Re-Engineering Philosophy for Limited Beings

PIECEWISE APPROXIMATIONS TO REALITY

WILLIAM C. WIMSATT

HARVARD UNIVERSITY PRESS
Cambridge, Massachusetts, and London, England 2007

Contents

CHAPTER ONE

Myths of LaPlacean Omniscience

Why is it that academics who claim to seek the truth want to pretend that they have always had it? What are they paid for, anyway?

This behavioral paradox is found throughout academia, but is nowhere stronger than in the sciences. We preach with pride that science is fallible and testable—contrasting it with religion and other dogmas. And we should. But we hide with embarrassment papers that are undercut or shown to be erroneous by later work. (Try to get tenure with such a paper!) We commonly don't report negative results. Errors are okay only if they are someone else's or belong to prior generations. Less reflective scientists teach undergraduates only those parts of their science that are "established fact," deferring until graduate school anything disputed or under active investigation. These same instructors then complain that beginning graduate students don't seem to notice when something is wrong or missing, or how to find questions worth researching. But has anyone tried to teach this skill? We publicly celebrate fortuitous discovery, as if to convince humanists and artists that science too is a creative liberal art, but discoveries to be celebrated are those that required a prepared mind and laboratory. We ought to celebrate the well-crafted research plan even if it failed, and not when dumb luck let us succeed without preparation. We should thereby *reinforce what our students can do* to help them to succeed.

Many of the contradictions emerge from our idealizations of science and philosophical views that see knowledge only in certainty. Even when these viewpoints have been given up explicitly, their residue remains. I address these issues in overview in the first three short chapters. In the rest of the book I elaborate an alternative picture.

The focus on truth is understandable. The sciences are the best-crafted repositories of truth on this planet. Their results are important for our technology, and for the public validation of our craft. At least as important are the cognitive technologies, material procedures, and social structures we have crafted for seeking and checking our results. But the way we talk about science and teach it not only fails to communicate this fact, but actively misrepresents it. An alien power seeking to undercut our civilization could do no better than to teach most of our science and its history and philosophy as we do now. This could have been written in 1962 by Thomas Kuhn, or Stephen Brush in 1974. What is striking is that it needs saying now more than ever.

These chapters are primarily about method in science; particularly, the strengths, weaknesses, and proper interpretations of reductionist approaches. This is not as limiting as it might seem. Reductionism is the dominant methodology of "big science" today, percolating widely through the sciences, with methodological strategies applying still more broadly to areas where analytic methods are found. But our analytic training leaves us with a set of bad reactions—that we don't really understand something until we have broken it down into the smallest possible pieces. This broader methodology of reductionism has important implications, but not those that many philosophers have assumed. Philosophers once put these methods at the core of a formalistic foundationalist philosophy of science. But this view is fundamentally flawed by idealizations we make about how we reason. As real beings, we "deviate" from these idealizations about computation, cognition, inference, decision making, and the structure of our conceptual schemes.

One important mistake is the belief that reductionist or analytic methods eliminate or analyze away what is being analyzed or reduced. This produces a false opposition between those working at different levels of organization, or between those using "humanistic" and "scientific" approaches to the phenomena. The aim of what are called reductionist explanations in science is not to atomize phenomena to the lowest possible level, but to be able to articulate and understand entities, events, and processes at diverse levels, and to give explanations involving heterogeneous relations among them in producing complex phenomena. We can get a robust and lasting appreciation of processes at higher levels of organization in their own terms that is not compromised by having lower-level accounts. And large parts of our reasoning are profoundly heuristic at our working levels, as we craft and modify our procedures. It is here that we must start.

Many of these paths were first sketched four decades ago by Herbert Simon. I have explored them often, and philosophers now traverse them in increasing numbers.[1] We also need a context of real problems, keeping us honest in ways that idealizations and toy examples common in philosophy often fail to do. In practice we use surprisingly powerful heuristic tools rather than such idealizations to structure our search. These new tools are sources of inspiration for an alternative, robust, naturalistic, and scientifically motivated realist philosophy of science. With this focus, scientific processes are seen as end-directed or teleological activities, generating a cognitive engineering of science, with natural resonance with design analyses in engineering (Chapter 12)—a fruitful path so far ignored in our profession.

Realism for Limited Beings in a Rich, Messy World

I seek methodological tools appropriate to well-adapted but limited and error-prone beings. We need a philosophy of science that can be pursued by *real* people in *real* situations in *real* time with the kinds of tools that we *actually* have—now or in a realistically possible future. This must be a central requirement for any naturalistic account. Thus I oppose not only various eliminativisms, but also overly idealized intentional or rationalistic accounts. In these chapters I advocate an approach that can provide both better descriptions of our activities and normative guidance based on realistic measures of our strengths and limitations. No current philosophy of science does this fully, though increasing numbers are moving in that direction. A philosophy of science for real people means real scientists, real engineers, historians or sociologists of real science and engineering, and real philosophers interested in how any of the preceding people work, think about their practice, think about the natural worlds we all inhabit, and think about what follows reflectively and reflexively from these facts.

This view involves a species of realism, though not of the usual sort. It fits current scientific practice and illuminates historical cases better than other approaches, and it has implications for how to do philosophy.[2] It neither has nor seeks the stark simplicities of current foundationalist theories. This philosophy must be based from top to bottom on heuristic principles of reasoning and practice, but it also seeks a full accounting of other things in our richly populated universe—including the formal approaches we have sought in the past.

This project is a philosophy for messy systems, for the real world, for the "in-between" (see the frontispiece), and for the variegated ecologies of reality supporting and increasingly bent to our science and technology. *Pace Quine,* this is ontology for the tropical rainforest. The "piecewise approximations" of the book title is unavoidable: we are, must be, and can be realists in our science and much of our practice. But our realism, like our practice, and even our inferential consistency, must be piecemeal and usually satisfied with a local rather than a global order. We aren't God and we don't have a God's-eye view of the world. (In this piecemeal world, we don't even have a *gods' eyes* view.)

