Reductionism and its heuristics: Making methodological reductionism honest

William C. Wimsatt

Abstract   Methodological reductionists practice 'wannabe reductionism'. They claim that one should pursue reductionism, but never propose how. I integrate two strains in prior work to do so. Three kinds of activities are pursued as "reductionist". "Successional reduction" and inter-level mechanistic explanation are legitimate and powerful strategies. Eliminativism is generally ill-conceived. Specific problem-solving heuristics for constructing inter-level mechanistic explanations show why and when they can provide powerful and fruitful tools and insights, but sometimes lead to erroneous results. I show how traditional metaphysical approaches fail to engage how science is done. The methods used do so, and support a pragmatic and non-eliminativist realism.

We take it for granted that human activity, including science, is purposive. Yet we have ignored this fact in our analysis of reductionistic activities. Or perhaps we have too easily taken for granted the time-honored aims of the unity of science and onto-logical economy. But these are aims of the metaphysician, and not, save perhaps in foundational projects of physics, those of most scientists. Scientists pursue more local goals, at several layers of increasing locality. I investigate them here. First I argue that there are two types of reduction, differentiated by function, which I call successional reduction and inter-level explanation. I here follow the lead of Thomas Nickles' classics 1973 paper, but in characterizing them functionally, I add other distinguishing features. The first kind of reduction plays crucial functions in theory succession, and I discuss it in passing to focus on the second. The second involves the use of identities and localizations to generate inter-level accounts of upper-level phenomena. The use of identities and localizations is a powerful way of generating new predictions, and

W. C. Wimsatt (✉)
Philosophy, The University of Chicago,
Chicago, IL 60637, USA
e-mail: wwim@midway.uchicago.edu

the locus of many useful heuristics. I show how this works by exploiting the rich consequences of posited mechanisms, thus justifying their use in functional terms. This part of the paper elaborates lines first sketched in my 1976a and b.

But this is not the popular understanding of reductionism, which many see as a claim to *Nothing-but-ism*; an attempt to deny or discredit the explanatory potency of upper level entities, phenomena, and causes. To justify this would require satisfaction of extremely strong conditions, (to argue that a "whole is nothing more than the sum of its parts"). These conditions are never met for real systems, and rarely for properties, but understanding the consequences of partially meeting them helps to understand how this view can arise and help to diagnose its failures. (Wimsatt, 1985b, 1997, 2000, 2007). This view and eliminativism have similar appearances but different causes. They rest upon two different mistakes arising from conflating different reductionist activities. Their analysis is revealing about methods of discovery and problem-solving used by reductionists.

Analysis of these methods, their strengths and their limitations—calibrating them as tools—deserves close attention, and reductionistic heuristics are the primary focus of the second part of the paper (Wimsatt, 1980, 1985a).This is perhaps the proper scope of what is mis-named methodological reductionism, and far more of use to practicing scientists is to be learned from it than from traditional debates about reductionism.[1]

In closing I shortly consider the significance of this approach through heuristics to what are traditionally taken as more ontological or epistemological problems as a broader paradigm for how to approach philosophical problems.

1 The failure of a unitary structuralist account of reduction

From the mid-1930's to the mid-1970's, philosophers commonly regarded reduction as a relation between theories, expressed in terms of theoretical vocabularies, laws, and "bridge principles". Nagel's (1961) influential analysis reflected contemporary orientations moving from logic and formal systems to the "language of science": one theory (or part of it)[2] reduced to another if theoretical vocabulary for its entities and properties were definable, and its laws (logically) derivable from that of the other—connected by empirical identifications, correlations, or reconstructive definitions. (Nagel's additional pragmatic conditions were seldom noted—formal properties were seen as central.) Schaffner (1967, 1974, 1993) extended Nagel, allowing approximations and "strong analogies" in connecting theories which didn't match exactly, agreeing in some predictions but diverging in others: when one theory succeeded another,

---

[1] What is called "methodological reductionism" in the philosophical literature could be better named, "wannabe reductionism". It appears to be the view that we don't know whether reductionism is correct, but let's pursue our research as if it were. Fine! But then we are never given any hints as to how we should act in the laboratory, or what strategies we should follow in building our models. And these writers appear to have no interest in finding out. So in their actions, they reveal themselves as more metaphysician than reductionist, and not particularly relevant here. My aim is to flesh out the forms of what one actually does when using reductionist methodologies. Only then is it appropriate to reassess the metaphysical and epistemological claioms made for it.

[2] Although it is entirely consistent with the traditional Nagel model to consider reduction of parts of theories, this was more practiced than discussed. In spirit, it would have admitted less than total success in pursuing "the unity of science". Writ larger, it heralds patchwork, piecemeal and other "disunity" views, though it is in effect anticipated and richly delineated by Waismann as early as 1951. Wimsatt 1976b is another early elaboration of this view.

or a higher level theory was explained and corrected by a more exact lower-level account. Successive reductions to the most fundamental theory or lowest compositional level supposedly indicated the derivative character and *in principle* dispensibility of the things reduced—or so earlier philosophical accounts claimed.

In principle claims in an empirical area should invite suspicion: whatever else they accomplish, they reliably indicate that it hasn't (yet?) been done in practice. And how are *in principle* claims established—outside of logic or mathematics? We often hear in the same breath talk of the "computational" world view (e.g., Dennett, 1995), We know what it means to be computable—in basic operations mappable onto the natural numbers. But what can this mean in the empirical sciences? Can we specify the alternative possible states of matter in any well-ordered and denumerable way that *also* maps productively onto our theories at different levels? (Dennett sketches the first—a DNA sequence analogue to Borghes labyrinth/library, but gives no hint as to how to do the second: how the relevant similarity classes realizing higher level properties and regularities are to be found.)[3] Do scientists use in principle claims? If so, how? If not, what else do they do? "Methodological" reductionists, paradoxically, seldom discuss methods—no actual reductionistic problem-solving heuristics, or anything from the supposedly irrelevant "context of discovery". These are bothersome lacunae: if a scientist is reductionistic, shouldn't this affect their practice? If meaning reflects use, then practice should be key.

The unitary Nagel–Schaffner account of reduction has dissolved, leaving a polyphonic disunity. Studies from different sciences have become needlessly decoupled, seldom citing one another, but agreeing in focusing more on actual scientific practice. Wimsatt (1979) and Hooker (1981) provide extensive earlier reviews, while Allchin (1998), Bechtel and Richardson (1993), Darden (1991), Lloyd (1994), Sarkar (1998), Schaffner (1993), Waters (1994), and Wimsatt (1976b, 1987, 1992, 2006), study diverse cases, with often complementary results. Biology is a worthy methodological paradigm: an unproblematically compositional science, it has seen the richest complexity and greatest diversity of mechanistic approaches. In evolutionary, ecological, and behavioral studies, and increasingly in developmental genetics, we also see the growing sensitivity to context and interaction required to dealing with embodied and embedded minds and social entities. Biology also connects with the physical sciences, not only compositionally through chemistry, but also in the use of rich mathematical tools. Callebaut (1993) tracks the recent influence of biological thought and practice on philosophers of science. This account aims at the nature of mechanistic explanations throughout the cognitive, social, and physical sciences and engineering, and the cognitive tools used in their synthesis, starting with inspirations and paradigms from biology.

---

[3] This is not a worry about supervenience, but about level-blindness. Causal equivalence classes expressible compactly and projectably in terms of lower- or intermediate-level properties may exist (like the classes of particles falling into similar momentum classes in the derivation of the Maxwell–Boltzmann distribution of statistical mechanics), but in Dennett's library, we are provided with no lower-level way of identifying them. And for a mapping from DNA sequence space to phenotypic properties like Dennett envisions, we now know that there cannot be one without bringing in proteomics and the compositional dynamics of the cell (Moss, 2002) and the spatial arrangement and compositions of biological structures dynamically interacting to acquire influences and structure from the environment "all the way up."

2 Differentiating types of reduction for a richer vision

The perceived unity of reduction was an artifact of focus on structural or logical rather than functional features, when interests in reduction served foundationalist aims of increasing philosophical rigor, epistemological certainty, and ontological economy. These philosophical goals rarely matched those of scientists (Schaffner, 1974), who pursue not one but two different kinds of 'reductions' (Nickles, 1973; Wimsatt, 1976b). *Inter-level* or mechanistic *reductive explanations* serve fundamentally different functions than *same-level* reductive *theory succession*, resulting in structural differences between them that had gone unnoticed.

Inter-level reductions are compositional. They localize, identify, and articulate mechanisms that explain upper level phenomena relationships and entities. Reductive accounts in complex sciences are usually inter-level: explaining the behavior of gases as clouds of colliding molecules, of genes in terms of the actions of DNA in its milieu, or Mach Bands in the visual field in terms of lateral inhibition in neural networks.

Intra-level reductions are common in mathematically expressed theories and models. They localize formal similarities and differences between earlier versus later or more approximate versus more exact theories of the same phenomena through mathematical transformations. This aids succession and elaboration of the later theory and delimits conditions for safe heuristic use of the former. Insufficient similarities can frustrate intra-level reductions, and engender replacements or elimination of posited entities.

Most accounts have conflated these two kinds of reduction. Their respective use of identificatory versus similarity relations produce many structural differences discussed below. Scientific reductions of *either* sort are not the global and complete systematizations traditionally envisioned by philosophers (but see Waismann, 1951). They are usually conditional, partial, local, and context dependent—for inter-level reduction, on the specific mechanisms involved, and their associated *Ceteris paribus* conditions; or for successional reduction, on the character and conditions of approximation used.

Aggregativity , a claim that "the whole is nothing more than the sum of its parts", is the proper opposite to *emergence* (Wimsatt, 1997, 2000, 2007). Aggregative relations are compositional, like *inter*-level reductions, but meet additional special and very powerful conditions. So they are rarely satisfied. Aggregative system properties are degenerately simple cases of reduction where the organization of parts doesn't matter: they are unaffected by organizational rearrangements and have no mediating mechanisms. Various forms and degrees of partial aggregativity are much more common and interesting. In all systems a few properties (like mass) are aggregative, and some (usually many more) are not. Mechanistic models often *start* with many aggregative simplifying assumptions, but we add organizational features to increase their realism and explanatory power, and the respects in which they are aggregative disappear.

Older entities and properties in some *intra*-level replacements are dispensible because they fail to map even approximately into the newer theory. If aggregates were dispensible, it would be for different reasons: they map too easily; they are not required *in addition* because they are "nothing more than" the reducing things. But whole systems could be aggregative only if all of their properties were aggregative. That seems unlikely in the extreme. And focusing on the parts from which composite systems are aggregated does not make them go away. (The situation is symmetric: the parts are also "nothing less than" decompositions of the whole.) Moving from parts to wholes is, in effect, just a perceptual scale-shift. These different overlaps between

the two kinds of reduction and aggregativity invite confusions between them—e.g., the mistaken beliefs that in inter-level reductions, reduced things are eliminated, or are nothing more than the sum of their parts. I now consider these three kinds of relationships, particularly the second, in more detail.

3 Successional reduction

Successional reductions commonly relate theories or models of entities which are either at the same compositional level or they relate theories that aren't level-specific, such as the broader theories in the physical sciences. They are *relationships between theoretical structures where one theory or model is transformed into another* (usually via limiting approximations) *to localize similarities and differences between them.* Since such derivations involve approximations, they aren't truth-preserving. (Logicians and physicists take quite different things as 'derivations'.) The later, *more exact*, or more complete theory is said to *reduce in the limit* to the other. (Contrast inter-level reduction: *less exact* higher-level accounts are said to *be reduced to* lower level ones—crucial differences first noted by Nickles, 1973). So special relativity reduces in the limit to classical mechanics as $v/c \rightarrow 0$. This can be done either by letting velocity of the moving entity, $v \rightarrow 0$ (a "realistic" or "engineering" approximation for velocities much smaller than the speed of light, c) or by letting $c \rightarrow \infty$ (a counterfactual limit yielding Newton's conception of instantaneous action at a distance.) Nickles also noted this possibility of multiple alternative reductions between the same theories serving different ends.

