variable) on a map of the region. More recent GIS technology, utilizing diverse variables from remote sensing images has enormously elaborated this kind of approach. And Bruno Latour's forthcoming *Oligopticon: Seeing Paris as a Whole* (1998??, in French??) details maps used in the “control centers” of diverse components of city infrastructure to recognize what is going on and to respond to it: police, fire, ambulances, electricity, telephone, water, and traffic flows are all represented and managed in their overlapping perspectives on the metabolism of the city. These maps are interesting both because they must represent processes as well as objects [Much like a PET scan] but also for their roles in active modulation of those processes. [Thus, traffic representations are used to compute transit times which—flashed to signs on the highway—lead drivers to alter routes to avoid congestion.]

perspectives are not compositionally orderable relative to each other, so we can't order the perspectives relative to each other the way we order compositional levels of organization. We don't have the kind of systematic and complete knowledge needed to reconstruct most larger parts. And we often can't tell how much detail is necessary in the reconstruction. (At best, we have to study different instances to determine the "don't care" conditions allowing multiple realizability for entities of that type. It is not desirable to put in everything, even if it were possible!) Thus we are forced to deal with higher-level, heterogeneously characterized parts *from this multiplicity of different perspectives*, and to characterize their relations to one another to solve most of our problems.

An anatomist and a physiologist might each have charts of their study organism on their wall: The anatomist's is a familiar arrangement of skeletal, muscular, and organ systems—the modern descendant of Vesalius's engravings in *De Fabrica* extended to other size scales. The physiologist's looks more like a program flow chart for how different parts of the various physiological systems interact (connections between elements of respiratory, metabolic, circulatory, nervous, and other systems). The physiologist's chart will usually seem far less complete.³ At more microscopic levels, the anatomist will have switched to a diagram of a typical cell (usually an "ideal type", not an actual cell, with cellular ultra-structural components represented—nucleus, membrane, mitochondrion, endoplasmic reticulum, ...), or perhaps to electron micrographs of these components in real cells. The physiologist will by now be looking at a flow chart of biochemical pathways (such as the Krebs cycle in primary metabolism). Both are spatial decompositions of the same organism (or part of it) into parts, but parts in anatomical and physiological charts won't look anything alike, and no information for how to map one class of parts to the other is provided. (The biochemist's "space" is a topological "reaction space", with a reaction's location in a network indicating its causal antecedents and descendants, NOT the place in the cell where it occurs.) The more recent recognition that biochemical reactions commonly take place in specific locations supported by elements of cellular ultra-structure is now forcing the first systematic articulation of these perspectives.⁴ But tying them together at the level of biochemical reactions and cellular ultrastructure does not—contrary to what one might expect—automatically zip anatomy and physiology together all of the way up. Instead, each pair of anatomical and physiological hooks at ascending levels must be separately and laboriously engaged, often in the conjoint context of neighboring hooks of the same or different sorts above, below, and on all sides.

Higher level parts (and their parts, ...) usually have well delineated *intra*-perspectival relations to one another. But they are hard to relate to other parts *across* perspectives in a systematic fashion, either compositionally or causally, though many problems demand it. Thus the dynamics of interaction of biochemical elements where some components are at hand, some need biochemical synthesis from a precursor two stages up the pathway, and some need to be genetically synthesized, demands analyses of interactions between components in diverse topological and spatial locations in the two decompositions. Components must be both spatially local (so they can interact physically) and topologically local in biochemical reaction space (a reaction exists involving them). This requires

³I've never seen a "whole body" chart of this type—at least at a molar level. (There are biochemical charts of all of the major metabolic pathways.) For partial charts, see Grodins' (1963) description of the cardio-vascular regulatory system, or the hierarchial chart of interactions controlling ovulatory cycles of communally living rats in Schank and McClintock, 1993.

⁴Recognition of the critical role of membrane structure in the theory of oxidative phosphorylation could count as the first fundamental hybrid of anatomical and physiological theories—See Allchin, 1991, and Bechtel and Richardson, 1992.

joint use of both decompositions. These compositional and causal problems reflect two kinds of complexity—*descriptive* and *interactional* complexity—which bedevil the articulation of our different partial theories or perspectives of the organism to produce an integrated account of what it is doing and how. This *problem of conceptual coordination* is the norm for anything other than extremely local and discipline-specific questions in biology. Problems of functional localization represent a special case of this more general phenomenon.

This is an incommensurability—not between competing theories, but between *complementary* theories which we must use together. We don't solve the problem by choosing one or another of the theories—we need all of them! *Philosophical theories of incommensurability do not address this kind of case, but it is far and away the most common kind of incommensurability we deal with in science.* And such cases are encouraging: heuristics for dealing with them suggest useful tools for articulating relations between different paradigms across time as well as across disciplines. **Real** problems of incommensurability are both systematically underestimated (in frequency) by philosophers who have missed this kind of case, and systematically overestimated (in intractability) because we have more experience (and success) in dealing with them than they suppose. These kinds of integration and coordination issues are major foci of Schank's (1991) work, where he argues that computer simulation (of cross-perspectival problems) is a useful—even essential—tool for solving such problems.