But then the first part of the title is only half-truth: to re-engineer the whole of philosophy in a human image is still ambitiously global. I don't do all this. I sketch how to do it for significant parts of philosophy of science and closely connected areas of science. This captures new phenomena and reconceptualizes old in ways that fit more naturally with how we proceed. I hope that others find these results sufficiently suggestive to use, extend, and add to the tools I describe here to employ them elsewhere.

"Re-Engineering" appears in the title as a verb: this view of science and nature is constructed largely (as with all creative acts) by taking, modifying, and reassessing what is at hand, and employing it in new contexts, thus *re*-engineering. *Re*-engineering is cumulative and is what makes our cumulative cultures possible. And any engineering project must be responsive to real world constraints, thus realism. Our social, cognitive, and cultural ways of being are no less real than the rest of the natural world, and all together leave their marks. But putting our feet firmly in the natural world is not enough. Natural scientists have long privileged the "more fundamental" ends of their scientific hierarchy, and pure science over applied—supposing that (in principle) all knowledge flowed from their end of the investigative enterprise.

Not so: *Re-engineering* also works as an adjective, and has a deeper methodological role. Theorists and methodologists of the pure sciences have much to learn about their own disciplines from engineering and the study of practice, and from evolutionary biology, the most fundamental of all (re-)engineering disciplines. Our cognitive capabilities and institutions are no less engineered and re-engineered than our biology and technology, both collections of layered kluges and exaptations. We must know what can be learned from this fact about ourselves to better pursue science of any sort.

Social Natures

Philosophers were once deaf to claims that cognitive science, evolutionary biology, and issues of practice were centrally important to *our own* methodology. In those days, psychology led only to "psychologism," the study of history only encouraged committing the "genetic fallacy," and "sociology" seemed worse than socialism. Knowledge after all was mind-independent and timeless, and minds were individual and both abiological and asocial. Today such opinions seem quaint; such philosophers find themselves increasingly embattled, circling the wagons in defense of Reason. We show our own disciplinary biases and force them on others: the various "philosophies of X" often seem to be more about arguments internal to philosophy than "of" anything. In much of our broader humanistic intellectual community, the social rootedness of our activities is taken for granted, the real world is seen as a construction, and reason is but a chattel of interests. But this course also swings too far. How can we bridge these divergent perspectives?

I want to re-psychologize, re-socialize, and re-embed us in the world, where we reason about that world as well as about how we interact with and reflect upon it. Can we still be recognizably philosophical while letting the subjects of "philosophies of" shine through much more clearly and inspire new philosophies, rather than merely exporting our same old "philosophical" disputes to these new territories?

Contrary to some philosophical views, "empirical contingencies" are crucially important to philosophy. We are embodied socialized beings: evolved and developing in a world conditioned by our sociality and technology. We steer, often unreflectively, through it with values that are uneasy combinations of history, religion, science, pseudo-science, and the latest fads—most recently the ideological preachments and short planning horizons of a "free" market. Any adequate account of reason must see it as the adaptation that it is: fallible, but self-correcting. And self-correcting not just through reason alone, but in the way that DNA is self-replicating—when embedded in a larger supporting complex that is both of the world and self-continuing in the world.

Those who want to recognize our social natures rarely see biologists as useful allies. But Levins and Lewontin (1985) argue that organisms often actively construct their environments, and urge a rich developmental and evolutionary *relationalism*, treating organism-environment interactions as real (and often primary) without succumbing to relativism.[3] And Paul Ehrlich (2000) speaks of "human natures" because culture is part of human nature, and different cultures make different natures. He thus recognizes that the Hobbesian state of nature never existed. We were social throughout our primate history—long before language or tool use. The social contract is a creation myth of western political science. Our culture is midwife to and product of our extended juvenile period of learning, before we mature and our habits become less plastic. Yet this richly relational world is still a natural world, and we must study both biology and cultures to understand it.

But how do we study nature? How do we organize it? How do we avoid chaos, disorder, or unfathomable confusions in our account of natural complexities?

Heuristics as Adaptations for the Real World

Philosophers have established ways of creating order working from a few key idealizations about decision making and rational or logical inference. These principles crop up continually; normative frameworks for almost everything else they do, and (supposedly) everything else *we* do. These assumptions are deeply *generatively entrenched,* widely used pivotal assumptions that play a role in generating our philosophical theories of almost everything else.[4]

So is this a problem? Unfortunately, yes: these particular idealizations demand unrealistic degrees of knowledge, unlimited inferential powers, or both (see Appendix D on LaPlacean demons). They don't fit our behavior. As normative rather than descriptive principles, that's not necessarily a problem. But they make assumptions that we *cannot* satisfy, and that *is* a problem. How can these provide goals toward which we should orient ourselves? How could they be correct even as normative claims? Should we give them up? We hear that to do so would be to open the doors to irrationalism, relativism, or other unnamed horrors precipitating total cognitive collapse and chaos. *But the very success of our cognitive and social adaptations in the real world belies those fears. It is an existence proof that there are better idealizations or models to be made.*[5]

Where should we look instead? I study the heuristic techniques scientists use to explore, describe, model, and analyze complex systems—and to "de-bug" their results. These are plausibly more carefully tuned and evaluated descendants of more broadly used inferential practices. So we start with our actual practice—but seeing these practices for their strengths as evolved cognitive adaptations rather than as compromised attempts to pursue our ideals. *These "deviations" are not failings, but the source of our peculiar strengths in this uncertain world of complex, evolving beings, technologies, and institutions.*

Considering our cognitive powers and limits and the social contexts and embodiments of our decision processes yields many benefits.[6] As philosophers we can analyze new scientific activities left unexplored or dismissed because our overly idealized models have lacked the resources to analyze them. Theories of science and other human activities can be both broader in scope *and* more sensitive to detail while giving better—more effective, detailed, and realistic—normative guidance for how to proceed. As scientists, more realistic accounts of these activities help us to sidestep avoidable errors, find and fix them when they do occur, and pursue our goals more effectively. Thus my intended audience is philosophers—to get them to change their practice—and scientists—in the hope that these ideas might be genuinely helpful in their practice. There *is* room for a theory of practice: a genuinely philosophical theory with lessons in how to do philosophy as well as for philosophy of science.[7] I present case studies in and of this practice. I will return to some of the ways we could expect changes in Chapter 3.