Localizing similarities (by reduction) and differences between theories (in the transformations used and in things lost in the reduction) serves multiple functions aiding succession of the newer theory (Wimsatt, 1976b, extending Nickles, 1973). It legitimates use of the older theory where they agree (as $v \rightarrow 0$); co-opts its evidence for the newer theory; may establish conceptual connections between them (as $c \rightarrow \infty$); and locates places to pursue confirmation, test, and elaboration of the newer theory where they disagree. Limiting conditions show where the older theory is a valid approximation, and how rapidly it breaks down as the conditions are relaxed.[4]

Successional reduction is a kind of *similarity* relation between theoretical structures or their parts. So historical sequences of reductions would usually be intransitive, and for two different reasons: Most importantly, reductive transformations involve work. Successional reductions aren't done gratuitously. Since their scientific functions are served by relating the new theory to its *immediate* predecessor, there is rarely any point in going back any further. (Whatever the claims of historians, no scientist sees *scientific* relevance in tracing special relativity back to Aristotelian physics!) Second, differences accumulate, since changes in successive theories would most likely target parts of the theory dealing with different domains or ranges of phenomena. Understanding and localizing the net effects of cumulative re-mappings would exceed thresholds of comprehensibility, becoming too great to manage (Wimsatt, 1976a, b, 1979). So successional reductions are intransitive by default, but transitivities are also thwarted by cumulative differences.

If mappings are too complex to construct transformations relating *immediate* successors, successional reduction fails and the older theory and its ontology may be

---

4 This collapses a longer list in Wimsatt (1976b, Fig. 1).

replaced or discarded.[5] *So this kind of reduction can be eliminative* (of older theoretical entities and relations)—*but characteristically is so only when it fails* (Wimsatt, 1976b), Ramsey (1995) and Batterman (1995) elaborate such reductions, Sarkar (1998) approximations, and Wimsatt (1987) related uses of false models.

4 Inter-level reduction

Inter-level reductive explanation is the primary focus of the first part of this essay. *Inter-level reductions* needn't relate theories at all, though they may, as in the relations between classical thermodynamics and statistical mechanics. Darden and Maull (1977) urge that we often see creation of a single theory tying together two domains without having prior theories at either level. Such theories are usually multi-level articulated mechanisms rather than something with laws and boundary conditions (Wimsatt, 1976b). Reductive explanations are driven by referential identities (Schaffner, 1967; Wimsatt, 1976b) or localizations (Darden, 1991; Bechtel & Richardson, 1993)—not by theoretical similarities. Inter-level reductions explain phenomena (entities, relations, causal regularities) at one level via operations of often *qualitatively different* mechanisms at lower levels. So emergence claims fit here. Such "articulation-of-parts" explanations (Kauffman, 1971) are paradigmatically causal and reductive—in biology and elsewhere (Glennan, 1996, 2002). They are compositional—upper and lower-level accounts refer to the same thing, as a whole and as a set of configured interacting parts. Unlike similarity relations in successional reductions, these references are *transitive across levels*, though the explanations may not be transitive due to different explanatory foci at different levels. Mendel's factors are successively localized through mechanistic accounts (1) on chromosomes by the Boveri–Sutton hypothesis (Darden, 1991), (2) relative to other genes in the chromosomes by linkage mapping (Wimsatt, 1992), (3) to bands in the physical chromosomes by deletion mapping (Carlson, 1967), and finally (4) to specific sites in chromosomal DNA thru various methods using PCR (polymerase chain reaction) to amplify the number of copies of targeted segments of DNA to identify and localize them (Waters, 1994). Recent accounts emphasize the lability and context-dependent interactions of different genetic units and sub-units, moving away from tightly localized compact mappings. Accounts of the construction and action of genes become simultaneously more mechanically detailed while also becoming less reductionistic (e.g., Beurton, Falk, & Rheinberger, 2000; Moss, 2002). This is both predictable and explicable as models move with increasing sophistication from more atomistic and aggregative to more structured, interactive, and emergent (Wimsatt, 1997, 2007).

---

[5] What about entities or properties several smooth successional reductions back? If successional reduction functions primarily to aid in localizing similarities and differences between a new theory and its immediate predecessor, this issue should not arise for scientists, however much it might intrigue historians or philosophers. And the increasing difficulty of similarity judgements across greater reaches should make judgements variable and local. Few biologists would have problems with saying that the genes we study are the same as Mendel's factors, despite several revolutionary changes in how we conceive them since. Dominance and recessiveness would fare less well: Dominance is now treated as a matter of degree, and recognized as much more contextual (for epistatic effects) and relational (for multiple alleles), so an allele can be dominant for one allele and recessive for another. Bateson spoke of dominance and recessiveness as a "law", but Mendel recognized exceptions, and likely chose traits with strict dominant-recessive patterns to get clear results: intermediate hybrids are more easily misclassified.

Identities and localizations are powerful hypothesis generators for an articulatory reductionist, suggesting new predictions at one level from properties or relationships at the other—or others, since mechanisms at other new intermediate levels often emerge as important. Ample causal cues for how to construct explanatory accounts arise from the operation of the mechanisms through which the lower-level parts realize the upper level entities and phenomena (Wimsatt, 1976a, 1976b, 1992; Darden, 1991, Bechtel & Richardson, 1993; Glennan, 2002). *These mechanistic accounts not only explain how the connections work when they do, but give the resources for predicting and explaining when they should fail.*[6] "Correspondence theories" lack these resources. They look "empirically equivalent" to identificatory theories (as claimed by Kim, 1966) only in static *ad hoc* comparisons made after the fact or in the metaphysical timeless present. Work in scientific discovery motivates more realistic and revealingly different dynamical accounts of these processes.

Consider an example from the early history of genetics. This period is covered in close detail by Lindley Darden in her excellent book (1991) on theory construction and discovery. Mechanisms were not her focus then as they are now, so her exposition would match the account here more closely.[7] Earlier (Wimsatt, 1976a, 1976b) I focused on inter-level identities because I argued that the converse of Leibniz's law on the indiscernibility of identicals (If two things are identical, any property of one is a property of the other) drove the construction of inter-level theories and gave a principled way of interpreting claims of *in principle* translatability. Identities are still important, but the story is more complex. Scientific discussions often better fit processes of "localization", as noted by Darden (1991), Bechtel and Richardson (1993), and Sarkar (1998). I accept their amendations. In most cases, talk of identification comes down to localizations.

In 1966, Jaegwon Kim questioned the status of the psycho-physical identity theory, but his argument would apply to any multi-level reductive explanatory account. (He compares it with dualism and epiphenomenalism—perhaps idiosyncratically special to philosophical discussions of the mind-body problem.)[8] Kim claims that the only evidential basis for an identity is observed correspondences between elements or phenomena of the two domains, and if so, that an identity theory is empirically indistinguishable from dualism or epiphenomenalism, since both of these also depend upon (presumably the same) correspondences. But then identities would have no special scientific support. Indeed, they seem an unjustifiably strong inference entailing

---

[6] In this they parallel a function of successional reductions: to validate a less-exact theory by delimiting its range of applicability, and explaining how and why it breaks down.

[7] Her account provides much to draw on. I use also my discussions of the development of linkage mapping as modeling in 1987, and, focusing on the account of interference, in much closer historical detail, in 1992.

[8] When philosophers say "theory" in this context, they mean the class of theories satisfying the condition in question, or sometimes any instance of that broader class. My position fits none of his characterizations. A reductive mechanistic explanation should not deny the causal efficacy of or eliminate higher level entities or properties, including their powers to affect lower level phenomena. My account shares this feature with classical images of dualism, but is at odds with philosophical accounts of identity or epiphenomenal theories that locate all of the causal movers at the lower level. Another artifact of classical discussions is the use of correspondences or "bridge principles" to link entities, properties, and relations of the two domains—in this case an artifact of viewing the process of reduction as one of tying together two theoretical vocabularies with "translation" statements. Once one leaves the linguistic mode, this no longer fits. The idea that there are separable "bridge principles" seems a strange way to talk about the identifications and "localizations" of multi-level mechanists constructing an inter-level theory in the sense of Darden and Maull (1977) or any modern mechanist.

Table 1 Relations between chromosomes and unit-characters arising from the chromosome theory of Mendelian Heredity (From Darden, 1974, p. 114.)

Chromosomes	Unit-characters
1. pure individuals (remain distinct. do not join)	1. pure individuals (remain distinct, no hybrids)
2. found in pairs (in diploid organisms prior to gametogenesis and after fertilization)	2. found in pairs (in diploid organisms prior to segregation and after fertilization)
3. the reduction division results in one-half to gametes	3. segregation results in one-half to the gametes
4. PREDICTION: random distribution of maternal and paternal chromosomes in formation of gametes	4. characters from maternal and paternal lines found mixed in one individual offspring: independent assortment of characters
5. homologous characters **A** and **a** combine to give: **AA: 2Aa: aa**	5. Mendelian ratios of unit-characters **A** and **a** are: **AA: 2Aa: aa**
6. PREDICTION: abnormalities in chromosomes (e.g., abnormal sex chromosomes found by Bridges)	6. PREDICTION: corresponding abnormalities of characters (e.g., abnormal sex-linked characters)
7. paired homologous chromosomes in normal, pure-bred organisms	7. PREDICTION: dual basis for each character in normals (two alleles for each gene in normals as well as in hybrids)
8. chromosomes fewer than character number	8. PREDICTION: more than one unit on a chromosome; such units linked in inheritance
9. PREDICTION: the state of a chromosome related to a dominant character is different from the corresponding state of its homologue related to a recessive	9. dominant unit-characters are associated with a corresponding recessive
10. PREDICTION: recombination of parts of chromosomes or of the units on the chromosome	10. PREDICTION: more combinations of linked units than number of chromosomes

an open-ended list of correspondences, covering any property of either of the things to be identified. Kim suggests that to prefer an identity theory over a weaker "correspondence only" theory is only justified by an aesthetic preference for ontological simplicity. This fundamentally misrepresents the use of identifications and localizations in science, and turns a rich and empirically fruitful strategy into a non-empirical aesthetic demand, so I want to examine it more closely.

On Kim's "inductivist" account of identificatory statements, the open-ended number of correspondences entailed should recommend that we make identifications only after we have discovered many correspondences—if at all. When the ontological weight of all the Doppelgänger properties and correspondence rules gets too great, we allow ontological collapse and simplification. Is this what scientists do? Not at all. We find identifications and localizations made as soon as possible—tentatively and inexactly at first, but with increasing refinement over time. Are these scientists irresponsible? No! They are doing something else.

Consider the list of 10 "correspondences" between Mendelian factors and chromosomes of Table 1, taken from Darden's Ph.D. dissertation in 1974.[9] These are associated with the Boveri–Sutton hypothesis that Mendel's factors are located on

---

[9] Darden (1991) has a fuller list broken down into parts appropriate at different temporal stages of the debate. Starting with her initial list (given here), I found over the next decade another 43 predictions

chromosomes. [10] This might at first seem to confirm Kim's thesis. But it soon became apparent to both of us in discussion (Darden was then at Chicago) that when the Boveri–Sutton hypothesis was proposed, only a few of the correspondences could possibly have been known and used as a basis for the proposed localization. These are items 1, 2, and 3 from Table 1.[11] So what about the others?