This situation provides useful diagnostic criteria for the kinds of complexity we must deal with in biology and the human sciences. Nothing illustrates this better—though not his primary intent—than the complexities and qualifications Schaffner is forced to deal with in his magisterial review, *Explanation and Discovery in the Biomedical Sciences* (1993). Each system and problem—all of the real cases which he brings up—could be exemplified simultaneously (though in different ways) in analyzing organisms of any one of countless biological species. I attempt to describe when these kinds of problems of conceptual coordination can be expected to happen, and suggest (enter the systematic errors!) a range of fallacious inferences we would be more likely to make, and which—with this new understanding—we have a basis for avoiding. (This is the premier domain of functional localization fallacies!) I give concepts for understanding the nature of inter-perspectival disputes in complex sciences; particularly for seeing how theorists with different paradigms can all be partially right, but fundamentally “out of register”. Bringing such views “into focus” simultaneously to compare and integrate them would be useful throughout the human sciences.

Traditional relativists have tried to recognize truths in different perspectives while in effect assuming they are each complete and all competing. They then seem inexorably pushed into saying that if these can *each* be right, then (in a sense) *nothing's* right, and (therefore) *anything* goes. But if we see this as the problem of how to integrate complementary partial-truths, the demands on reality are strong and multi-faceted, but not impossible. Accepting a common referent is a powerful tool in seeing how to reconcile and integrate the different views. If the contributions of different perspectives are partial, there is *room* for them to complement one another. Seeing them as complementary rather than contradictory suggests how to focus and resolve disagreements, or at least how to localize and understand their sources—a precondition for any such resolutions. I urge this therapy for relativist disputes almost everywhere.

The Ontology of Complex Systems (1994):

Use of multiple complementary views or perspectives on an object is continued in essay 7, elaborating an account of levels of organization I first gave in (1976a). Reductive explanation in complex sciences presupposes an account of levels of organization, but no-one has ever offered a non-trivial analysis. We tend to think of levels as collections of objects, but they have a more molar order as well. (Not any collection of objects will do!) They are proper subjects of ontological investigation just as the more “atomistic” abstract things normally embraced by philosophical ontologists—objects, events, causes, properties, and the like. Levels are pivotal parts of our hierarchical view of nature. Size and time scales of particular characteristic causal processes have a surprisingly central role in individuating them. These might seem merely contingent features, but they are sufficiently central to the genesis and properties of levels to make a conceptual difference. Levels have interesting properties bearing directly on problems of explanation, reduction, emergence, evolution, and the nature of living and thinking beings. They give us important handles on the nature of macroscopic theoretical structures and on many of the more molar questions in philosophy of science.

Some properties are features of *any* transition between levels. This includes features treated by philosophers of psychology as if they were special to the mental realm: emergence, “supervenience”, the conceptual and dynamical autonomy of special sciences, multiple realizability, and the “anomalousness” of the mental. Such features apply equally for the relations of chemistry to physics, molecular biology to chemistry, membrane biophysics to molecular biology, the cellular biology of nerve cells to membrane biophysics, neurophysiology to the cellular biology of nerve cells, and so on all of the way up, and (in many cases) for important intermediate levels in between.⁵ Understanding how they emerge in terms of relations between levels puts them in a different light. They are not alternatives to mechanism: unpacking intermediate levels of organization—and these kinds of relations between levels—is an integral part of articulating both specific mechanisms and more general mechanistic explanations of higher level phenomena.

I take it for granted that each new level of organization has characteristic emergent properties (essay 9). And some systematic trends emerge with increasing size: higher levels become less well-defined as their increasing complexity leads to greater interpenetration of levels (at least up to that of the biosphere).⁶ As boundaries between levels break down, we have the emergence of new kinds of macro-entities in and ranging through the biological, psychological, and social realms. One centrally important new type is a *mechanistically explicable* family of things which I call *perspectives*, a special case of which we see in essay 6.⁷ A “perspective”—something much like a “point of view”

⁵Thus understanding relations between the “levels” of chemistry and molecular biology—between bonding properties of simple molecules and the structural character and behavior of macromolecules—involves at least getting from the “*primary structure*” or linear sequence of an assembled protein, through “*secondary structures*” (alpha-helices, beta-sheets and others) into which these fold, to the “*tertiary structure*” or 3-D stereochemical configuration which yields the shape and active sites of the macromolecular machine. (And assembling primary sequence requires a working cell—Moss 1992.) Many proteins have still higher levels of relevant organization. Hemoglobin, for example has at least two more, since it is a tetramer, and (in sickling variants of the molecule), can form still larger super-crystals, which deform red blood cells and have still higher level effects.

⁶We might still discover new levels or principles forcing recognition of new levels where we had seen only an unclear mixture of size and dynamical scales. Ecologist C. S. Holling (1992) urges that new instances of scaling relations are found within ecosystems.

⁷See the glossary—Appendix B for more on these italicized terms.

which we may think of as subjective, or as part of *our* theory of a system—may actually be intrinsic to the system. Since Nagel (1974), perspectives are seen as paradigmatically unanalyzeable and “anti-objectivist”, so this is surprising. (Some—the theoretical perspectives of essay 6, or the ecological niche—are relatively objective. Others are paradigmatically subjective. Yet they have a lot in common.) Understanding what kind of things “perspectives” are and how they relate to levels of organization gives useful purchase on naturalistic anchors for some of the mental dimensions of the human sciences.