Nature as Backwoods Mechanic and Used-Parts Dealer

The worldview of logical empiricism was built with minimalist tastes, axiomatizations, on the inspiring constructive triumphs of the new symbolic logic, and on the more rigorous standards of proof that accompanied them. The computational worldview descending from it builds on algorithms—powered and inspired by the marvelous and increasingly powerful engines we have built to execute them quickly and reliably. The algorithmic view adds to the axiomatic an appreciation of problems of computational complexity, and procedures and subroutines make explicit the advantages of a natural modular and hierarchical organization. But the algorithmic view, an improvement, supposes or tries to construct a well-defined structure to the problem-space (Simon, 1973), and an abstract crispness in the engines for exploring it.

These views are each appropriate to many tasks, though neither is as general as commonly supposed. There is order in our practice, raw material for a generalizable account. These *in principle* claims deflect our attention, and don't help here. So I will systematically ignore the kinds of *in principle* claims often arising in these two contexts and much beloved by philosophers to see how we do—and should—deal with situations we face *in practice*.[8]

Our ancestors didn't adapt in this complex world with simpler but still formally respectable deduction systems. Biology is full of diverse kinds of inductive pattern-detectors. The first problem was finding order, not testing it, and finding it according to rough error-prone procedures. If we look at ourselves as biological social cognitive beings, we see that our responses to problems of adaptation in a complex world are crafted with heuristics. Insofar as these are products of selection processes, biological and cultural dimensions of our reason must also be heuristic. *Heuristic principles are most fundamentally neither axioms nor algorithms, though they are often treated as such. As a group, they have distinct and interesting properties.*[9] *Most importantly, they are re-tuned, re-modulated, re-contextualized, and often newly re-connected piecemeal rearrangements of existing adaptations or exaptations, and they encourage us to do likewise with whatever we construct.* We are just now beginning to come to grips with that picture: Nature as a reconditioned parts dealer and crafty backwoods mechanic, constantly fixing and redesigning old machines and fashioning new ones out of whatever comes easily to hand.[10] So much of engineering—even of "new" designs—is *Re*-engineering. This has profound implications for the character of evolutionary products (and of evolutionary change; see Chapter 7).

Are heuristics merely an interesting subset of the class of all possible algorithms, as Dennett (1995) suggests? If we imagine a list of all executable procedures, all truth-preserving procedures as well as all heuristics would be on it. This broader class is sometimes called the class of algorithms. This picture has a misleading air of necessary truth. It encourages us to treat heuristics as derivative from algorithms—a mistake. (We might as well treat texts as derivative from alphabetic sequences.) The truth is quite the reverse. Heuristics are both causally and phylogenetically prior to our algorithmic abstractions. Evolution opportunistically favors results, using any available method, no matter how crude, that is cost effective in its context relative to other available alternatives. Recognizing the heuristic character of our reasoning solves many otherwise intractable problems (Aloimonos and Rosenfeld, 1991; McClamrock, 1995).

The context-free elegance we are taught to value is of no broader foundational moment than the contexts these methods were designed for. Traditional foundationalisms combined the need for systematization and the reliability of socially cross-checkable proof processes and mistook them for an account of the internal character of reasoning (Hutchins, 1995; Gigerenzer, 1993). But formalistic foundationalism fails miserably as a process model for cognition, for the construction of theories or devices, or for any biological adaptive processes. It has done little better for philosophy. Working practices with axioms and algorithms can be seen as limiting special kinds of heuristic methodologies. Well adapted to appropriate problems (the deductive exploration of formal models), the success of formal methods in those instances has no generalizable foundational significance. A naturalistic worldview of the genesis and operation of functionally organized evolving systems is heuristics all the way down—as far down as entities or configurations of them are products of selection. (Biological adaptations share the central features of heuristic procedures—see chapters 4 and 5; appendixes A and B; and Wimsatt, 1980b).

Error and Change

So should we just substitute heuristics for axioms as new fundamental rules, acting on whatever comes easily to hand, and proceed as before with a neo-classical foundationalism? Not quite! We also need adjustments elsewhere: a change in the scope of (and expectations from) acceptable inference rules; a new conception of foundations, which must become dynamical and potentially changeable; a more error-tolerant conception of the organization of our conceptual structures; and a new treatment of error. Our adaptive mechanisms must be capable of detecting and responding to—nay, feeding on—errors at different levels and across varied contexts, and exploiting parallelisms and redundancies eschewed by formalists to detect errors and fine-tune responses. One of the most common and effective uses of deductive arguments in science is to detect and localize errors and refine concepts.[11]

With a new focus on error and change, a richer theory must address not only how generative structures work, but how they are changed or modified in midstream, and what patterns such changes should follow—*for there are such patterns, important ones, and they make such an endeavor worthwhile.* These are totally ignored in traditional approaches, but are central to a fuller reading of the nature of developmental and evolutionary processes—whether biological, cognitive, or cultural, including processes of scientific change. *Generative entrenchment* provides a heuristic and dynamical reading of an attractive feature of foundational theories: the remarkable generative power of a few well-chosen assumptions, structures, or processes, and the consequences of this power for freezing-in the assumptions that engender it. This crucial idea and its implications are sketched in Chapter 7, suggesting the form of a foundationalism for non-foundationalists—scientists, engineers, and students of adaptive complexity across the disciplines.