4.1 Items 7 and 8, were predictions from chromosomes to factors

Item 7: It is not clear whether Mendel or Mendelians before 1902 recognized a diploid basis for traits in homozygotes. Mendel's own notation represented ratios as **A:2Aa:a**, rather than the modern form, **AA:2Aa:aa** suggested by Bateson in 1902, for reasons of symmetry. But if factors were located on chromosomes, there must be a dual basis in homozygotes.

   Item 8: Since there were many more characters than chromosomes, presumably many characters must map to each chromosome. (This was one of the most fruitful of predictions, since it leads naturally (with item 4 below) to predict the violation of Mendel's "law" of independent assortment for factors on the same chromosome, which should always be inherited together. And with qualification of that regularity suggested by item 10 below, this in turn provides part of the basis for linkage mapping).

4.2 Items 4, 5, and 9 would have been predictions from factors to chromosomes

Item 4: Mendel's law of independent assortment would require that if different factors are located on different chromosomes, they should assort independently of each other (breaking up the parental "sets"), so that there could be a random assortment of their characteristics in the offspring. Determining this would require visible differences in homologous chromosomes—not available in 1902. McClintock (with Corn) and Stern (in Drosophila) accomplished this by adding translocated parts from other chromosomes as "markers" in 1931.

   Item 9: Since dominant and recessive factors are different, this would predict differences in the corresponding homologous chromosomes.

   Item 5: The 1:2:1 ratio observed in the inheritance of factors must also have been present for chromosomes, but as with the prior two predictions, this could not have been observed until one had found chromosomes with distinguishable markings observable cytologically.

4.3 Two "correspondences", items 6 and 10 were apparently predicted ex nihilo

Material bodies are subject to various insults and modifications, so with a localizing identification, making factors physical, anything happening to either would happen to both. This provides two kinds of speculative predictions:

Footnote 9 continued
(many not "correspondences") that emerged from the analysis of chromosomal mechanics and linkage mapping up thru the early 1920's. See also my 1992, 1987.

[10] See Boveri (1902) and Sutton (1903). E. B. Wilson, whose lab Sutton had just joined, and where he heard Bateson talk of Mendelism, was clearly also on the way there.

[11] The issue with 5 is whether the different chromosome homologues could have been recognized to be in different states. No morphological differences could have been apparent in those chromosome preparations in the species studied at the time, and without this a 1:2:1 ratio for chromosome pairings would have been meaningless.

Item 6: Abnormalities in chromosomes should appear as abnormalities in characteristics, inherited with the same patterns as the chromosomes. Darden notes the occurrence of abnormal numbers and combinations of sex chromosomes studied by Bridges (in 1916). Boveri's (1902) observations on the consequences of abnormal somatic mitoses putting different numbers of chromosomes in some cell lineages, generating somatic mutations (such as mosaic mongoloidism) in local parts of the body also fits here, indicating the causal role of chromosomes in generating characteristics.

Item 10 (suggested by both Boveri and Sutton within the next year) If chromosomes could exchange segments, the linkage associations of factors should change, and there would be more combinations of linked units than chromosomes. (This anticipates recombination.)

Looking at this list yields two interesting facts: First, Predictions go both ways. Neither level is privileged as a source of knowledge or constraints on hypotheses, thus contravening a common belief among reductionists that knowledge flow is bottom-up (Wimsatt, 1976a, b). Second, especially striking is the large number of predictions arising from putting what is known of chromosomes and factors together, and predicting matches for any unmatched properties. The logical strength of identity claims (which appears to make them radically underjustified) provides a veritable fountain of new predictions—just what one wants if one looks for theoretical fruitfulness.

But *are* identificatory or localization hypotheses so radically underjustified? They seem so when one focuses as an "inductivist" on the many potential properties of objects and their relations and interactions with other objects. All seem potential topics of prediction—with an enormous number of degrees of freedom. But this is an unrealistically "disembodied" view of nature. We know a great deal about how the objects we study behave. Properties are not free to vary independently. If material objects or configurations are responsible for the upper-level phenomena, just a few observed correspondences provide a strong filter for arrangements or configurations of objects and interactions that could produce these results. These articulated constraints are what make mechanisms such powerful explanatory tools in the compositional sciences. The robustness of objects (Wimsatt, 1981) generates multiple richly articulated properties, and thereby provides most of the logical power arising from an identificatory or localizationist claim. This works surprisingly often even when some of the causal details are unknown or incorrectly specified. Higher level phenomena may be critically affected by some of the details or their realization but not affected by many or even most of the others (Wimsatt, 1994). So we have a nearly open-ended set of tests filtering hypotheses suggested as we learn about the system. This is what we find in the history of genetics.

5 Correspondences vs. articulations of mechanisms

Philosophers talk as if the data used to construct theories were sets of correspondences, but this is not the usual set of entities from which scientists assemble their accounts. They may provide early starting points, (as with items 1–3 of the table 1), but as we proceed we are given not correspondences, but pieces of mechanisms. We infer what happens when we perturb or work mechanisms in certain ways. From them we may infer or construct correspondences after the fact, but these correspondences

are richly context-dependent. This has several implications which severely undercut these philosophical accounts:

(1) Given a list of correspondences, they can be derived from the identifications and localizations of parts within a working set of mechanisms, but not conversely.

(2) New correspondences can't be predicted from existing ones without the identifications and localizations of parts within and use of the machinery. Any form of "correspondence only" theory would be predictively sterile. (Kim confuses metaphysical with scientific theories, but indeed, practice is so far from this abstract specification that one wonders whether any point is served by this form of metaphyical specification of the problem.)

(3) Identities, localizations, and lower level mechanisms can predict correspondences which are too unobvious or misleading to have been found in any other way: what is to be predicted at other levels is a function of understanding of the mechanisms. Thus Morgan found it necessary to warn in 1915 that although the genes appeared to play a role in specifying characters, each of which had spatial locations, the spatial location of the characters actually had nothing to do with the locations of the respective genes in the linkage map. As chromosomal mechanics develops, still richer interactive properties of multiple mechanisms are employed, suggesting increasingly diverse and unintuitive things to look for at each of the involved levels.

(4) Things will *look* more like correspondences after the fact because identifications and localizations assist in the segmentation of properties, behaviors, aspects, and entities (a point made also by Darden, 1991). Increasingly entrenched elements of the theory makes things look observational that weren't seen as such before. In early stages, linkage maps are obviously constructions, but later, new frequencies of recombination plus the maps allow the experimenter to just "read off" the location of newly found genes (Wimsatt, 1992; Kohler, 1994).

(5) Thru this activity, identities or localizations are refined constantly in ways that are clearly indicated and constrained by the data and the structure of parts of the machinery that are now robustly established. A static account of theory comparison totally misses this dynamics.

(6) A focus on correspondences turns attention away from pivotally non-correlational evidence. Roux's (1883) hypothesis about the significance of nuclear division figures (the complex dance of the chromosomes as they replicate in mitosis) provided crucial theoretical background for localizing Mendel's factors on chromosomes. The newly found cell-cycle and resources devoted to this repeated dance of the chromosomes was striking. Adaptationists assume that nature does nothing in vain, and the more complex the organic apparatus and processes, the more important the function served. The dance of the chromosomes got them regularly and reliably assorted equally into daughter cells in development. Roux used the details of the cellular parts and various mechanical hypotheses to argue that the point of mitosis must be to divide a very large number of elements contained in or on the chromosomes equally in two and to get them reliably sequestered into the two parts of the dividing cell. This influenced accounts of both chromosomes and hereditary elements from then on, but did not itself have an origin in correspondence data.

(7) The emphasis on correspondences also ignores contributions made by arguments ruling out other alternatives. Comparison of alternatives with the favored hypothesis to *generate* predictive tests was a rich feature of the development of chromosomal mechanics. Boveri's (1902) paper elegantly deployed the results of diverse

experiments to argue that the results of his "natural experiments" with simultaneous fertilization of eggs by two sperm could not be explained by any other features of the situation than generation of an abnormal combination of chromosomes. Other alternatives proposed in between 1909 and 1920 (Carlson, 1967; Wimsatt, 1987) generated a steady stream of responses from the Morgan school with new predictions (by my count, at least another 43) derived by comparing consequences of the proposed alternative mechanisms.

This all suggests that the focus on correspondences in philosophical theories of mind is misplaced, and that a better understanding of the development of explanations articulating mechanisms would better serve accounts of reduction and of the mind-body problem.

6 Identities and localizations

Localizations are logically weaker than identities. With two things that are identical—one and the same thing however differently known or accessed—*any* property of one is a property of the other. *But we can do interesting things with somewhat less. Exact localizations of two things to the same spatio-temporal region preserve all spatio-temporal properties of identities,*[12] *and thus all of their local causal relations arising from these spatial relationships.*[13] Localizations preserve identities for many kinds of properties, even when relatively imprecise.[14] Spatio-temporal location and contiguity are such causally rich factors in determining an entity or property's causal role, that they are also rich indicators of interaction when exploring how a system is causally organized to produce the higher-level properties of interest.

Identities or localizations are transitive when compositional claims preserve boundaries, and *functional localization fallacies often result when they don't*—e.g. Moss (1992). (We may move boundaries for good reasons and explanations may be intransitive for other reasons—e.g., different interests at different levels.) But commitments may be revealed indirectly: explanatory transitivity (and its tie to compositional

---

[12] There is a hidden scale-dependence here: "same spatio-temporal location" in a scientific investigation is not to a spatial or temporal point, but to a bounded S–T region. If boundaries differ, or are known with different precision, possibilities of mismatch arise. But even approximate localizations often preserve more robust property identifications, or may allow identifications still regarded as sound as qualified due to the boundary mismatch. These provide targets for further refinement.

[13] However when indirect indices are used to estimate location, more distant causal interactions can become involved. Linkage mapping uses frequency of crossover events between genes to determine their linear ordering along the chromosomes and "map distances" between them. These map distances are subject to various mechanistically explicable metrical distortions of actual distances along the physical chromosome. These involve not only various intra-cellular interactions between chromosomes in crossovers, but such unintuitively connected confounding variables as the temperature during gamete formation (which affects crossover rate) and fitnesses of the various genotypes produced (which act as multipliers of the ratios of genotypes arising from genetic processes to give the genotype frequencies actually observed) (Wimsatt, 1992, 1987).

[14] Thus in dichoic listening experiments, mental processing is localized to one hemisphere or the other by determining differences in response time when the stimulus is presented only to one ear or the other. It utilizes different transit times of neural signals along paths of different lengths, but the differences are large enough that it does not matter if the neurons are not stretched along geodesics, or that the depolarization pulse may travel at different speeds in axons of different diameters, or experience local speed variations along the axon.

relations) is still reflected in the *modus tollens* form. Reductionists inevitably blame *failure*s to explain upper-level phenomena in lower level terms on incomplete, incorrect, or differently bounded descriptions of the relevant system at one level or another, generating mismatches.

Here we have a crucial difference between successional and inter-level reductions. Failed successional reductions may eliminate objects or properties of the older theory. But failed inter-level reductions have the opposite effect: they make upper-level objects and theory *indispensible*! How else to organize the phenomena? *Inter-level reductive explanation, successful or not, is never eliminative.* Eliminative (inter-level) reduction is a mythical invention reflecting older aims of ontological economy since abandoned (Wimsatt, 1979). Some discussions in philosophy of mind suppose that successful reductions of upper level phenomena, entities, and relationships are eliminative, but *no such beasts are found in the history of science.*[15] *Nor is there reason—in terms of scientific functions served—to expect it in the future.* I believe that contrary claims arise through conflations of successional and inter-level reduction.