“Causal thickets” arise in turn as boundaries between perspectives begin to break down with still further increases in complexity. They do much to explain the character of methodological disputes in psychology and the social sciences. We should expect methodological and conceptual disputes exactly when boundaries among divergent perspectives break down, as well as when they lay claim to the same territory. Ambiguous boundaries between perspectives, how to tell when we have—or assume—perspectives rather than causal thickets, and recognizing new emergent perspectives at higher levels are conceptually challenging aspects of new work on embodied consciousness and social cognition (Thompson Rosch and Varela (1993), McClamrock (1994), and Hutchins (1995)). Another example is the productive fusion of economics, history, cognitive anthropology, psychology, and cultural evolution (Boyd and Richerson 1985) required to make sense of what economic historian Douglass North (1991) calls “institutions.” Or the increasing confluence of paleontology, macroevolution, and evolutionary genetics with developmental biology and genetics (Raff 1996). Here we have whole disciplines as nodes providing increasingly rich contexts for one another. These would have to fuse with developmental psychology and psycholinguistics to engender a new evolutionary theory of development, in which both the internal process-architecture and the rich informational structure of the environment will be seen anew in terms of their intrinsically relational structure. These three mega-disciplinary aggregates will themselves be linked—providing connections to and boundary conditions for each others’ problems.⁸ These aggregations or confluences of disciplines for particular problems do not unify them generally, and thus do not unambiguously locate the problems. Most people are still trained in particular disciplines, so in solving these complex problems, we need to be able to address questions like: “Even though language is a complex multi-disciplinary entity—thicket-like if anything is—under what circumstances, and for what purposes can we get away with studying or approximating it from the perspective of just one of its attendant disciplines?” We can do it sometimes, and just as surely, we can’t at others. Are there any generalizable ways to recognize when a different approach is called for? Can people trained within disciplines be expected to recognize it?

These three new kinds of ontologically macroscopic entities—levels, perspectives, and causal thickets—increase the coherence and plausibility of a generalized mechanistic view of the world, and without the eliminativism that has characterized so many past attempts. Paradoxically, *our usual conditions for a mechanistically acceptable explanation are seen to break down systematically in some cases in a mechanistically explicable fashion, and—therefore—without fundamentally compromising mechanism.* Finally, this paper uses concepts of robustness, drawing on essay 2 to delineate an alternative philosophical methodology. Recognizing when logical structure, crisp

⁸Though speculative, each of these disciplines are now reaching out to their neighbors in unprecedented ways, and seeing their respective problems as related. Problems of human development, the nature of cultural evolution, and its relation to biological evolution require this kind of massive cross pollination and cooperation. Evolutionary psychology, while seriously threatened by oversimplification as it is now developing, is at least moving psychologists in useful directions, integrating both downwards with biology and upwards with the social sciences.

definitions, and universal generalizations are inappropriate to some problems, I propose other ways of proceeding. This paper is methodologically self-referential—that is, it uses the new methods it recommends. It is thus the most philosophically radical paper here—with deeper implications for the methods and practice of philosophy, and for strategies in the analysis of key concepts and problems of the human sciences.

Reductive Explanation—A Functional Account (1976):

This methodological perspective gives a new fundamentally and deliberately more approximate framework in which to understand the use of exact and inexact tools alike—including those of essay 8. This essay comes the closest to traditional accounts of reduction: at least it spends more time explicitly discussing philosophical literature on the topic. It also proposed several theses then uncommon or unknown, but since widely accepted.⁹ Modulating Nickles (1973), I argue that the “classical” search for *the* logical form of reductions conflates two fundamentally different activities.

In theory succession, a newer and better theory replaces or reduces to an older one. Both either deal with phenomena at the *same* compositional level of organization, or (for some physical theories), lay claim to phenomena at all levels. If structures of older and newer theories are too dissimilar to construct reductive transformations (serving various problem-solving functions elaborated here), there is no reason to keep the ontology of the older theory, and there can be wholesale eliminations. (By contrast, if there *is* a reductive transformation, the ontology of the older theory will tend to be preserved as limiting case approximations of the ontology of the newer one. And the limiting cases will usually be, or include, the ontology of our world.)¹⁰

The other activity called “reduction” need not relate theories at all, but refers instead to the explanation of upper level *phenomena* and regularities in terms of lower-level *mechanisms*.¹¹ *It is*

⁹Various authors draw piecemeal on parts of this paper (or the overlapping 1976a), but often seem to miss its points. Essays 6, 7, and this introduction should help to motivate its different perspective. Hooker (1981), Kitcher, (1982), Cartwright, (1983, 1990), the Churchlands (1986, 1996), and Waters, (1990) have since offered views overlapping it in different respects, and Bechtel and Richardson (1993), McCauley (1996), Ramsey (1995), Sarkar (1992, 1995), and Glennan (1992)) provide extensions, related analyses, and further support.

¹⁰Ramsey (1995) elaborates this kind of reduction with cases where a “less good” theory is a limiting special case of a more general, fundamental, or better theory, but derived from it afterwards in a new way to *apply it* to a special case. The ontology of the special case could even be “created” by the reduction! He also exploits connections with engineering practice. Batterman (1995) finds new complexities which may affect both types of reduction: mathematical singularities blocking deduction, and hybrid intermediate theories (notably so called “semi-classical theory” relating quantum mechanics and classical physics) capturing phenomena not generable from either “parent” theory.