The roots of the views offered here are diverse. The rich backwoods of evolution (Darwin's "tangled bank") is a heterogeneous, multi-level tropical rainforest, with converging overlapping branches, and patterns of intersecting order, residents, and connections at a variety of levels, but no *single* stable foundational bedrock that anchors everything else. Yet this multiple rootedness need not lead to "anything goes" perspectival relativism, or an anti-naturalist worship of common sense, experience, or language. It yields a kind of multi-perspectival realism anchored in the heterogeneity of "piecewise" complementary approaches common in biology and the study of complex systems (chapters 9 and 10). Here an overlapping diversity of roots, assumptions, approaches, and methods is fruitful (Chapter 4). But tracing foundations, sensible for more simply rooted systems, becomes both more difficult and less important. One could argue that multiple rootedness only poses *practical* problems for tracing foundations; and, *traditionally,* as philosophers, we are only interested in matters of principle! This is a serious mistake. Several chapters consider arguments from principle, and give reasons for rejecting, deflecting, or reinterpreting them in specific contexts (especially chapters 5, 9, 11, and 13). Like false models and qualified generalizations, *in principle* arguments are often honored more in the breach than in the observance. But we need to consider the role of *in principle* arguments more generally—especially where these spring from idealizations concerning our rationality. We start this in Chapter 2.

Organization and Aims of This Book

These essays, written between 1971 and 2002, are about a fifth of my published work. This chapter and the next two, chapters 7 and 13, and the introductory materials to parts II–IV are published here for the first time. Chapter 12 has been substantially expanded from earlier versions with new scientific examples.

Chapters 2 and 3 elaborate in more detail the fallibilist, realist, and heuristic perspective I urge. After the introductory section, the following chapters form three groups.

First (in Part II) is a series of general methodological essays on some deep and pivotally important heuristic principles of inference, principles that are used over and over again in the chapters that follow. They are, in effect, both heuristic and foundational (if "heuristic" is unfamiliar or problematic, Chapter 5, on heuristics, might be the best way to start). These chapters provide perspectives and principles that should be of much broader use in philosophy. They have relevance to the analysis and construction of diverse complex evolved structures—from macromolecules, to musical composition, to social networks and methodologies.

Next (in Part III) is a collection of chapters on the nature of reduction and emergence—a specific set of methodological problems central to modern science and particularly relevant to the study of complex systems, which is a constitutive theme of my work generally. These topics are strongly informed by the methodologies advocated in the first section, and provide further examples of their application. (Indeed, grappling with these issues reinforced for me their centrality.) Reductionist methodologies commonly take for granted the perspectives or levels of organization they utilize and focus on how they are related. These chapters—particularly 9, 10, and 12—also deal with how we recognize or individuate perspectives, levels, and decompositions in the first place.

The introduction to Part IV and Chapter 13 look backward to theoretical physics and engineering simulation in a pre-computer age. Incidents there led me away from my naive reductionism to this heuristic-based realist view of science and engineering. It deals with relations between pure and applied science, and what philosophers have to learn from a closer look at the engineering perspective.

Substantial introductions to the different sections track the most important themes in the chapters, and attempt to create a new order among them. For some major elaborations of their arguments, you might wish to return to the introductions after reading the chapters. This chapter is itself the introduction to the next one, which expands on two themes: the nature and differences between our uses of idealizations in science and in philosophy, and the fundamental reorientation toward error required in this fallibilist and heuristic view of adaptation and knowledge. In Chapter 3 I advance proposals for how to view the philosophical fallout. Further meta-reflections are found particularly in Chapter 10 and in the final sections of chapters 12 and 13.

The first two appendixes concern heuristics, elaborating Chapter 5, and complementing the rest of the book. Appendix A lists important properties of heuristics crucial to explaining characteristics of their use. Appendix B lists 20 different heuristics used in conceptualizing, model-building, experimental design, and other activities in the analysis of systems and their behavior. These heuristics are reductionistic or have that effect under specified circumstances. They may be helpful in analyzing a system, but carry with them characteristic biases associated with ignoring or downplaying context. They flesh out specific portable methods in how a reductionist might approach a system or problem—a "toolbox" for methodological reductionists. Appendix C is a limited glossary for key concepts or assumptions used in the text. Appendix D reviews unrealistic "in-principle" idealizations found in philosophical or scientific arguments, assuming powers that are infinitely (or sometimes just indefinitely) greater then we could possibly achieve. Their obvious falsity points to the need to re-engineer philosophy for limited beings.

CHAPTER TWO

Normative Idealizations versus the Metabolism of Error

Inadequacies of Our Normative Idealizations

Scientists use idealizations, often conflicting ones, for various ends. Planets and stars are point masses in deriving Kepler's laws, as are molecules for the ideal gas law, while some mid-sized bodies (springs) are extended but massless in the simplest treatment of the harmonic oscillator. Geophysicists, stereochemists, and race-car designers, respectively, use other idealizations for these same objects sharing none of these assumptions. Organisms may be ageless, identical, and sexless in some ecological models, sexed and variably aged in others; discretely variable in specified ways in genetic models; and continuously variable in a small set of specified directions in evolutionary optimization models. Different sets of false assumptions—idealizations—play crucial roles in teasing apart different aspects of the causal structure of our world and permeate model-building, which is the cutting edge of theory construction. They are implicit in and motivate experimental design. They are used in acquiring, describing, and analyzing data. Models or idealizations increasingly replace theories as foci for studying interactions among these three central scientific activities.