Analyses of inter-level reduction presuppose (and should provide) correlative analyses of levels of organization, whose objects, properties, and relationships are articulated across the various levels (and perspectives). The characteristic properties of levels and perspectives are analyzed and explained in (Wimsatt, 1974, 1976a, 1994). These analyses show that robust (multiply-detectable) higher-level entities, relations, and regularities (Wimsatt, 1981, 1994) don't disappear wholesale in lower-level scientific revolutions. Transmute, add (or occasionally subtract) dimensions, or turn up in different ways—yes, but disappear—no. Eliminativism rests upon exaggerated or incorrectly described accounts of unrepresentative cases.

Explanatory mechanisms do not themselves suggest, nor are they directly legitimated by, exceptionless general laws operating at their level of organization. Such generalizations would require unmanageably complex antecedent conditions and be of uselessly narrow scope, or have unintuitively related disjunctive antecedents and be even less useful (Cartwright, 1983). With a tolerance for noise, we can do better—as nature does. Useful, simple, broadly applicable generalizations about composed systems are richly qualified with *Ceteris paribus* exceptions explicable in terms of mechanisms operating under an open-textured variety of applicable conditions (Waismann, 1951, Glennan, 2002). The exceptions are intelligible when ordered as ways of making the mechanisms malfunction, but these are unlikely to be exhaustively specifiable in advance. Mechanisms are not translatable into laws (Wimsatt, 1976b).

There are reasons for this mechanistic complexity. In systems formed by selection processes, a population of individuals varying in many properties must generate heritable fitness differences if evolution is to be possible. Selection works probabilistically on context-dependent fitnesses, arising from life trajectories filled with a mixture of regularities and contingencies, so systems under selection must have "noise-tolerant" architectures. But then the relevant targets of analysis are not exceptionless laws, but robust context-sensitive mechanisms with lots of tolerance but many exceptions nonetheless. Any "universals" would be what I have (1991) called "sloppy, gappy" universals, but they are better understood,with their contextuality and exceptions, by

---

[15] The case of the replacement of phlogiston by oxygen is often cited as an inter-level elimination. It was not. At the time of that conflict, both phlogiston and oxygen were described in molar terms, so it was a case of elimination in theory succession. The redescription of oxygen and other elements in atomic terms was a second revolution, coming after the Dalton atom in 1807.

understanding the mechanisms involved, their normal operations, and their characteristic modes of failure.[16]

7 Organization, multiple realizability and supervenience

Opponents of reductionism usually fear what Dennett (1995) calls "greedy reductionism": explaining upper level things directly in lower-level terms without intervening mechanisms mediating the emergence of qualitatively different phenomena—as Dennett says, "without cranes." Greedy reductionism goes with "nothing but" talk. Neurophysiologist Roger Sperry worried that to reductionists "...eventually everything is held to be explainable in terms of essentially nothing." (1976, p. 167). But this doesn't happen in explanatory reductions. One moves to successively smaller parts in successive reductions, but in each transition, much of the explanatory weight is borne by the *organization* of those parts into larger mechanisms explaining the behavior of the higher-level system. A trip several levels down leaves nested articulatory structures of successively broader reach as one goes from there up to the starting point. These structures mediate the appearance of the properties and relationships at higher levels. Delineating this organization is the major work of and provides the explanatory power of reductions. For this reason, they should perhaps be called articulatory analyses, or when they explain the appearance of qualitatively new kinds of properties and relations, *articulatory reductions*.

Those parts' properties causally relevant to the behavior of the larger mechanism also provide criteria for *multiple realizability* and *functional equivalence*—roughly, *any part(s) whose causal articulations realize those (functional) properties in that context would do*. This works at all levels. Elements in the same column of the periodic table have some similar properties due to similar occupancies of outer electron orbitals, allowing limited substitutability in some contexts. Thus oxygen and chlorine are both potent oxidizers, so one might be substituted for another in very limited contexts (e.g., both are used in compounds for cleaning and sterilization).[17] Several levels higher (presupposing the whole protein transcription machinery of the cell) codon synonymy in the genetic code yields the same amino-acid sequence for different nucleic acid sequences (with the correct t-RNA's and a long list of other usually unspecified conditions). Amino acids are often intersubstitutable for others within the same broad hydrophobic ("water hating"), hydrophilic ("water loving"), and "neutral" activity classes in protein primary structure without disturbing protein function. Assuming action of relevant chaperone molecules and other unspecified conditions to catalyze proper folding, proteins fold up in a three dimensional tertiary structure so as to put hydrophobic amino-acids inside (away from water) and hydrophilic bonds outside (nearer water). Tertiary structure folds non-contiguous regions of the primary

---

[16] Some of this is present in my 1976b, but there I expressed views that looked more sympathetic to law-like regularities. There I was opposing views that micro-level regularities would totally undercut and displace macro-level accounts. For that I need just to point to the robustness of the upper-level mechanisms. This is not to claim either that they are exceptionless or even nearly exceptionless, but just that they work sufficiently frequently and systematically to be selected.

[17] But in the living realm, this limited intersubstitutability doesn't go very far: oxygen-based metabolism utilizes oxygen in very many compounds and reactions, with each interaction making somewhat different demands on it which chlorine cannot satisfy, so chlorine is poisonous to oxygen-breathing animals.

structure back on itself, forming active sites where the protein interacts with other molecular complexes. Substitutions within these broad equivalence classes is generally okay as long as the substitution is not within an active site (where the different characteristics of the side chains of the specific amino-acids and detailed interactions on a finer scale can become crucial).

Notice the frequent parenthetical qualifiers in this discussion of rough functional equivalence classes: *For differentiated parts (and their correspondingly differentiated functional roles), rich contexts generate a host of required conditions at various levels for functional equivalence, and things may be functionally equivalent for some roles (or performance tolerances) but not for others, or "fine tuned" to different degrees for a given function* (The complexities of supposed multiple-realizability in neuropsychology are charted by Bechtel & Mundale, 1999). Supervenience and multiple-realizability do not pose problems for a reductionistic analysis. Rather they are intrinsic to it—inevitable consequences of stabler levels of compositional organization (as argued in Wimsatt, 1981, and more fully in 1994). The explanation of and specifiable limitations on multiple realizability is a natural product of articulatory reductions.

Supervenience seems mysterious in a philosophical landscape of exceptionless laws whose antecedent conditions and modes of multiple realization seem to engender unbounded open-ended and unpredictable complexity. But articulatory mechanisms give rich and counterfactual-supporting causal structure to these tendencies and qualifications, and the mystery disappears. Advocates of supervenience marvel at the lack of systematic translations. The false furor over these concepts is an artifact of the widely held assumption that one should deal with laws rather than mechanisms. (For more on mechanisms, see Glennan, 1996, 2002.)

8 Aggregativity, emergence and reductionistic biases

What if a property 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.* This is MULTIPLE-REALIZABILITY in spades. But it is so broad that we simply factor organization out of the equation and treat the system property as a simply additive function of the monadic properties of the parts. There are such properties—those picked out by the great conservation laws of physics: mass, energy, momentum, charge, etc. These properties *are* strictly aggregative. For them, "nothing-but-ism" is justified. In every sense the whole is nothing more (nor less) than the sum of its parts. As far as we know, that's all. Even the shape (or to sound more exciting, the stability!) of a pile of rocks is not aggregative, since it depends on how individual rocks are arranged. Aggregative properties 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. (These 4 conditions are elaborated and applied to different examples in Wimsatt, 1987, 1997, 2000, 2007.) For *these* "aggregative" properties, we *do* 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 unreasonable shortages in what we got back. (Wimsatt, 1997).[18]

Other system properties, i.e. almost all of them, depend upon the mode of organization of the parts and are thus emergent—and (to reductionists) also mechanistically explicable. (To most scientists, emergence and reduction are not opposites: non-linear, organizational, and context-dependent interactions are reductively explicable *and* emergent.) Inter-level reductions deal with such properties, so why is reductionism treated as if it were "nothing but-ism"?

The conditions for aggregativity require *invariance* of system properties under different decompositions, rearrangements, and aggregations of and appropriate scaling relations with additions or subtractions to the system's parts (Wimsatt, 1987). Invariance under some of these operations but not others, or under limited conditions, or within specified tolerances, makes partial aggregativity a multi-dimensional degree property. Thus aggregativity, a concept with near-zero extension, becomes a partial ordering scheme of great richness for classifying kinds of emergence and context-dependence. But this isn't just a scheme for classifying ontological types. Scientific theories attribute properties and relationships to entities of the theory. Changes in these theories often result in changes in the posited aggregativity of different properties of the system. In developing explanations we start with simpler models that simplify or ignore higher-order interactions. They are thus *more* aggregative than later richer elaborations. Almost no properties are aggregative in all respects for all decompositions, but many are approximately so for some. Such *decompositions* are particularly simple and fruitful: more nearly factoring systems into modular parts with monadic, intrinsic, context-independent properties. These decompositions make for fruitful theories. We tend to see these parts as robust instances of *natural kinds*, these properties as "natural", and too easily and incorrectly, regard such systems as "nothing more" than collections of these parts (Wimsatt, 1996, 2007).

Such decompositions show varying success for different problems. Decompositions with more solutions grab attention, and are more often overused (e.g., Moss, 1992). Powerful reductionistic problem-solving heuristics bias us towards underestimating or ignoring effects of context (Wimsatt, 1980, 1985a,b, 2007). We may accept "nothing but" statements which are really context-bound and approximate as if they were truly general and unqualified. Poorly chosen decompositions, or more commonly, good ones which are overused or overinterpreted, produce functional localization fallacies. These fallacies, (e.g., attributing a whole system property—or function—solely to an important part of that system), systematic biases, and conceptual confusions (Bechtel & Richardson, 1993) arise from overextensions of even very powerful decompositions.[19] The most striking example in modern biology is the hegemony of the "selfish gene". This grew out of a powerful insight (Hamilton's elaboration of kin

---

[18] Nagel (1961) discusses some of these criteria, but largely to point out the ambiguity of the phrase "the whole is equal to the sum of the parts". He does not use the 4 criteria as a systematic tool for determining dimensions of non-aggregativity as in Wimsatt, (1996, 2000, 2007).

[19] Probably nowhere is this more dangerous with what is for us the most robust and obvious individuals—individual persons. This indicates one place where a modest eliminativism might have some rational purchase. While an eliminativism of person-agents is incomprehensible, it is possible that in the ultimate view, persons might be shaved or thinned a bit, referring some properties to lower levels of organization, but more importantly, referring some properties to higher levels. "Perspectival focus" (see its discussion in part II below) *must* lead us to overestimate the autonomy of person-level agents.

selection theory—Hamilton, 1964) into a monolithic conceptual imperialism that has far over-reached its productive use.[20] Analyzing complex systems often requires simultaneous use of different decompositions, boundaries, and contexts—increasing the likelihood of errors arising when one focuses only on a single level or single class of causes. We need our heuristics, but must use them here with special care (Wimsatt, 1974, 1994). Ockham should have invented a safety-razor!

Philosophers start in the wrong place to understand most debates in the complex sciences, taking the levels of the reduced phenomena and reducing apparatus as givens. But there aren't always well-defined levels. It is often unclear how to localize a key phenomenon, generating arguments about its "real" level. (One cannot debate whether a reduction has happened when the topology of the theoretical landscape is so contested that one cannot even agree on what is to be reduced to what.) Such uncertainties are particularly common for properties of the mental and social realms, but are common even within biology. In the domains of complex functional localizations and perspectives (Wimsatt, 1976a, 1994, Fig. 2), even well defined perspectives can be so affected by rich boundary-crossing interactions that many localizations are contested and order degenerates into causal thickets. Memory—a property of molecules, neural circuits, tracts, hemispheres, brains-in-vats, embodied socialized encultured beings, or institutions? Is consciousness intra-neural, intra-hemispheric, intra-individual, or (what level) social? With well-articulated theory and better-mapped terrain, evolutionary biologists have argued for decades over the bearers of fitness and the level(s) for analyzing altruistic behavior: from genes thru chromosomes, cells, individuals, families, kin groups, to local populations (Wimsatt, 1980, 1981, Lloyd 1994). This is not just confused thinking, but a breakdown of concepts of well defined objects honed on simpler cases from the physical sciences. Should the fractionated human sciences be expected to have a clearer picture?