¹¹There may be theories at the various levels, but not necessarily. The aim is usually to get lower level mechanistic explanations of higher level phenomena (an “inter-level theory”—Darden and Maull, 1977). With complex systems it is harder to find well articulated theoretical structures true enough to the phenomena at two different levels for them both to be called theories. Statistical mechanics and classical thermodynamics, and classical and molecular genetics represent exceptions

never eliminative. The idea that there is a kind of *theory reduction* (“eliminative reduction”—favored by some philosophers of mind) in which higher level objects and relations are systematically eliminated in favor of lower level ones, rather than extended, modified, or transformed, arises from a conflation of these two distinct activities. *No such beast as eliminative reduction is to be found anywhere in the history of science, and there is no reason, in terms of the scientific functions served, to expect it in the future. It, and its aims, are largely misconceived philosophical inventions. Robust higher-level entities, relations, and regularities do not disappear wholesale in lower-level scientific revolutions—our conceptions of them transmute and add new dimensions in interesting ways, yes, but disappear, no.*

§ § § Excursion on Eliminative Reduction: § § §

As always, the story is more complicated than it appears, but this is a good first approximation.¹² *First*, an eliminative view in inter-level cases would be more plausible if all system properties were simple aggregates of parts properties—this would motivate “nothing but” style eliminativist talk. But as Essay 9 shows, these conditions are essentially never met. Confusions between consequences of total aggregativity and what to expect of inter-level reductions would generate illegitimate support for eliminativism. (This confusion is abetted for philosophers by a tradition with views like Russell’s logical atomism, in which derived terms (including theoretical terms and material objects) were mere “logical fictions” and *in principle* dispensable.)

Second, the phlogiston/ oxygen case often quoted by eliminativists to support their views was *not* a case of theories or phenomena at different levels of organization as one might suppose, but one of successional reduction at the same level, where ontological elimination *can* occur. (This dispute took place before Dalton’s atomic theory, so during their competition, phlogiston and oxidation theories were accounts offered at the same macroscopic level: oxygen was an extract of or species of air.) So elimination of entities in this case does not bear on the issue of eliminations for inter-level accounts.

Third, (here it gets more complex) most real cases of scientific progress in understanding an inter-level mechanism or coordinated cluster of phenomena involve *both* successional and inter-level change and reduction. So elimination through inter-level scientific progress is possible, but it would appear as a displacement of one set of conceptions of macro-level things by another—not by *any* conceptions of *micro*-level things. More controversially, most such “replacements” will preserve so much of the older phenomenology that changes will not be obvious in most operations at the macro-level. Wimsatt (1986a, 1999b, 2002) provides a plausible case: the fundamental reconstruction of the innate-acquired distinction necessary for consistency with new theory and data in evolutionary and developmental biology and ethology. If this reconstruction is regarded as a replacement, it is unconvincing as a paradigm for the eliminative replacement of folk psychology. The replacement for the innate-acquired distinction preserves the vast majority of its traditional consequences. If anything it *increases* their number, and for the first time explains their relationships in an integral fashion in such a way as to make the distinction more robust than before—although it is drawn along radically different lines. To be sure, there are some crucial differences, but the net appearance is not of eliminating a traditional discipline or perspective. New macro-level conceptions will both have to

here. That explanations generally are explanations of phenomena rather than of data or linguistic entities has since been argued by Bogen and Woodward (1988).

¹²See also McCauley’s excellent (1996) discussion of the issues surrounding theory coevolution and eliminativism.

capture any *robust* phenomena of the older set and be appropriately related to or anchored at the lower level.

A *fourth* alternative provides intriguing possibilities for agreement.¹³ Conceptual change can attenuate or expand what we attribute to a level. This is because functional localization fallacies can lead us to mis-assign levels for phenomena. Conceptual change can make us aware of this and lead to reassignments. I'm quite happy to explain Mach bands (the heightened contrast at color change boundaries in the visual field) in terms of the activity of lateral inhibition networks—giving them a lower-level anchor, though we still have to deal adequately with the phenomenal colors which *show* the increased contrast at the boundaries. But some mental phenomena will also be correctly reassigned to higher levels. (See Hutchins (1994) and others on social cognition). Indeed, biases of reductionistic research strategies (essay 3 and appendix A) *predict* that in periods dominated by reductionistic views, more properties should be mislocalized at too low a level than at too high a level. It is thus plausible that many properties attributed to individual psyches are in fact psychologically reified projections of social practices—a tendency exacerbated by our “methodological individualism”. This seems especially likely for claims that large fractions of our internal thought processes are deductive inferences. We construct valid inferences, but usually with a great deal of external help—ranging from the development of our language, logic, and mathematics, to our use of external tools (ranging from written language to computers) to check and elaborate the results. The current “computational” world view suffers from an exactly parallel malady—projecting downwards many features of a high level abstraction.

So could a level—e.g., the traditional mental realm of individual psychology—evaporate totally? Not likely: but our theories of mentality or cognition may diffuse *sideways*, as well as downward or upward, as people like Frank (1988) argue the inadequacies of simplistic faculty psychology which put cognition, conation, and affection into separate black boxes, all distanced from the motoric (which I predict will also come to be seen as central). In our current framework, cognition has grabbed too much credit, which should be better shared with other mental faculties. (See the bias-producing phenomenon of “perceptual focus” in Wimsatt, 1980b). I believe that the *apparent* personal level will get thinner: thinned to the stuff that's “really” there—robustly accessible relatively directly at that level.

This now introduces more focus to the dispute: Now we can ask whether there *are* any robust phenomena at the upper or intentional level. I think that there are lots of them, including phenomenal colors (not *qualia*, if those are **defined** as things which can't be explicated in any other way) and a variety of phenomena loosely characterized as intentional. If we do away with beliefs, intentions, and propositions, it could only be in favor of things or assemblages of them which behave an awful lot like them in a wide variety of contexts. I don't feel the same way about propositional attitudes—which in their elaborations of the 1960's and 1970's seem more the constructions of philosophers than the beliefs of any “folk”. On phenomenal colors, see Evan Thompson's “Novel Colors” (1993). For intentionality and the self, I think that there are non-eliminative ways of reading most of what Dan Dennett argues in his *Consciousness Explained* (1991). I won't justify these claims here. They are made to locate the natural extensions of my position in a broader conceptual geography.