Idealizations commonly play different roles for scientists and philosophers. Scientists treat models as simplifications of or useful counterfactual transformations of nature—visions ultimately to be corrected and elaborated in an ongoing theoretical, experimental, and observational dialectic with our world. This dialectic is socially, culturally, and technologically arranged, like other activities in our "carpentered" world. But it is clear that when the model doesn't fit reality, as it commonly doesn't, we should make changes in the model to do better. We don't expect reality to change to fit the model!

Philosophers sometimes seem to want the reverse: for human agents, we assume that they should change to fit the models. How can this be? Philosophers use idealizations too—to simplify and generalize analyses, to abstract away from particular details peripheral to the point at hand, and often to urge certain norms of behavior. This *normative* role gives idealized models of rationality and inference an ambiguous status. As implicit accounts of the scientist as rational agent—as logical thinker and utility maximizer—these models often play as presuppositions in the background while the analysis targets something else: decision under uncertainty, the idea of a conceptual scheme, theory reduction, supervenience, causal explanation, or the nature of language. These idealizations too are models, though we rarely act as if they were.[1] Calling them "constitutive ideals" or suggesting a special normative or generative role doesn't hide the fact that they embody assumptions and conditions. It is sensible to ask whether and how well these assumptions are realized in that part of the world they are applied to. While they are sometimes inspected directly, more commonly these assumptions are invisible. They aren't treated as if they were part of the analysis. Problems with them are often misidentified or incorrectly localized as due to some failure in the proposed analysis.

Why this strange behavior? I assume that models are kinds of abstract structures or sets of assumptions that can have either a broadly descriptive use, as in science, or to specify a standard for behavior, in which case it is used normatively. (Scientific models are actually used in various ways—see Chapter 6—but here they will all be called "descriptive"). Unlike these uses of models, *a model treated as normative is not usually compromised by our failure to act accordingly*. We can *choose* to follow a normative rule or not—the usual contrast between rules that guide action and causal laws. So violations of a rule may indicate our own failures rather than problems with the rule. But this is not enough: falsely describing human behavior is too weak a test for normative status! Many norms are commonly followed, often better than our best predictions for systems of comparable complexity. Consider Dennett's (1971) rationale for the "intentional stance," to assume that a potential opponent is a rational agent with reasonable beliefs is a powerful predictor of their behavior. Blaming the agent *is* the most likely response to deviant behavior in many contexts. For a well-designed set of norms, this is appropriate. But blaming the agent just places the initial burden of proof; else we could never question or reject a normative model. And sometimes we must. The right rules? Sometimes these are not even *good* rules for us to follow.[2]

Philosophers treat rationality as a normative ideal deeply rooted in our conception of ourselves and all that we do. But in our complex world of limited and fallible agents, with imperfect, noisy, and incomplete information, is it rational to try to be *perfectly rational?* Herbert Simon (1955) first shocked decision theorists four decades ago with the claim that it wasn't. He urged "satisficing" (discussed below) rather than "maximizing" as a basis for rational choice, according to his "Principle of Bounded Rationality." Darwin would have liked his views.[3] Philosophers and economists have resisted this move, favoring extreme idealizations and often deifying them as necessary or conceptual truths. We are still far from realizing the full impact of Simon's revolution. *The normative status of common philosophical idealizations of the scientist-as-agent interacts with various common cognitive biases to serve us poorly. This status hides how badly these idealizations perform as descriptions of our actual behavior and also give misleading advice about what we should do about the deviations we do detect.*

Consider an example of the cobbled together but adaptive way heuristics can act to our benefit. Festinger (1957) found that the perceived utility of a chosen alternative increased after making the choice, apparently just through the act of choosing. *This would give a bias towards answering "yes" to "Did you maximize utility in your choice?" if the question were asked after the fact, even if he or she had not maximized utility.*[4] This behavior seems more delusional than rational, but this "irrational bias" should normally improve the performance of individuals who have it over those who don't.

It may be advantageous to change utility assessments after the fact if there is some cost associated with changing your mind after making or announcing your decision. There often is. A subconscious biasing process favoring a chosen alternative could act as a lock-in device—useful if preferences are psychological variables fluctuating in time.[5] We notice decisions if the choice is important and problematic, so the alternatives being weighed are generally close. In these cases it may cost more to change after committing than to go with a suboptimal choice. Festinger's example involved vacillation between two potential mates—something likely to scare both away! Could this cunning unreason be meta-rational? Such a "utility distorting" device in our cognitive apparatus may serve us better, even if we are "misled" in our assessments.[6] Indeed, *is* this being misled? How (and when) *should* we evaluate the "goodness" of choices? Traditional decision theory just assumes that utilities are fixed. Perhaps calling it a "distortion" is an artifact of a static view of choice. Other such cognitive biases and their effects are discussed in chapters 5, 4, 7, 11, 13, and Wimsatt (1980b).

The status of idealizations is clearer for scientists. When we describe planetary orbits as elliptical we ignore deviations caused by the planets' gravitational attraction for one another, and a host of smaller effects required for more accurate prediction and explanation (Peterson, 1993). Human population growth is logistic (a simple model of exponential growth with resource limitations) if we ignore the age-structure of the population, and treat individuals of different ages as if they were alike. But they aren't: limited resources and other biological and cultural variables have effects that change with age. Women of 2, 32, and 62 years have different birth and death rates responding differently to changing resources. Some changes are intrinsic to the natural life cycle and others are modulated by culture. We accept these inaccuracies when they serve our purposes. For more detail and accuracy, we modify models, predictions, and collect more data on age-structure and other things as required. Modeling is discussed further in chapters 6 and 7.[7]