With these multiple boundaries and ambiguities, for any investigation, we must track changes at two bounding reference levels and in-between: (1) bottom—the lowest level at which parts must be characterized to explain the phenomenon of interest (so an explanation of "position effect" sees an operon as a genetic control structure and how it works, but not the quantum mechanics of the electron orbitals yielding bonding of the repressor molecule.) (2) top—placing system boundaries out far enough to include all relevant parts of key mechanisms. And the boundaries required to explain different aspects of the behavior of a part of a system may differ. Successful "multi-level" articulatory reductionist analyses utilize resources at various levels, spanning these to determine appropriate loci for objects, processes, and phenomena, and explicate their relations. With lower level progress, we often forget higher levels (Sarkar, 1998). Paradoxically, reductionists frequently must expand the boundaries of what was originally taken as the system for analysis to include structured aspects of its environment in order to complete their tasks—becoming, in a way, more holistic as they succeed.[21] Most reductionistic "successes" which "stay local" and avoid

---

[20] The first signs of the waning tide are to be found in various places, from the increasing interest in group selection arguments, increasing discussions (and discoveries) of cooperative behavior, to the increased recognition of the importance of epistasis and context in gene action, and the generative role of environmental conditions and larger contexts and patterns in the realization of developmental programs in developmental genetics.

[21] Two recent examples here are the (re-)discovery that the cell as a whole is the appropriate unit to understand gene replication (Moss, 2002), and Scott Gilbert's (2001) proposal that environmental causes of gene activation in development are sufficiently rich that "eco-devo" or even "eco-evo-devo"

such a boundary expansion are won only through redefinitions of the problem into a sub-problem solvable by looking at only a tractable subsystem of the original. This is acceptable when the redistricting is recognized, but only too often the successes are presented as if the original problem has been solved.

The original problem of heredity from Aristotle to the beginning of the 20th century involved identifying the presumably single elements, factors, or stabilizing conditions which were both responsible for the generation of traits in ontogeny, and for their intergenerational transmission. As what came to be known as "classical" or "transmission" genetics emerged through the work of Morgan and others in the early 20th century localizing "genes" to the chromosomes, it became a scandal that transmission genetics explained in rich detail combinatorial aspects of the transmission of traits without having anything to say about how they were generated. Morgan himself complained of this feature of Mendelism in 1909, even though in 1911, he executed his now-famous reversal to localize a factor yielding a "white eye" mutation to the X-chromosome. This led to the development of classical genetics in terms of the articulations of chromosomal mechanics in which patterns of inheritance and co-inheritance of traits were explained by and often predicted from features of chromosomes and their behavior.

Opposition to the Drosophilist's growing research program through the next decade continued to revolve around its lack of connection with development or mechanisms for producing traits. All of the major attacks and proposed alternatives by Bateson, Castle, and Goldschmidt, the three best known critics of the program, were attempts to reintroduce cytological or developmental roles for the hereditary factors. They were not alone. The success of transmission genetics without solving this problem generated an uneasy equilibrium that was resolved only when Muller (1922) pointed out that the different mutations they studied (with radically different disruptions of function) were equally heritable. So the causal pathways of the development of traits had to be distinct from the causal pathways of intergenerational transmission. Thus was the quasi-autonomy of transmission genetics secured, and investigations of development (already moving away on their own) seemed plausibly disconnected from the problem of heredity.

There was nonetheless a particular delight when the Jacob–Monod post-DNA model of gene regulation in the *lac*-operon suggested a beginning handle on how gene expression was adaptively controlled. Explaining development was still far off, but at least this was a plausible beginning. The "modern synthesis" of evolution and genetics had taken place without the significant inclusion of development. This unnatural rift has been righted only with the emergence of "evolutionary developmental biology" in the last two decades. Here restricting the boundary of a biological system under study gave an apparently successful local reductionistic solution, but one with a puzzling separation of heredity and development. This is only now being reversed in a boundary and disciplinary expansion which creates a reintegration of data and disciplines which is broader, richer, and tighter than heretofore imagined, one that re-recognizes the causal roles of higher-level complex entities.

We now turn to the problem-solving heuristics which are the tools of reductionistic analysis. These same tools, sometimes lightly modified or supplemented by others appear to be used in sophisticated reductionistic analyses across the

---

Footnote 21 continued
(adding ecology to evolutionary developmental biology) is the correct conjunction of disciplines for many of what were taken as problems in developmental genetics.

disciplines, pointing to a methodological integration by problem-types which cross-cuts and undercuts at least the more extreme claims to the disunity of science (Wimsatt, 2007).

9 Part II: Reductionism and its heuristics

Taking a functional account of scientific activities as purposive and end-directed seriously naturally leads one to want an account of scientific problems, research programs, procedures, techniques, heuristics, solutions of various kinds,[22] and anything else that "wears its purpose-relative character "on its face", as these do. This leads to a wide literature which I won't survey here, except to note two authors worthy of closer attention: Darden's underappreciated (1991) book (already drawn on above) has relevant material all through. And Tom Nickles has written many interesting and important papers on a cluster of key ideas in this area: on heuristics (distinguishing between "epistemic" and "heuristic" appraisal, forthcoming), on characterizing problems (1981), his intriguing, right-headed and important idea of "problem reduction" (1976), and normal science and case-based reasoning (2003). (These are just a sample of a wider corpus). His 1976 paper (which connects traditional talk of reduction centrally with heuristic ideas) deals at a slightly higher level (a local aim of science) with the same issues raised here in talk of heuristics (specific tools used to accomplish "problem reduction"). See particularly property #4 of heuristics below.

9.1 The concept of a heuristic

I use the term "heuristic" in a broader sense than usual in the artificial intelligence literature, but closer to Herbert Simon's original (and less formal) use when he got the term from mathematician Georg Polya's *Patterns of Plausible Inference* (1954). Polya used the term to denote tactics that were often useful in specific contexts to get solutions. These tactics were not derived directly from a theoretical understanding of the solution space, but were "informal" "rules of thumb" derived perhaps from a partial theoretical understanding, or perhaps which were known only empirically to work often enough to make them worth trying.[23] In AI, heuristics are formal procedures or inference rules, as is appropriate for components in computer programs, which can play this role. Simon (1966) suggested that they constituted a "logic of discovery".[24] But with the properties below it is only a small move to see, e.g., the extended waggling bait-like tongue in the open mouth of the angler-fish as a heuristic procedure for luring smaller fish which become prey. And to a migrating animal, the presence in a strange environment of a conspecific is a useful

---

[22] This is not the place to survey solution types, but it is a worthy endeavor, e.g., (e.g., analytic solutions arising from formal methods, and numeric solutions of various kinds, approximations (giving a solution within a given tolerance, when an exact solution is not required), local solutions (those working, usually as approximations and only in a neighborhood or under a range of conditions). But lest this begin to sound like a list with plausible closure, a localization is a kind of solution, but so is an existence proof, and there are also empirical "existence proofs" (e.g., finding a phylogenetic "missing link"). Indeed there may be as many solution-types as problem types.

[23] Simon (1973) offers an account of "ill-structured" problems which I find revealing and appropriate for the knowledge backing the use of at least many heuristics.

[24] This was not a logic in the sense of a deductive argument form. Rather it was a relatively efficient form which (because of its ease and efficiency) constituted a 'rationale of discovery.'

indicator that the environment is suitable: organisms characteristically stay longer in places that are suitable for them, and quickly move on in environments which are not. ("Suitable environment" is species-specific. Thus the particular value of a conspecific—of any kind, whether potential mate, competitor, or neither.) Here a heuristic is a pattern for action wherein a kind of action (behavior) is characteristically undertaken under specifiable kinds of circumstances to achieve an end, or as part of a larger plan designed to do so, and succeeds sufficiently frequently to justify its use. These "action-patterns" have characteristics explaining why they are so widely adopted, calibrated, and combined in larger methodologies to correct for biases and increase robustness. These properties are shared with adaptations, so heuristics are plausibly seen as problem-solving specializations of a broader class of adaptive tools.

I believe that all heuristics share the 6 properties below.[25] These properties make sense and articulate well with one another. The first three are required both to see how heuristics differ from truth-preserving algorithms in a traditional deductive methodology or a more current "computational" one and to understand their particular strengths. They stem from an analysis of reductionist methodologies to explain some puzzling blindnesses in the construction of incorrect models of group selection by leading theoreticians (Wimsatt, 1980). This account was generalized in Wimsatt (1985a,b), (adding item 4 suggested by Bob MacCauley). Item 5 is added in Griesemer and Wimsatt (1989), and item 6 in Wimsatt (2006). Item 6 draws a closer link between evolutionary accounts of adaptation and heuristics, gives consilience with Hull's "populational" view of scientific theories, and explains a puzzling difference in classifying heuristics between Lenat's (1982) account and mine, in which Lenat recognized many more variants than I would have.[26] This list may not be exhaustive.

9.1.1 Six properties of heuristics:

(1) By comparison with truth-preserving algorithms or other procedures for which they might be substituted, heuristics make *no guarantees* that they will produce a

---

[25] I do not claim or care whether they are necessary or sufficient. A computer scientist I know claims a heuristic is just a truth-preserving algorithm not yet proven to guarantee results. His example was the method of "alpha-beta" pruning of sub-optimal branches in a search tree for chess. After years of supposing that it did not guarantee results, a proof was found that it did, (never lopping off the branch that contained the optimal solution). For a heuristic that works often in a class of cases, it is tempting to ask whether additional features could be added as antecedents to guarantee success in that more restricted class. This should work in a deterministic world but that does not guarantee that the list of conditions would be short or have any natural unity. Unless it did, it would violate condition 2. So I choose not to take this option, which would also rule out taking probabilistic rules as heuristics.

[26] Lenat had a large number of heuristics procedures in his work on theorem-proving, far more than I had discovered, but when I looked more closely, I saw that he had about 60 different heuristics where I saw just one with slightly different antecedent conditions. His multiplication of heuristics was forced by the formal AI simulation context (and the famous "literal mindedness" of computers). It led me to wonder whether the computational context was the right one for elaborating heuristics after all. If one agrees with Lenat that there is a very large number of heuristics, adding condition 6 gives consistency with the intuition that there is just one (or a far smaller number) of distinct heuristics which are more informally understood. Perhaps our molar informal sense of heuristic (with an only intuitively specified range of application) is incomplete, and actually properly grounded by Lenat's more precise delineations. But I am more inclined to regard our phenomenological reasoning as analog, perhaps to be explicated by models which are at least part connectionist, and that more precise delineations are just fine tunings evolved in a satisficing manner (Simon, 1955) when they become sufficiently common and important. (This would better fit their evolutionary origins.)

solution or the correct solution to a problem. A truth-preserving algorithm correctly applied to true premises *must* produce a correct conclusion. But one may correctly apply a heuristic to correct input information without getting a correct output.

(2) Heuristics are more "*cost-effective*" than the procedures for which they may be substituted, (commonly truth-preserving algorithms, but possibly other heuristics less suited to the situation), in terms of demands on memory, computation, or other limited resources, under their (possibly incompletely specified) conditions of use. (This of course is why they are used.)