§ § §

¹³This possibility precipitated out of a conversation with Bob McCauley and Paul Churchland at a PSA reception in 1996.

Discussions of eliminativism have grabbed the stage, but these distinctions do more. The functions of successional versus inter-level reduction reveal and explain unnoticed differences in their logical form and assumptions, and point to the use of a number of detailed and powerful heuristics for theory construction for inter-level reductive explanation. The analysis of essay 8 thus delivers on a promissory note of increased relevance and usefulness to scientists. It suggests that they should *not* look to construct eliminative theories but look for resources at both upper and lower levels to constrain, and more positively, to use in connecting with and refining the other level, or with new phenomena and entities at levels in between. Purely functionalist approaches favored by advocates of the “special sciences” are also compromised: knowledge of lower-level mechanisms is a crucial source for the refinement of upper-level accounts, and conversely. This is beautifully supported in Bechtel and Mundale’s (1999) discussion of “multiple realizability”. The view argued here makes specific suggestions for how to use these resources in constructing and refining identificatory or localizationist hypotheses.¹⁴ It is also *symmetric*: it does not privilege lower level accounts over higher level ones in the development of reductionistic explanations. Normally, each has things to contribute to the development of accounts at the other level, and usually each needs revision to fit—a phenomenon suggested by Schaffner in his model of reduction as early as 1969, and elaborated here and in my 1976a as a co-evolutionary process. These views have attracted growing support (and significant elaboration) since. See e.g., Darden and Maull (1977), Churchland, (1986), Bechtel and Richardson (1992), and McCauley (1996).

This paper was prescient in other respects: I advocated a non-eliminativist causal interpretation of Salmon’s (1971) “statistical relevance” account of explanation years before he did. His account (and attack on the deductive nomological model of explanation) dovetails more naturally with a search for causal factors which we articulate into mechanisms than with a search for laws. But to make sense of how scientists deal with mechanistic explanations, we must also modify his scheme (as he did not) to bring our search for causal mechanisms in line with our bounded abilities: in giving mechanisms, scientists stop short of exhaustively complete accounts when the yield from additional causal factors gets too small, too rare, or both. (This is in effect a satisficing version of a cost-benefit rule). *Mechanisms just are truncated, partially and roughly characterized, manipulateable, and relatively modular and generalizeable patches of the causal fabric of the world.* They are the creatures of a backwoods mechanic or an engineer—not of a mathematician or foundational physicist—though they are used by model-building theoretical physicists in all other areas. Mechanisms are portable and applicable in different contexts, and subject to *Ceteris Paribus* qualifiers in an explicable way (see Glennan, 1994 on mechanism, and Mikkelsen, 1997, on the role of *Ceteris paribus* in the analysis of counterfactuals).

This picture resonates with Nancy Cartright’s later rejection (for different but convergent reasons) of laws as central to causal explanation.¹⁵ In (1992), I extend this account: to understand the

¹⁴Lindley Darden, Bill Bechtel and Bob Richardson, and Sahotra Sarkar urged this correction of the views expressed here and in my 1976a. (See their 1991, 1993, and 1996, respectively). They point out the explicit target of a developing inter-level account is often the (lower-level) **localization** (and mechanistic explanation) of an upper-level property, object, or relationship not an inter-level identificatory hypotheses. Localizations are indeed often the target. However, since they require the identity of spatial properties, they guarantee the relevant causally relevant properties.

¹⁵“Mechanisms” in my account play a similar role to “capacities” in hers, with this difference: “mechanisms” key more naturally to the ways we give, modify, test, and elaborate mechanistic

Morgan school's differences with Haldane over the role of general models of linkage, one has to assume that "laws" played a distinctly derivative role for them—as templates which they expected to be falsified, and used primarily as tools for detecting anomalies, which pointed to deeper mechanisms. We are thus led to a picture of theory structure in genetics in which mechanisms are primary, and are revealed not directly through laws but through a tissue of successive anomalies to tentatively advanced models and generalizations, a view I call—with a bow to Nancy—"particularistic mechanism." It fits mechanistic investigations and explanations more generally.

This paper also looks similar to Kitcher's (1981) explanatory unification account, though the metaphysics is quite different. The flavor here is more strongly realist and mechanist than Kitcher's, and gives explanatory unification with a strictly local flavor as a *consequence* of our activity rather than as its aim. We unify in terms of mechanisms, looking for the same or similar ones in diverse places, rather than in terms of laws. We also seek to integrate our accounts of phenomena by connecting them as richly (as robustly) as possible into the causal network, but do so without use of exceptionless generalizations, laws, and inference rules as Kitcher does.