When we describe decision makers as maximizing expected utility, is it equally clear that, except rarely, they cannot possibly do so, that this is an incomputable task? Such models are not even in the right ballpark for the structure of most of our decisions. Our world is too complex, and our abilities too limited for such a model: we have neither the knowledge nor the computational facility required even to formulate correctly all the terms in the equation for expected utility, much less to determine the parameter values. Nonetheless we are given the common philosophers' de facto idealization that we are LaPlacean demons—omniscient and computationally omnipotent computers.[8]

Consider other common idealizations: virtually no real scientific theories are globally free from contradiction, and few real scientific inferences are truly deductive or truth preserving. Approximations in the theory and in its derivations belie both of these claims, and real scientific theories are laced together with approximations.[9] Sloppy idealizations about birth and death rates don't raise a hair, but what could be more important than birth and death? In science, we at least try to identify where these approximations will get us into trouble; yet contradictions in theories and approximative inferences are skeletons in almost all philosophical closets. *Philosophers internalize their denial and have built such magnificent normative edifices upon them that they don't seriously consider the possibility that they are false—not just in little ways, but in ways fundamentally compromising their approach to the rationality of human behavior and the practice of science.* I'm not against the rationality of science, I'm all for it. But this just is a bad conception of it.

Satisficing, Heuristics, and Possible Behavior for Real Agents

What other options are there? Consider Herbert Simon's (1955, 1996) "satisficing" account. An agent has a level of aspiration or of satisfaction and evaluates alternatives by whether they fall above or below this level. The level of aspiration is set and modified through experience. Alternatives aren't given in advance, but generated, or acquired (and sometimes lost) sequentially in real time. Decisions are made according to heuristic rules—rules as simple as "choose the first alternative which meets or exceeds your level of aspiration, unless you already know one to be probably better." This model demands much less in computational and cognitive resources than rational decision theory, and fits plausibly with decision mechanisms we actually use. Unlike the maximization model, one need not compare—or even generate—all choices (Goldstein and Gigerenzer, 1996). Satisficing also integrates learning: aspiration levels are constantly modified by experience. There's no need to determine an exhaustive and mutually exclusive set of choices, or the comparative judgments required within it, to determine the chosen alternative. Giere (1988) applies satisficing to the decisions of scientists, arguing its advantages over more traditional Bayesian models. *The heuristic principles of reasoning ("heuristics") urged here are natural extensions of satisficing, or of Simon's broader conception of "bounded rationality."*

But with satisficing, wouldn't you accept lower standards and performance? You need not. And higher standards aren't always better. "Reasonable goal setting" commonly contributes to better performance (McClelland, 1973). "Who knows just *how* well we can do?" Indeed. But are these realistic standards? Can we try to high jump 20 feet—or a quarter of a mile? What improvements in diet, training regime, or genetic engineering are supposed to result in that? Another million years of hominid evolution? Not a prayer! Yet our implicit models of decision making are orders of magnitude more demanding of our grey matter, and less plausible. We seem not to want to recognize limitations on our powers of thought. Standards like that we don't need.

Nor is this a benign failure. Idealizations about reasoning enter surreptitiously into our technology as well, and cause real problems. Those who worry about "human engineering" have known and studied this concern since the Second World War. Cognitive psychologist Donald Norman puts it this way:

> Today most industrial accidents are blamed on human error: 75 percent of all airline accidents are attributed to "pilot error." When we see figures like these it is a clear sign that something is wrong, but not with the human. What is wrong is the design of the technology that requires humans to behave in machine-centered ways, ways for which people are not well-suited . . .
>
> When technology is not designed from a human-centered point of view, it doesn't reduce the incidence of human error, nor minimize the impact of the errors that do occur. Yes, people do indeed err. Therefore the technology should be designed to take this well-known fact into account. Instead, the tendency is to blame the person who errs, even though the fault might lie with the technology. (Norman, 1993, p. 11)

What kinds of machines are *we* supposed to be to run this technology? Machines who, in an airliner or nuclear power plant for example, keep track of hundreds of meter readings, detect any discrepancies, and decide correctly what to do about it. And even if and when we detect anomalies, idealized scenarios of "deciding what to do about it" fail to notice that we may treat aberrant readings as problems with the instruments rather than with the monitored processes. Instrument failures are far more common, and we may resist recognizing more baleful possibilities. Idealized Cartesian assumptions are again to blame: we will treat instrument readings as infallible (just when we should, of course!) *rather than adding dangerous delays as we try to check up on them or grapple with the implications of what they are telling us.* These are both *prima facie* reasonable reasons for delay if immediate action has a significant cost for which we may be blamed if it wasn't required.

These are common patterns of failure in after-the-fact analyses of major disasters. There is seldom a single error, but a series of compounding factors: intersecting errors of design, manufacture, maintenance, training, knowledge, and use. Peterson (1995) has revealing case studies of diverse software errors—errors that often led too much larger accidents. Often the disaster develops over time and could have been better contained if operators had made the "right" decisions;

judged by hindsight, of course. Medvedev (1991) is eloquent on Chernobyl, as are Petroski's many essays celebrating the role of failure in engineering design. Idealizations of our cognitive powers ignore not only our humanity, but also our biology—both our cognitive limits and our strategies for dealing with error. This is not just a biology of constraint, invoking biology to explain and apologize for what we cannot do. Evolution empowers us with reliable and error-tolerant solutions, and we can learn by looking more closely at nature's designs.