(3) Errors produced when using a heuristic are not random, but *systematically biased*. This has 5 parts: (a) The heuristic will tend to break down in certain classes of cases and not in others, but not at random.[27] (b) If we think of a heuristic as a kind of *mechanism*, understanding how it works should allow us to *predict* the conditions under which it will fail, and (c) may also suggest ways to fix it. (d) Where it is meaningful to speak of a *direction of error*, heuristics will tend to cause errors in specifiable directions, as a function of the heuristic and of the kinds of problems to which it is applied. (e) This systematicity of errors may leave characteristic *footprints* providing clues to its use in the derivation of a result, and aid in reconstructing the inference. (Wimsatt, 1980, analyzing data on errors in modeling assumptions in the units of selection controversy unearthed by Wade in his classic 1978).

(4) Applying a heuristic to a problem *transforms* the problem into a non-equivalent but intuitively related problem. (See Nickles, 1976 on "problem reduction"). But it is not the same problem, so beware: answers to the transformed problem may not be answers to the original problem. However cognitive biases operative in learning and science may lead us to ignore this possibility and assert or assume that we have answered the original problem (Wimsatt, 1985a,b), leading to inflated or premature claims about the power of an approach. In other contexts it may lead us to underestimate the creativity of a solution, thinking that it was already given or predetermined in the original problem formulation (so that we "had to follow the rule" in the way that we did (Wittgenstein, 1962; Lakatos, 1978)).

(5) Heuristics are useful *for* something—they are *purpose relative*. Tools designed as particularly useful for one purpose may be bad for another (Levins, 1968). This may help to identify or predict their biases: one expects a tool to be relatively unbiased for the applications it was designed for, but perhaps quite biased for others. One might also expect that increases in performance in one area will often (but not always) be at the cost of decreases elsewhere.

(6) Heuristics are commonly *descended from other heuristics*, often modified or differentially specialized to work better in different environments. So they commonly come in families, which may be drawn upon for other resources or tools appropriate for similar tasks, and they may show recognizable descent relations. (This is reminiscent of the variety of blacksmith's tongs one finds, with identical or near-identical handles but different jaws specialized for gripping different objects, all to similar ends and using the same principles, and, of course, all produced by blacksmiths (Richardson, 1978))!

---

[27] I don't want to assume that there can't be a random component to the success or failure of heuristics, (e.g., generated by uncontrolled variations in circumstances) but only to argue that there are always non-random components. A heuristic whose trials in diverse circumstances comprised a homogeneous reference class with respect to probability of success would violate conditions (b) or (c) or both.

I now consider a large family of heuristics used in reductionistic problem solving. They have similar biases arising from the problem situation that they share, but different forms appropriate to their different local tasks and objects.

10 The origins of bias in reductionistic research strategies

I suppose that a reductionist analysis offers a lower level mechanistic account of a higher-level phenomenon, entity, or regularity. To do so, one commonly decomposes a complex system into its parts, analyzes them in isolation, and then re-synthesizes these parts and the explanations of their behavior into a composite explanation of some aspect of the behavior of the system. Decomposition and recomposition (Bechtel and Richardson, 1993) is a "near-decomposability" meta-heuristic for reductionist problem-solving. (Simon, 1962, Wimsatt, 1974, 1987). In using this approach, we use heuristic strategies with systematic biases that ignore or downplay the context-sensitivity of the results and the importance of the environment.

Reductionistic problem-solving strategies have systematic biases. Many of these arise in choosing or in modifying the choice of a system for study: Figure 1 depicts a studied system (solid boundary) with 3 inner subsystems, interacting in its environment with 4 other systems. But the system and its environment (the part of the world interior to the dotted boundary) is not fully described. The environment is drawn only partially overlapping these systems, and the simplifications in modeling guarantee that even in the "captured" parts, only some of the interactions are represented. Another important system with inputs affecting several of them and their interactions with the system studied is entirely left out. These lacunae may lead to inaccurate assumptions about inputs to the study system, and its effects on them (richly explored in simulated ecologies in Taylor, 1985)—missing alternative more natural or revealing choices for the boundaries of the system being studied.

But where do these biases come from? I believe that they are inevitable consequences of a reductionist methodology. So should we avoid reductionist approaches? No, or at least not necessarily, but we need to study the origin and nature of these
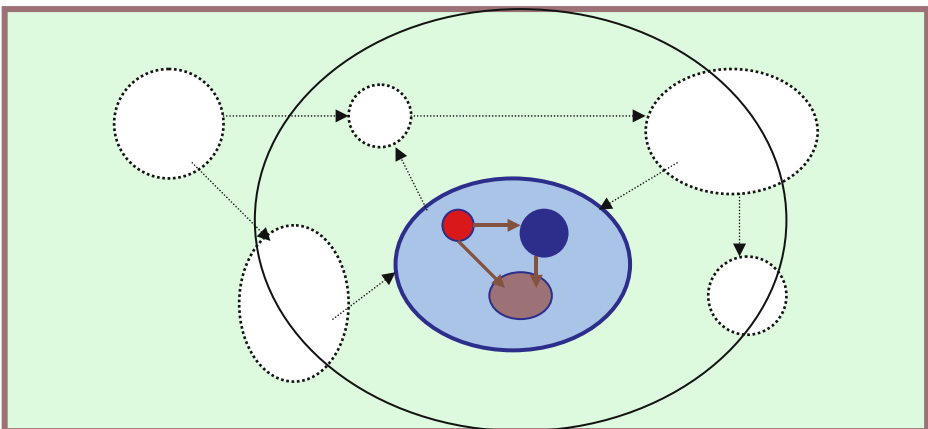

Fig. 1   Reductions tend to misplace boundaries and ignore what goes on outside

biases so that we can recognize and correct for their effects. I formulate their origin as a 3-step argument:

(1)   One starts by choosing, designating, isolating, or constructing a system for analysis. This immediately partitions the world of study into that system and its environment:

$$\{W \rightarrow [S, E]\}$$

(2)   We make Simon's (1955, 1996) "assumption of bounded rationality": any real world system is too complex to study in all of its complexity. We make simplifications—through selection of properties or objects for study, simplifying relationships between them, and assumptions about what variables must be controlled or randomized.

(3)   A reductionist is interested in understanding the character, properties, and behavior of the studied system in terms of the properties of its parts and their interrelations and interactions.[28]

From (3), a reductionist is primarily interested in entities and relations *internal* to the system of study. Simplifications will have to be made everywhere, but reductionists will order their list of "economic" priorities so as to simplify first and more severely in describing, observing, controlling, modeling, and analyzing *the environment* than in the system they study.

To see the asymmetry, imagine a reductionist who started by assuming that the system is homogeneous and constant. But then where are the parts? And what's to analyze? But many reductionists would begin by making just these assumptions for the environment. *This systematic bias operates relative to the system boundary chosen, so changing the system boundary will change what simplifications are made and appear acceptable. So reformulating the problem for different choices of the system boundary, and comparing the results generated using them may be one of the most fruitful checks to detect bias. This systematic bias is characteristic of all reductionistic problem-solving strategies including those discussed below* (Wimsatt, 1980, p. 235). Each is used because its adoption transforms the initial problem into one that is easier to analyze and to solve. Each is an application of the above schema for making simplifications to specific scientific activity. Each leads us to downgrade or underestimate the importance of factors in the environments, beyond the boundaries of the systems we study. This characterization of the problem-situation shows just how general (and how important) heuristics arising from this situation could be.

11 Common reductionistic heuristics

Let us consider some more specific reductionistic heuristics. They are grouped under the kinds of activities or contexts that occasion their use. All of these focus on inter-level reductive strategies, though some touch on issues of aggregativity.[29] Heuristics 1–3 and 5–10 first appeared in 1980.

---

[28] This inclusive description probably captures analytic methods in general, including many we would not think of as reductionist when they are not applied in compositional contexts.

[29] For detailed and rich consideration of issues in successional reduction and heuristics involving limiting case reductions, see Batterman (2001). He also covers some issues in formal inter-level reductions. Some of the functions of false models discussed in Wimsatt (1987) also are strategies for managing intra-level reduction.

11.1 Biases of conceptualization

(1) *Descriptive localization*: describe a relational property as if it were monadic, or a lower order relational property. Thus, e.g., describe fitness as if it were a property of phenotypes or genes, ignoring the fact that it is a relation between organism and environment. (This localization may be justified/facilitated, and its strong assumptions hidden, by fixing the environment, making it artificially disappear as a source of causally relevant variables—see complementary heuristics 6, 8, and 9 below.)

(2) *Meaning reductionism*: assume that lower-level redescriptions change the meanings of terms, but higher-level redescriptions do not. This reflects a kind of *atomistic essentialism*. Thus we suppose that the meaning of 'gene' is changed when we discover the structure of DNA, but that the meaning of 'iron' is not changed when we discover that it occurs as a crucial chelating ion in hemoglobin. The result: philosophers concerned with meaning relations are inclined to a reductionistic bias.

(3) *Interface determinism*: Assume that all that counts in analyzing the nature and behavior of a system is what comes or goes across the system-environment interface. This has two complementary versions: (a) *black-box behaviorism*—all that matters about a system is how it responds (with outputs) to given inputs; and (b) *black-world perspectivalism*—all that matters about the environment is what comes in across system boundaries and how the environment responds to system outputs (e.g., Fodor's "methodological solipsism" or Searle's Chinese room). Either can introduce reductionistic biases when conjoined with the assumption of "white box" analysis—that the order of study is from a system with its input-output relations to its subsystems with theirs, and so on. The analysis of functional properties, in particular, is rendered incoherent and impossible by these assumptions. (Beckner's classic (1963) analyzes similar cases arising from specification of system boundaries.)

(4) *Entificational anchoring*: Assume that all descriptions and processes are to be referred to entities at a given level, which are particularly robust, salient, or provide an apparently combinatorial basis for the construction of other entities and properties. This is the ontological equivalent of assuming that there is a single cause for a phenomenon, or single level at which causation can act. Thus the tendency to regard individual organisms as primary and more important than entities at either higher or lower levels (cf. methodological individualism for rational decision theorists and other social scientists. Similarly for genes for some reductionist neo-Darwinians.) Note that this is reductionistic only relative to higher levels.

Corollary 4a: This heuristic can also lead to a project of "level completion" (like the human genome project) where the aim of investigation is narrowed to a single level, arguing that a complete description of entities or phenomena at that level will be sufficient to solve problems at that and other levels. cf. *perceptual focus* (#19 below) and *multi-level reductionistic modelling*.

11.2 Biases of model-building and theory construction

(5) *Modelling localization*: look for an intra-systemic mechanism to explain a systemic property rather than an inter-systemic one. *Corollary 5a: Structural* properties are regarded as more important than *functional* ones (since functional ones require reference to embedding systems).

(6) *Contextual simplification*: in reductionistic model building, simplify environment before simplifying system. Thus the environment may be treated as homoge-

neous or constant (in space or in time), regular in some other way, or random. This strategy often legislates higher-level systems out of existence, (see discussion of the "migrant pool assumption" in models of group selection in Wimsatt 1980) or leaves no way of describing higher-level systemic phenomena appropriately.

(7) *Generalization*: When starting out to improve a simple model of the system in relation to its environment, focus on generalizing or elaborating the internal structure, at the cost of ignoring generalizations or elaborations of the environmental structure. *Corollary 7a*: If a model doesn't work, it must be because of simplifications in the description of internal structure, not because of simplified descriptions of external structure.

11.3 Observation and experimental design

(8) Focused *observation*: Reductionists will tend not to monitor environmental variables, and thus fail to record data necessary to detect interactional or larger scale patterns.