Finally, to generate explanatory *levels* I invoked "cost-benefit" considerations a second time. A "cost-benefit" revision of Salmon's "screening off" relation (my "effective screening off") allows us to apply screening off to variables from inter-level theories and mechanistic accounts of the same entity at different levels of description without having the lower-level variables automatically screen off all higher-level ones since they are (*in principle*) parts of a more accurate theory and therefore better predictors. Levins' (1966) "sufficient parameters" (more realistic heuristic analogues to multiple realizeability and supervenience) makes higher level variables both theoretically *cost-effective* (thus things that upper level organisms want to detect and use), and pragmatically *causally effective*. This is natural once we get away from the unrealistic expectation that causal relationships require or entail generalizations which are both usable and exceptionless. In the real world (of folk-psychology, and of science as applied to real-world systems), we usually can have one or the other, but not both, and we will also almost always choose usability over exceptionless universality. Cartwright (1983, 1990) argues similar points, but embraces probabilistic theories of causality. This is an error which would undercut all of the techniques we use for "localizing faults" and figuring out what happened at the macro-level (See also Glennan 1994)).

When the same macro-state may be realized by a variety of micro-states (i.e., for compositional levels of organization), Salmon's original account of screening off introduces an unrealistic reductionistic bias favoring lower-level redescriptions. So do most philosophical accounts relating causes at different levels. With multiple realizeability, a cost-benefit account naturally leads (via "sufficient parameters" (essays 2 and 12) and "dynamical autonomy" (essay 2)) one to recognize higher level variables as *causally potent under some, but not all, conditions*. This explains when we choose to explain phenomena at an upper level, and when it is necessary or profitable to go to a lower level—for phenomena or regularities which are anomalous at the upper level. Blind attempts to construct exceptionless generalizations at any cost would produce cumbersome and useless structures blunting this necessary multi-level dialectic. Only this account correctly captures how and when working scientists choose to work at one level or another.¹⁶

explanations; "capacities" seem chosen in part to deal with her focus on foundational questions at the quantum mechanical level. **See appendix B?? for further discussion.**

¹⁶Others have since made related moves, using screening-off to secure the autonomy of upper-level phenomena, though I have problems with how they did it. Thus Brandon (1982) argues that

With accounts of the “dynamical autonomy” and robustness of upper level variables (essays 7, 2, and 1976a), this provides the only analysis consistent with a reductionist methodology while justifying upper-level talk in terms stronger than pragmatic convenience. Other accounts either make it unintelligible why one should ever give upper-level explanations if one has an acceptable lower-level theory, or go so far in justifying upper-level accounts that it becomes mysterious how lower-level factors could ever get the purchase on upper-level phenomena they so obviously have. Neither extreme is acceptable: both fail to capture the integrative articulation in a single mechanistic explanation of entities, phenomena, and causes from different levels—the basic form of explanations which are so common in modern scientific and engineering practice.¹⁷

Emergence as non-Aggregativity and the Biases of Reductionism(s) (1996):

Essay 9 [new in this collection; see the shorter 1997] takes up the idea of emergence as a relationship between a property of a system and properties of its parts. (A system commonly has some properties which are emergent and others which are not.) Philosophers often suppose that emergence implies irreducibility—or if they are reductionists, treat claims of emergence as counsels of ignorance. Scientists would not accept this: many properties they claim as emergent are not only reducible, but straightforwardly so. Mechanism is at its most convincing when it can give an intuitive explanation of a system property regarded as emergent in terms of properties of the parts of that system. But having that explanation doesn't deny emergence. Extending my (1986b), I argue that emergence indicates dependence of a system property upon the mode of organization of parts of that system. But how should we characterize this?

Given the many possible modes of organizational dependence, it is better to flip the problem: if the mode of organization of the parts *didn't* matter at all, the system property would seem to collapse as nothing more than an aggregate of parts' properties. What does this entail? Four plausible conditions are required for the system property to be *invariant* over a (specific) class of operations on the parts and their properties. Meeting these conditions rules out any organization¹⁸ Emergence is

phenotype screens off genotype in selective explanations of evolutionary change, though his example invites a confounding of temporal with multi-level effects in the production of screening off. While attractive in other respects, McClamrock's analysis (1994) cannot do the job here—no analysis which does without the equivalent of an effective screening off relation can. (In fairness, he does not intend it to do this job, though perhaps he should.)

¹⁷Glennan (1992) and Schank (1991) each use insights from object-oriented programming to characterize mechanistic explanation and analysis. This should be the wave of the future. [Think of a mechanism as a potentially portable object in the sense of OOP—with standardized inputs and outputs—which, in the extreme case of engineered interchangeable parts, we can carry around and install, replace, or adapt it where it fits. Our sense of mechanism as an object and an articulated set of objects owes much to our engineering past. Our models of mechanisms do also, reflected in the OOP notion of an “object.”] Other recent analyses of inter-level explanations drawing strongly on scientific practice, e.g., Darden (1991) or Bechtel and Richardson (1992), complement the approach taken here.

¹⁸Why isn't aggregativity a kind of multiple realizeability? It is—an unusually strong form, but so strong and implying such a homogeneity of the parts' actions as to deny the *multiplicity* of

thus a failure to meet one or more of these conditions—blocking *aggregativity*. But organization may be internal to the parts as well as embodied in their relationships. So require the conditions to be met *for all possible decompositions of the system into parts* if the relationship is to be aggregative. There are diverse ways they can fail to be met—providing an interesting variety of kinds of emergence. Various examples are classified according to how they meet or fail these conditions. I'll return to the conditions shortly.

Aggregativity suggests a third (very extreme) sense of reduction which has been influential in a back-handed sort of way. Although seldom satisfied, our misconceptions about it—conflating it with inter-level reduction—probably motivate negative attitudes towards reductionistic accounts on the one hand, and the mistaken association of reductionism with eliminativism on the other. (These mistakes are endemic on both sides of almost every debate over reductionism.)