So what should we learn for normative idealizations? Kant's "[O]ught implies can" should be a constraint on all of our obligations. Given our unrealistic idealizations, this is perhaps the deepest principle of a heuristic approach: a test that idealizations must pass before we can accept them as providing normative guidance. We aren't talking here about apocalyptic descriptivism, fine-tuning our normative idealizations to adapt to all of the pebbles on the beach. What if no one can *ever* provide what they ask? By Kant's maxim, such idealizations are not good norms for us even to *try* to follow. *If we try to follow methods that require things far beyond our capacities, we may miss more effective tools appropriately tuned to our true abilities.* We have two such sets of tools—one sharpened and directed in hundreds of millions of years of animal, vertebrate, mammalian, primate, and hominid evolution. The other, a social and cultural evolution that has given us a second overlapping and largely complementary set. A broader formulation of the problems posed by Festinger's "utility distorter" may reveal better solutions than narrower "rationalistic" formulations that ignore or oversimplify more of our context and biological and psychological nature. Why not study these biological, psychological, social, and cultural tools for inference and problem solving, and include at least the more robust and less "accidental" (or avoidable) ones in our conception of the scientist as agent? We use more than logic and rational decision theory to get around in this world—even when it appears that in principle, we could get by with them alone. Epistemologists and ontologists like to get by with the minimum required, with clear and simply stated boundaries and no hard cases. But be wary of such promises: any claim that something can be done *in principle* is a tacit admission that it hasn't yet been done in practice!

The Productive Use of Error-Prone Procedures

Our common modes of inference aren't perfect. As with all inductive methods of the sciences, heuristic principles don't guarantee results.

And, through error or accident, we can *mis*use a truth-preserving rule of inference, thereby "voiding the warranty." So one could be a skeptic about any of our inferences. Exactly this line was used to argue that the senses don't give us infallible knowledge of the external world, and served various forms of idealism from the eighteenth to the early twentieth centuries. Philosophers often seem never to get past this point; they get hung up on certainty. Why certainty? Since Descartes, it has seemed to be the best way to avoid errors. We have engaged for over 350 years in futile searches for guaranteed ways. But what if a search for reliable knowledge is *not* best pursued as a search for guarantees?

Maybe errors are okay—or tolerable—as long as they aren't too frequent, and if we have good ways of remembering, detecting, and recovering from common ones; learning from the patterns of our mistakes; and effectively maintaining, refining, and teaching what we have learned. *More remarkable than our occasional failures is the fact that these common methods work so well as often as they do.* But with the above capacities, this isn't surprising: we can localize faults, fix them, and gain generalizable knowledge about how, when, and where they are likely to occur. Over time—evolutionary, cultural, and ontogenetic—this knowledge and these reliable practices accumulate.[10] Why not develop a philosophical theory that makes *these* capacities or similar ones the fundamental facts that we idealize from: the generative powers of evolved, developed, socialized, and acculturated human agents?

The most general maxim for those who study functionally organized systems is that we come to understand how things work by studying how, when, and where they break down. This is true for those who seek to test theories or models and, more basically, for those who wish to figure out how they work. We can't see how to test a theory until we *do* understand how it works. In the real world, knowledge of how to use or test a theory does not come packaged with its axioms! We who are surrounded by technical systems often forget or underestimate the learned wizardry embedded in the knowledge and practices of those who work on them—heuristic knowledge of breakdowns; their likely causes; and how to find, fix, and prevent them. Auto mechanics, doctors, engineers, programmers, and other students of mechanisms learn how to debug their preferred systems by localizing and fixing the faults that occur; both in their machines and in their procedures for working on them. *This works for the mind no less than for any of our other tools.* With that understanding, we can analyze, calibrate, and debug both our reasons and our reasoning.

Now we have come to the crux of the issue: *we can't idealize deviations and errors out of existence in our normative theories because they are central to our methodology*. We are *error-prone* and *error tolerant*—errors are unavoidable in the fabric of our lives, so we are well adapted to living with and learning from them. We learn more when things break down than when they work right. *Cognitively speaking, we metabolize mistakes!* Feedback mechanisms provide pervasive and often automatic means of error correction at all levels of biological and cognitive activity. We are particularly tuned to detecting and working with violated expectations, and more generally, with differences.

Errors are often sources of creative elaboration. Model building, experimental design, and software design exhibit this lesson in rich detail. "Bugs," incidental features of programs, models, designs, or things designed for another purpose often stimulate or initiate a useful result or capacity. Mistakes or non-adaptive features may get special attention because they aren't functional, and then inspire new ideas. The creative role of errors and "neutral" traits also emerges from a closer look at how theories and experimental designs are actually applied and implemented.[11] (Published reports of experiments would lead you to believe that everything was planned in advance!) Gould and Vrba's (1982) "exaptations" are features co-opted to serve new functions in new contexts, to provide new sources for evolutionary innovation and selection. They argued the need to replace the older term "preadaptation," which they thought inappropriately suggested a kind of prescience. *Exaptations* are fortuitous—not error exactly, but not intended either. Most organic features have been exaptations—the more entrenched ones many times over, each time midwifing a new kind of adaptive complex, and acquiring new layers of functional significance. Changes that make traits *vestigial* signal a change in design direction, and are often the mark of past exaptations. Histories of design are littered with changes in direction, both in nature and in our record of technological and cultural artifacts.

Our normative models should reflect how we learn, and we learn from our mistakes. Some errors are important, some aren't. Which ones are important may depend on your question. Some are easier to find, and some are very informative—errors with riches to be found. Heuristic principles leave tracks or footprints—signs of their application—in their systematic errors, failures, or biases through which we can recognize them (see Chapter 5 and appendixes A and B). Some tracks are diagnostic for their causes and others relatively non-specific, with alternative possible causes. These failures can be positive tools but they're not always easy to find. We can be quite resistant to seeing them. They may be too local or too global. Some tracks may be very small, or only found under rocks, or on the underside of leaves, or only visible in ultraviolet light; so if you don't know what to look for, and how, and where, you may not find them. That may take a specialist's knowledge of the area—thus our dependence on experts. So it is with the biases of most heuristics.