(9) Environmental *control*: Reductionists will tend to keep environmental variables constant, and will thus miss dependencies of system variables on them. (*Ceteris paribus* is regarded as a qualifier on environmental variables.) Mill's methods applied with this heuristic (vary the system variables one at a time while keeping all others (always including the environmental variables) constant) will yield apparent independence of system variables from environmental variables, though the right experiments won't have been done to establish this.

(10) Locality of *testing*: Test a theory only for local perturbations, or only under laboratory conditions, or only with specially engineered "model organisms" rather than testing it in natural environments, or doing appropriate robustness or sensitivity analyses to suggest what are important environmental variables or parameter ranges. (This is often a problem with equilibrium analyses when they fail to investigate behavior far enough from equilibrium to detect alternative equilibria or non-linear behavior. As such this is not peculiarly reductionistic, except when used to avoid appropriate studies of environmental variation.)

(11) *Abstractive reification*: Observe or model only those things common to all cases; don't record individuating circumstances. Losses: (1) sense of natural (or populational) variation; (2) lose detail necessary to explain variability in behavior, or exploitable in experimental design. (Raff, 1996 notes that evolutionary geneticists focus on intraspecific variability, while developmental geneticists focus only on genes which are invariant within the species, producing problems of methodology and of focus when trying to relate micro- and macro-evolutionary processes with each other or either with development. Similarly, cognitive developmental psychologists tend to look only for invariant features in cognition, or major dysfunctions, rather than populational variation, and design experimental controls so as to suppress it.)

(12) *Behavioral regularity*: The search for systems whose behavior is relatively regular and controllable may result in selection of unrepresentative systems which are uncharacteristically stable because they are relatively insensitive to environmental variations.[30]

---

[30] Mendel's selection of 7 out of 22 characters which are relatively constant and insensitive to the environment (done to increase reliability in scoring character traits) probably unintentionally resulted in unconscious selection against epistatic traits that would have shown variable expression in different

(13) *Articulation-of-Parts (AP) coherence*: Assuming that the results of studies done with parts studied under different (and often inconsistent) conditions are *context-independent*, and thus still valid when put together to give an explanation of the behavior of the whole. This can apply not only to intra-organismal parts and systems but at other levels—e.g. to the combination of behavior of individual species to analyze the behavior of an ecological community.[31]

11.4 Functional localization fallacies

(14) Assuming that the function of a part is to produce whatever the system fails to do when that part is absent, (e.g., spark plugs as "sputter suppressors"), or produced when that part is activated or stimulated. More generally, the error involves *reifying added or subtracted behaviors* of the system *as functional properties* of the manipulated unit. Gregory (1962) notes that things not done with lesion or deletion experiments may simply be the most obviously affected (rather than the most important). The part may have more importance to functions which are strongly canalized or with deficits not revealed under the testing conditions. Even if a part does realize the function attributed to it, it commonly does so only against a background of activities by other interacting components. Judgements of functional modularity are often insufficiently justified.

(15) Assuming *simple 1–1 mappings between* recognizeable *parts* and *functions*. This leads to problems in two ways: (1) ignoring pleiotropy: stopping search for functions of a part when you find one (e.g., the newly discovered region of hemoglobin implicated in NO^+ transport, because it was assumed that *the* function of hemoglobin was oxygen transport); (2) ignored division of labor (when a part's necessity is shown thru deletion studies, etc.) (missing other parts' roles in the hypothesized function because they are part of the constant context, so they are always there to provide it). See #9 above. Given the frequency with which we today see talk of "discovery of the gene for X"—a tendency blamed on classical geneticists, it is worth noting that from the beginning, Thomas Hunt Morgan and all of his group were more careful in describing the relation between genes and characters than many geneticists today:

> "*It is important to note that mutations in the first chromosome are not limited to any part of the body, nor do they affect more frequently a particular part. The same statement holds equally for all of the other chromosomes. In fact, since each factor may affect visibly several parts of the body at the same time, there are no grounds for expecting any special relation between a given chromosome and special regions of the body. It cannot too insistently be urged that when we say a character is the product of a particular factor we mean no more than that it is the most conspicuous effect of the factor.*" (Morgan, 1916, *italics in original*.)[32]

Footnote 30 continued

genetic backgrounds. This fortuitously increased the plausibility of his model, which ignores the possibility of epistasis (personal observation).

[31] Jeffrey Schank (1991) first noted this problem and that checking it gives a non-trivial use for computer simulation. He elaborates this further (2001), suggesting (personal conversation) that even this may not be sufficient, and that hybrid simulation/robotic methods may be required for greater realism.

[32] Given Morgan's early care when compared with some modern geneticists in avoiding simplistic functional localization talk, there is some irony that his claims denying any spatial correspondence in mapping from chromosome location to expression in the body—claims that would have seemed most

(16) *Ignoring interventive effects* and damage due to experimental manipulation. First noticed as relevant in neurophysiological studies, it occurs also in many other places (e.g. marking specimens in mark-recapture studies may affect their fitness).

(17) *Mistaking* lower-level *functions for* higher-level *ends*, or misidentifying the system which is benefited. Thus medical texts will often treat survival of the individual as if it were the end served by organic design, though to an evolutionary biologist, survival of the individual is just a means to the generation of fertile offspring. Level errors in assessment of function are common in units of selection controversies—either of the apocalyptic variety as with Dawkins (1976) who denies all units of selection at higher levels than the gene, or for eliminative reductionists, who want to deny the existence or significance of large domains of cognitive function. There are legitimate concerns as to what is the appropriate level for analysis in both disputes, but the extremists are commonly seriously wrong.

(18) Imposition of *incorrect* set of *functional categories* derived from making the wrong assumption about what the system is doing. (Common in psychology in contexts where it ignores development, ethology, ecology, or evolutionary biology.) Gregory (1962) has a mixture of serious and jocular examples, as when he considers whether the exhaust pipe of a car could possibly function as a hair-dryer.

11.5 Other important biases

(#'s 4, 11 and these can generate either reductionistic or holistic biases in different contexts.)

(19) *Extra-perspectival blindness or perceptual focus:* Assuming that a system can be exhaustively described and explained from a given perspective because it has been very successfully and powerfully so described. (Not all problems of biology are problems of genetics, or of molecular biology, physiology, or anatomy (to cite other past excesses) and (as we can now see from a safe distance), not all problems of psychology are problems of behavior. Perceptual focus can artificially inflate the number of properties attributed to a level of organization. Thus, the individual psyche—though perfectly real—has attracted social properties through improper (reductionistic) functional localization fallacies, and other phenomena better explained at lower neurological levels. This bias interacts with #4 to give *extra-level blindness*, which can be counteracted by doing *multi-level reductionistic modeling,* in which a process is modeled at several levels (commonly a level up and a level down from the focal level using reductionist heuristics like those here), with results which are *then* cross-checked to see if modeling assumptions made at the focal level make sense at higher and lower levels.

(20) *Tool-binding:* Becoming sufficiently bound to a specific (usually very powerful) tool that one chooses problems for it, rather than conversely ("The right job for the organism", rather than "The right organism for the job"!) This applies to theoretical models and skills as well as to material tools and "model organisms". This may be an efficient division of labor if mastery of the tool is very demanding, or if an unusually large accumulation of data concerning it makes it particularly useful (as with Drosophila and the other model organisms). It is problematic only when it facil-

---

Footnote 32 continued

secure—have fallen before the marvelous correspondences discovered with the HOX gene complexes. Morgan's warnings about the many-1 mappings in both directions between genes and characters have survived unchallenged however.

itates errors #4 or 16. If we look at the kind of systematic information a general tool provides about a broad range of objects, this may constitute a *perspective* (Wimsatt, 1994).

12 Part III: A heuristic methodology for philosophy

For most of the last century, formalist and foundationalist ideas have substantially influenced our own conceptions of what we are doing as philosophers, under the aegis of logical empiricism and analytic philosophy more broadly. This influence has persisted in less obvious ways even after their hegemonies have ebbed. Given the revolutionary progress that appeared in physics and logic about 100 years ago, this is not surprising. But the last third of the century has belonged to biology, and particularly to evolutionary and genetic perspectives on nature. If we look at organic design we see quite different principles operating (Wimsatt, 1981, 2007)—with robustness and error tolerance secured using designs that are contingent, and contextually (rather than generally) sensible and cost-effective. There is every reason to think they should apply also to our faculties of reason and our constructions of all sorts. Heuristics are a species of cognitive adaptations, and the study of heuristics both suggests and calls for an entirely different viewpoint in constructing philosophical methodology.[33] This needn't replace the current broader philosophical inspiration by various logical and more formalistic paradigms, but seems an appropriate complement to it that should give us broader reach and more appropriate tools for a whole class of problems where variations may be familial rather than accidental. These problems should be expected to crop up for products of evolutionary processes. This may sound uninterestingly narrow, but I include in this scope the three great designed systems of philosophical inquiry: *body, mind, and society*.[34] The heuristics I propose for this class of problems would include;

- Instead of looking for inexorable arguments, we look for robust tendencies; and for conditions under which those tendencies are more likely to be realized.
- Instead of looking for truths, we study errors, and how they are made.
- Instead of looking for context-free inferences, we study commonly used but context-sensitive ones.
- Instead of classifying them as invalid because content or context specific, we calibrate them to determine the conditions under which they work, or are "locally valid".
- We may look for argument schemata, but look for broad conditions where they are likely to work [like looking for the range of validity of a model], rather than trying to demonstrate their universal validity. In this way, we can espouse the use of formal methods, but as a tool for appropriate problems, not as architectonic principles.

---

[33] The closest parallel to this new turn is the inventive use by Robert Nozick of Herbert Simon's (1973) article, "The Structure of Ill-Structured Problems" as proposing methods for philosophy in trying to generate structure for philosophical problems. This move is in an appendix to Nozick's insufficiently well-received book, *Rationality* (1996).

[34] It is a great irony that many philosophers who are refugees from engineering see that as a close call from which they have happily escaped, and from which they learned nothing. And then they turn to the analysis of these three great designed systems without a clue about evolution or design, and convinced that they don't need any. See Wimsatt 2006.

- Counterexamples become revealing sources of information about limitations of a model, or suggestions for probing its depths; in either case, a tool to refine the model, not an argument for trashing the system, or something to be swept under the rug.
- It is often as important to try to refine, extend and generalize counterexamples as it is to do this directly for the original model. This may better illuminate the structure of failures of the original model, and thus point to a deeper way to construct a new one.[35]
- For heuristics, we are looking at the adaptive structure of our cognition, or specific features of our social organization, or specific characteristics of the problem domain, for either strengths or weaknesses, and the conditions under which these are realized. Thus there is (or we can often extract) a reference context that contains more useful information about the method. This then recognizes methodologically the importance of context-dependence.
- Rather than looking for universal theories or principles which are foundational to all other elements of a given domain, look for the conjoint application of robust principles which may be heterogeneous in application, but complement each other to give a broader and richer fit to the details of the situation.
- Look for generative ways in which empirical results, constraints, and conditions may have broad application to extend or support philosophical viewpoints, looking for the kinds of support that come from the above principles rather than entailments or similarly tight linkages. This should include studies of concept and meaning creation, change, and stabilization.

Heuristic methods permeate and constitute the vast majority of methods that we have. It is time that we make a central place for them in our philosophy.

References

Allchin, D. (1998). *Resolving disagreement in science*. book manuscript, in press.

Batterman, R. (1995). Theories between theories: Asymptotic limiting intertheoretic relations. *Synthese, 103*, 171–201.

Batterman, R. (2001). *The devil in the details*. Oxford: Oxford University Press.

Bechtel, W., & Richardson, R. (1993). *Discovering complexity: Decomposition and localization as scientific research strategies*. Princeton: Princeton U. P.