Mathematical biologist Jack Cowan loves to describe the difference between biophysicists and theoretical biologists. (A university president once said to him: “*You both use a lot of math and physics to do biology—you must be doing the same thing. Why shouldn't I merge your departments?*”)

“*I'll tell you the difference,*” Cowan said, “—*take an organism and homogenize it in a Waring blender. The biophysicist is interested in those properties that are invariant under that transformation.*” One couldn't get a more graphic image of the difference between aggregativity and emergence.¹⁹

What if some properties of the parts and system were invariant no matter how you cut it up, aggregated, or rearranged its parts? *For such properties, organization wouldn't matter.* There are such—those picked out by the great conservation laws of physics: mass, energy, charge, etc. As far as we know, that's all. These meet *very* restrictive conditions: *for any decompositions of the system into parts*, these properties are *invariant* over appropriate rearrangements, substitutions, and reaggregations, and their values scale appropriately under additions or subtractions to the system. For *these* “aggregative” properties, we *are* willing to say: “The mass of that steer I gave you was nothing more than the mass of its parts”. And we blame the butcher—not vanished emergent interactions—for any shortfalls. If these four conditions (informally stated above) are met *for all possible decompositions of the system into parts*, aggregativity must be an extremely demanding relationship, one seldom found in nature. (Even Waring blenders only disrupt organisms down to the macromolecular level, preserving organization at lower levels, which is why they—used with centrifuges to separate out fractions of varying densities from the homogenate—have such a large role in isolating specific macromolecules.)

Emergence is thus extremely common—much more so than normally supposed. Is this then too weak a notion of emergence? It fits the intuitions of most scientists I know, who want both their

realizations! With aggregativity, there's no tendency to separate system property from sum of parts properties; they are simply identified.

¹⁹Cowan was chair of the very successful new department of Theoretical Biology, descendant of the historic Committee on Mathematical Biology, at Chicago. President John Wilson merged the departments, and better-funded biophysics took over. They had little in common. The theoretical biologists soon found appointments in other places in the university where Waring blenders were not used. And of course, the biophysicists weren't *uninterested* in biological organization: disrupting cells and tissues made it easier for them to isolate the macromolecules whose organization they were interested in.

reductionism and their emergence, and who agree with its classification of particular cases. Even with stronger notions of emergence to explore, the specificity and power of these four conditions make them extremely useful tools of analysis. Were a stronger notion needed (something not yet demonstrated), this notion could “clear the brush” of the many cases it captures to focus discussion on those which remain. Aggregativity also gives a powerful handle for diagnosing certain important kinds of errors. Doing so first requires a short discussion of approximations.

Since these conditions are expressed in terms of *invariance* of the system property under operations on the parts, one can easily produce a family of quantitative criteria for approximate or local aggregativity, where variations of the system property within an “error tolerance” of $\pm \epsilon$ are tolerated for various values of ϵ . Essay 6 uses a similar move for different degrees of near-decomposeability or modularity in systems. Adding such “tolerances” to an analysis is a useful strategy for qualitative concepts in a messy, inexact, and approximate world having many regularities and stable patterns, but few exceptionless generalizations. (We need “sloppy gappy generalizations”. This is quite consistent with determinism. See essay 5 and the discussion of deterministic chaos (with pictures!) in Wimsatt, 1991. Essay 7 also supports this view.)

Tolerances are essential because we often use quantitative and formal qualitative frameworks as templates for pattern detection and matching which Nature may meet in varying degrees. With a particularly adaptable framework which can be fitted onto nature in diverse places in different possible ways, we may try a variety of such mappings, looking for “best fits”.²⁰ The complex ontological objects—levels and perspectives (essay 7)—result from *natural* selection and stabilization processes involving “best fits” of such frameworks *in nature*. Our modeling and theory construction also make fundamental uses of such processes. The conditions of aggregativity are just such an approximate and adjustable framework, since one can construct orderings for how well each condition is met, and partial orderings for how well all four of the conditions are met across different decompositions of a system into parts.²¹ Indeed, to the extent that modularity exhibits some of the features of aggregativity, the importance of modularity in evolution (and in neural architecture) implicates these conditions causally (through selection) in the architecture of many natural objects (See Wagner and Altenberg, 1996).

Situations where the conditions are met (or approximately so) for some decompositions, but not for most, are particularly interesting. These properties *look* aggregative for those decompositions, but reveal themselves as emergent or organization-dependent for others. *The better a decomposition meets these conditions, (meeting more of them, more exactly, and over a broader range of conditions), the more nearly it factors the system into modular parts which can be characterized from*

²⁰How do we recognize when such a framework is only being used as a “curve-fitting” device? (Instrumentalists cannot even ask this one, since to them *all* theories are curve fitting devices.] We need criteria for telling realistic theories from curve-fitting ones in a world where there are *both* kinds of entities, and hybrids between them. Ramsey (1991) discusses the “semi-empirical method” in chemical kinetics—a provocative case of such a hybrid, and I treat (1992) a revealing juxtaposition of the two in Haldane’s (1919) self-conscious simultaneous use of causal and predictive non-causal models of linkage and mapping functions in the same paper.

²¹Giere (1988, p. 106) argues that many puzzles are solved by distinguishing two dimensions of approximation. A model is similar in some *respects* but not in others to the thing modelled, and accurate to different *degrees* in each of these respects. We get consilience if we treat the 4 conditions as his “respects”, and orderings within conditions as his “degrees”.

that perspective in terms of monadic, intrinsic, or context-independent properties. Such perspectives thus provide particularly simple and theoretically productive decompositions of the system into parts. We are more likely to see these properties as natural, and these parts as instances of natural kinds, as robust, and to regard the system as “nothing more” than the collection of these parts. Note the use of aggregativity here: an apparent foundational distinction between kinds of properties is transformed into a search heuristic for finding preferred, simple, “maximally reductionistic” analyses of systems. Given the biases of reductionistic problem-solving strategies (essay 3, appendix A), such decompositions lead readily to excesses of “nothing but” talk and disciplinary imperialism.

Aggregativity can aid the search for good parts decompositions (other criteria are also used!). What else does it have to do with inter-level reductive explanation (as “mechanistically explicable”) in essays 6, 7 and 8? The answer is—virtually nothing! (For a start, total aggregativity has almost no instances, and the latter has many.) But there is a curious operational paradox: *with total knowledge of a system, the two senses are clearly distinguishable, but when we know very little about a system, our strategies in approaching them look very similar.* This is the source of many confusions! The simplest theories we can construct for the interaction of a number of parts tend to model them in a very homogeneous and aggregative fashion. These are “first-order approximation” models we construct (because they are very simple) for systems we don’t yet know much about (Wimsatt, 1979, 1980b). We thus use similar *starting points* in modelling diverse systems: some quite aggregative, others non-linear but still symmetric and homogeneous (tolerating rearrangements and substitutions of parts), and others which are mechanistic, highly differentiated, and whose key properties are highly organization dependent. With partial knowledge but predominant ignorance, it is thus less surprising that the properties of mechanisms and of aggregates should be easily confused—especially if the former are modular or partially aggregative.

This explains temptations to make unsupported and excessive claims of the sort common in early stages of a reductionistic analysis, and then to quietly take them back later. Total aggregativity *would* compromise claims for autonomous higher level properties and systems. *But it never happens.* Mechanistic explanations of phenomena commonly invoke mechanisms with highly differentiated parts and behavior which depends on their mode of organization. Some people see such explanations as very threatening. But they misidentify the enemy: only aggregativity, not mechanism, would justify the claims they fear. And that won’t happen. Analyzing these excessive claims and their biases (commonly associated with “eliminativist” positions) is especially important for fields and explanatory tasks whose major questions are still “in process”. They are just the kinds of evaluations and calibrations of conceptual tools we should seek for limited, fallible, and error-prone scientists. To assess these claims fairly we *must* recognize the limitations and incompleteness of our knowledge, the heuristic character of our tools, and the specific biases which likely issue from their application. This is then another point at which the LaPlacean image of the ideal scientist—who already knows everything—serves us poorly.

But then why should reductionist methodologies have appeared to be so successful? In part because they often *are* genuinely successful. But there is more. Heuristic principles characteristically transform hard problems into different but related ones which are easier to solve (essay 3). If we then solve them effectively, we will likely identify the new problem with the old one—saying, “Now that we’ve *clarified* the problem so that it can be solved, ...”, without noticing that it’s been changed! *In this way, quite substantial changes in a paradigm can be hidden—particularly a cumulative string of such changes, each too small to be regarded as “fundamental”.* Without denying the power of reductionistic approaches, this kind of *ex post facto* reification helps to generate excessively high opinions of them, and—in another context—the exaggerated belief that

work elaborating a paradigm or applying a theory is merely playing out options which are already given.

Positivists commonly underestimated the importance and difficulty of applying a theory. This is an important lesson one can learn by studying engineering. There is a mythology that engineers don't have any proper theories of their own: they just develop techniques to apply the basic theories of others. This is inaccurate. Essay #10 below, and the work of Walter Vincenti (1990) demonstrate just how conceptually rich these "mere techniques" can be. Engineers do often use or presuppose basic theories in various other sciences. So what should we make of this? To treat the applications or consequences as somehow already contained in the theory is a patently absurd view on a moment's sensible reflection! It diminishes the creativity of those who work applying and elaborating the paradigm, and overestimates its original power. These are common errors for philosophers. It may be natural for a foundationalist, but all philosophers seem to have tendencies in that direction—probably from too much attention to *in principle* arguments, which collapse the long train of argumentation and the many creative turns from where you are to where you want to go. Other things also contribute: the traditional focus on justification rather than discovery—especially if one never looks at cases of the latter—makes it easy to undervalue the difficulty and creativity of applying a theory in a new area or in a new way. By looking at such achievements only after the fact, when justification is the focus, they look not only easy but unavoidable.²² Sensitive contemporary reconstructions of incidents in the history of science are the best possible therapy for this Whiggish disease. And in this context, the rise of laboratory studies and analyses of practice should be liberating to the philosophical imagination.

²²The kind of *ex post facto reification* discussed in the last several paragraphs is central to Wittgenstein's concern in *Remarks on the Foundations of Mathematics* (and the *Investigations*) when discussing how we follow rules. Before the fact, it appears that we have a choice in how we apply a rule, but after the fact it appears that we *had* to go in the chosen direction. This important psychological phenomenon arises when we assimilate the new application to our understanding of the rule. This assimilation is central to the perceived sense of continuity we have in many of our activities. As an epistemological question—how much and why we had to continue as we did—the discussion of generative entrenchment in essay 11 elaborates why and how some epistemic actions may become effectively irreversible after they are performed, and the consequences of this fact. Finally, the lovely chapter 10 by Hoffrage and Hertwig on hindsight bias as a product of adaptive processes (in Gigerenzer et. al. 1999) brings this whole topic successfully within the scope of heuristic inference.