Bechtel, W., & Mundale, J. (1999). Multiple realizability revisited: linking cognitive and neural states. *Philosophy-of-Science, 66*, 175–207.

Beckner, M. (1963). Metaphysical presuppositions in the description of biological systems. In R. Gregg & F. T. C. Harris (Eds.), *Form and strategy in science* (pp. 15–29). Dordrecht: Reidel.

Beurton, P., Falk, R., & Rheinberger, H.-J. (Eds.), (2000). *The concept of the gene in development and evolution, historical and epistemological perspectives*. Cambridge University Press.

Boveri, T. (1902). On multipolar mitosis as a means of analysis of the cell nucleus. English translation of Über mehrpolige Mitosen als Mittel zur Analyse des Zellkerns, *Verhandlungen der physikalish-medizinischen Gesellschaft zu Würzburg, 35*, 67–90. (Reprinted in B. H. Willier, & J. M. Oppenheim, (1974). *Foundations of experimental embryology* (2nd ed.). New York: MacMillan)

Callebaut, W. (1993). *Taking the naturalistic turn, or how real philosophy of science is done*. Chicago: The University of Chicago Press.

---

[35] This is strongly reminiscent both of *in passim* suggestions of Kuhn (1962), but even more as powerfully elaborated by Lakatos in his *Proofs and Refutations* (1978). Lakatos gives a much richer sense of the productive use of heuristics in this work on mathematics than he does in his subsequent (in my opinion, less successful) attempts to generalize it to science.

Carlson, E. O. (1967). *The gene: A critical history*, Philadelphia: W.B. Saunders.

Cartwright, N. (1983). *How the laws of physics lie*. London: Oxford University Press.

Darden, L. (1974). Reasoning in scientific change: the field of genetics at its beginnings. Ph.D. Dissertation, Committee on Conceptual Foundations of Science, The University of Chicago, 1974, p. 114.

Darden, L. (1991). *Theory change in science: strategies from mendelian genetics*. New York: Oxford University Press.

Darden, L., & Maull, N. (1977). Interfield theories. *Philosophy of Science, 43*, 44–64.

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

Dennett, D. (1995). *Darwin's dangerous idea*. New York: Simon and Schuster.

Gilbert, S. F. (2001). Ecological developmental biology: Developmental biology meets the real world. *Developmental Biology, 233*, 1–12.

Glennan, S. (1996). Mechanism and the nature of causation. *Erkenntnis, 44*, 49–71.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*, S342–S253.

Gregory, R. L. (1962). Models and the localization of function in the central nervous system. Reprinted In C. R. Evans & A. D. J. Robertson (Eds.), *Key papers: Cybernetics*. London: Butterworth, 1967.

Hamilton, W. D. (1964). The genetical foundations of social behavior, I and II. *Journal for Theoretical Biology*, 2–59.

Hamilton, W. D., (2000).

Hooker, C. (1981). Towards a general theory of reduction. *Dialogue, 20*, 38–59, 201–236, 496–529.

Kauffman, S. (1971). Articulation of parts explanations in biology and the rational search for them. In R. C. Buck & R. S. Cohen, (Eds.), *PSA-1970 Boston Studies in Philosophy of Science, 8*, 257–272.

Kim, J. (1966). On the psycho-physical identity theory. *American Philosophical Quarterly, 3*, 227–235.

Kohler, R. (1994). *Lords of the fly*. Chicago: University of Chicago Press.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago: The University of Chicago Press.

Lakatos, I. (1978). *Proofs and refutations*. Cambridge: Cambridge University Press.

Lenat, D. B. (1982). The nature of heuristics. *Artificial Intelligence, 19*, 189–249.

Levins, R. (1968). *Evolution in changing environments*. New Jersy: Princeton University Press.

Lloyd, L. (1988). *The structure and confirmation of evolutionary theory*. New York: Westview Press, reprinted 1994, NY: Princeton University Press.

Morgan, T. H. (1916). *A critique of the theory of evolution*. Princeton: Princeton University Press.

Moss, L. (1992) A kernel of truth? on the reality of the genetic program. In D. Hull, M. Forbes, & K. Okruhlik (Eds.), *PSA-1992* (vol. 1, East Lansing, MI: The Philosophy of Science Association, pp. 335–348).

Moss, L. (2002). *What genes can't do*. Cambridge: M.I.T. Press.

Nagel, E., (1961). *The structure of science*. New York: Harcourt, Brace, Jovanovich.

Nozick, R. (1993). *Rationality*. Princeton: Princeton University Press.

Nickles, T. (1973). Two concepts of inter-theoretic reduction. *Journal of Philosophy, 70*, 181–201.

Nickles, T. (1976). Theory generalization, problem reduction, and the unity of science. In A. Michalos & R.S. Cohen (Eds.), *PSA 1974* (Boston Studies in the Philosophy of Science, vol. 32, pp. 31–74). Dordrecht and Boston: Reidel.

Nickles, T. (1981). What is a problem that we may solve it? *Synthese, 47*, 85–118.

Nickles, T. (2003). Normal science: from logic to case-based and model-based reasoning. In T. Nickles (Ed.), *Thomas Kuhn* (pp. 142–177). Cambridge: Cambridge University Press.

Raff, R. (1996). *The shape of life*. Chicago: The University of Chicago Press.

Ramsey, J. (1995). Reduction by construction. *Philosophy of Science, 62*, 1–20.

Richardson, A. (1978). *The practical art of blacksmithing* (originally published in 4 volumes, 1889–1891). New York: Weathervane Press, Crown Books.

Roux. W. (1883). *Über die Bedeutung der Kerntheilungsfiguren*, Leipzig, 1883. Translated and reprinted as On the significance of nuclear division figures: A hypothetical discussion. In B. Voeller. (Ed.) *The chromosome theory of inheritance*. New York: Appleton Century Crofts, 1968.

Sarkar, S. (1998). *Genetics and reductionism*. Chicago: The University of Chicago Press.

Schaffner, K. (1967). Approaches to reduction. *Philosophy of Science, 34*, 137–147.

Schaffner, K. (1974). The peripherality of reductionism in the development of molecular biology. *Journal of the History of Biology, 7*, (1974b), 111–139.

Schaffner, K. (1993). *Discovery and explanation in biology and medicine*. Chicago: University of Chicago Press.

Schank, J. C. (1991). *The integrative role of model building and computer simulation in experimental biology*. Ph.D. Dissertation, Conceptual Foundations of Science, University of Chicago.

Schank, J. C. (2001). Beyond reductionism: Refocusing on the individual with individual-based modeling. *Complexity, 6*, 33–40.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99–118.

Simon H. A. (1962). The architecture of complexity. Reprinted in Simon 1996 as Chapter 7.

Simon H. A. (1966). Scientific discovery and the psychology of problem solving. In R. G. Colodny (Ed.), *Mind and cosmos, pittsburgh studies in the philosophy of science* (vol. 3, pp. 22–40). Pittsburgh: University of Pittsburgh Press.

Simon H. A. (1973). The structure of Ill-structured problems. *Artificial Intelligence, 4*, 181–201.

Simon H. A. (1996). *The sciences of the artificial* (3rd ed.). Cambridge: MIT Press. Earlier editions, 1969, 1981.

Sutton, W. S. (1903). The chromosomes in heredity. *Biological Bulletin of the Marine Biological Laboratory at Woods Hole, 4*, 231–248.

Sperry, R. W. (1976). Mental phenomena as causal determinants in brain function. In G. Globus, I. Savodnik, & G. Maxwell (Eds.), *Consciousness and the brain* (pp. 163–177). New York: Plenum.

Taylor, P. J. (1985). Construction and turnover of multi-species communities: a critique of approaches to ecological complexity. Ph.D. Dissertation, Department of Organismal and Evolutionary Biology, Harvard University.

Wade, M. J. (1978). A critical review of the models of group selection. *Quarterly Review of Biology, 53* (#2), 101–114.

Waismann, F. (1951). Verifiability, In A. G. N. Flew (Ed.), *Logic and language* (first series, pp. 117–144). London: Blackwell.

Waters, C. K. (1994). Genes made molecular. *Philosophy of Science, 61*, 163–185.

Wimsatt, W. C. (1974). Complexity and organization. In K. F. Schaffner & R. S. Cohen (Eds.), *PSA-1972 (Boston Studies in the Philosophy of Science, volume 20)* (pp. 67–86). Dordrecht: Reidel.

Wimsatt, W. C. (1976a). Reductionism, levels of organization and the mind-body problem. In G. Globus, I. Savodnik, & G. Maxwell (Eds.), *Consciousness and the brain* (pp. 199–267). New York: Plenum.

Wimsatt, W. C. (1976b). Reductive explanation: A functional account. In A. C. Michalos, C. A. Hooker, G. Pearce, & R. S. Cohen (Eds.), *PSA-1974 (Boston Studies in the Philosophy of Science, Vol. 30)* (pp. 671–710). Dordrecht: Reidel.

Wimsatt, W. C. (1979). Reduction and reductionism, invited review article In P. D. Asquith & H. Kyburg, Jr., (Eds.), *Current research in philosophy of science* (pp. 352–377). East Lansing, Michigan: The Philosophy of Science Association.

Wimsatt, W. C. (1980). Reductionistic research strategies and their biases in the units of selection controversy. In T. Nickles (Ed.), *Scientific discovery-vol.II: Case Studies* (pp. 213–259). Dordrecht: Reidel.

Wimsatt, W. C. (1981). Robustness. reliability, and overdetermination. In M. Brewer & B. Collins (Eds.), *Scientific inquiry and the social sciences* (pp. 124–163). San Francisco: Jossey-Bass.

Wimsatt, W. C. (1985a). Heuristics and the study of human behavior. In D. W. Fiske & R. Shweder (Eds.), *Metatheory in social science: pluralisms and subjectivities* (pp. 293–314). Chicago: University of Chicago Press.

Wimsatt, W. C. (1985b). Forms of aggregativity. In A. Donagan. N. Perovich, & M. Wedin (Eds.), *Human nature and natural knowledge* (pp. 259–293). Dordrecht: Reidel.

Wimsatt, W. C. (1987). False models as means to truer theories. In M. Nitecki & A. Hoffman (Eds.), *Neutral models in biology* (pp. 23–55). London: Oxford University Press.

Wimsatt, W. C. (1992). Golden generalities and co-opted anomalies: Haldane vs. Muller and the drosophila group on the theory and practice of linkage mapping. In S. Sarkar (Ed.), *The founders of evolutionary genetics* (pp. 107–166). Dordrecht: Martinus-Nijhoff.

Wimsatt, W. C. (1994). The ontology of complex systems: levels, perspectives and causal thickets. In R. Ware & M. Matthen (Eds.), *Canadian Journal of Philosophy*, Suppl. Vol. #20, pp. 207–274.

Wimsatt, W. C. (1997). Aggregativity: reductive heuristics for finding emergence. In L. Darden (Ed.), PSA-1996, v. 2 [*Philosophy of Science*, Supp. Vol. #2, 1997], pp. S372–S384.

Wimsatt, W. C. (2000). Emergence as non-aggregativity and the biases of reductionism(s). *Foundations of Science, 5*, 269–297.

Wimsatt, W. C. (2002). Functional Organization, Functional Inference, and Functional Analogy. In R. Cummins, A. Ariew, & M. Perlman (Eds.), Substantially revised and expanded version of 1997a for a collection on Function. Oxford. pp. 174–221.

Wimsatt, W. C. (2007). *Piecewise approximations to reality: Engineering a philosophy of science for limited beings*. Cambridge: Harvard U. P. Contains 1974, 1976, 1981, 1985, 1987, 1994, 1996.

Wittgenstein, L. (1962). *Remarks on the foundations of mathematics*. London: Blackwell.