For a global perspective, there's the Charles Addams cartoon of the early 1960s: Two pith-helmeted types have pulled into a large shallow depression, and parked their jeep in a southwestern desert landscape punctuated with saguaro cactus and occasional mesas. They are now taking a rest break. One says to the other: "I guess you're right. It was just one of those wild reports." Your viewpoint is back a way, and *you* can see that they have just parked in a giant footprint. Hard for them to see, but just what we don't want to miss: so fundamental that it affects everything in sight, but hard to find a place to stand to view it—or to see where it is not. Here a specialist's knowledge may not help, because it may lead us to focus too closely. It is just this kind of mistake we make with our models of rationality. As with the cartoon, it may sometimes be better to stand back from the search.

What we really need to avoid is not errors, but significant ones from which we can't recover. Even significant errors are okay as long as they're easy to find. With enough time, or enough of us working on the problem from enough different perspectives, or all of the above, we assume we will find them—maybe even all of them, though we'll never know that for sure. This last qualification slips out effortlessly, like a mumbled incantation. But let's not get distracted. We've been sitting in the skeptics pew for 350 years, so it is time to move on to more productive activities.

This positive view of error yields a richer and more natural model of our rationality and decision making than those recently used by decision makers, economists, and philosophers (Nozick, 1993, and Williams, 1996, offer other arguments to this conclusion). This new picture is full of context-sensitive heuristic rules of procedure efficiently tuned to generate solutions for specific situations, with learned (and often *labeled*) warnings as to when breakdowns may occur. Less elegant, but it usually works better. More accurate, not only in characterizing the decisions we make, but also in describing and analyzing the processes by which we reach them. And well tuned to the adaptive complexity we find in nature. The recommendations of this new model are different, though often refreshingly familiar from other "real world" contexts. They have been and should continue to be a lot more fruitful, and this model *doesn't* lose its normative force by being descriptively accurate. That's good for a change.

CHAPTER THREE

Toward a Philosophy for Limited Beings

The Stance and Outlook of a Scientifically Informed Philosophy of Science

An adequate philosophy of science should have normative force. It should help us to do science or, more likely, to find and help us avoid sources of error, since scientific methodologies are by nature open-ended. Without being normative it is not a philosophical account. Mere descriptions of scientific practice, no matter how general or sensitive to detail, will not do. Without normative force, studies of methodology, however interesting, would translate as a catalogue of fortuitous and mysterious particular accidents, with no method at all. So the "special sciences" (Fodor, 1974) can't be *too* special, as I'll argue below.

These lessons should be keyed much more closely to actual scientific practice than is common in philosophy. We should be prepared to do field work in the sciences. We need to learn to speak the language, but even more to understand and to value their practices as a native, while still retaining the conceptual and methodological interests of the philosopher. We will sometimes need *knowing-how* (and *knowing-why,* a comparable knowledge of, access to, and sometimes even identification with their internalized values) as well as *knowing-that,* to be able to write about it competently.[1] And we should be activists, reapplying these lessons to the practice of science. Don't be afraid to *do* science, if that seems to be called for to get the necessary perspective!

We shouldn't be afraid to give advice to scientists, but we must be humbler about it. This is the price of privilege. We are not "keepers of the logic of science," revealing foundational conceptual truths prior to any possible science. Our advice should be contextual and sensitive to feedback, not a priori pronouncements offered *ex cathedra*. (No one ever asks for such help, or takes it willingly!) Such an a prioristic view is fundamentally incompatible with the values of science, and self-compromising for anyone committed to adopting coherent and compatible views of scientific and philosophical method. To use the two together, we need a different view.

If we can understand the science from the inside while retaining a philosophical perspective, we can gain a new and important viewpoint on scientific practice. However, this is ultimately just another kind of empirically informed and thus empirically infirmable theoretical perspective. We will be occupationally more self-reflective, but not immune to error. Statisticians are uncommon among mathematicians for their extensive attention to real-world cases and reasoning processes. Like them, we should be *theorists* of reason, and at the same time among the most applied *therapists* of reason.[2] Like them also, we should be as interested in the detection of pattern as in what inferences we draw from it. We should explore the apparently unsystematic kluges and patches as well as paradigmatic cases of good science to find and criticize new handles for the phenomena. In this, statisticians are finally discoverers or inventers of reason—or at least the first taxonomists of new reasoning processes that they find.

This new attention to scientific practice suggests the image of many philosophies for the many sciences. But for all their particularity and context sensitivity—hallmarks of "the special sciences"—and the new *disunity of science* movement, these new lessons *are* widely projectible across the sciences. In our flight from a monolithic and exceptionless logic of science we should not miss the many techniques that are wide but not universal in scope—a "toolbox of science." We advocates of the former unity of science movement have just sought too much projectibility in terms of all-too-simple and context-free unqualified universals. That's not to be had. But there are many sound *ceteris paribus* generalizations cross-cutting traditional disciplines. *Maxims of reductionistic problem-solving* and *maxims of functional inference* apply generally to these kinds of problems, with qualifications appropriate to different areas. Neither set captures all inferences in given disciplines. The historical sciences share many methodological approaches. The problems and methods arising from the conjoint complementary use of individual and populational (or qualitative and quantitative) data yield insights across many disciplines. Other disciplinary special knowledge includes how tools commonly used together interact for that specific subject matter.

We need a new organization of methodological knowledge around families of heuristics that are applied to characteristic kinds of problems across disciplines, and how they interact with one another, with specific natural constraints, and with other practices for managing our subject matters. This yields breadth greater than that of the special sciences, but only via a new taxonomy. Classifying together families of related heuristics should simplify the task of going from footprints to heuristics, and then back to the probable other biases and corrective therapy (complementary heuristics) required. (A beginning list and classification, particularly of reductionistic heuristics, is found in Appendix B.)

Ceteris Paribus, Complexity, and Philosophical Method

The projectible maxims should tell us new things about *philosophical* methods as well. Studies of scientific methodology can be philosophically innovative. Witness the revolutionary systems of great past philosopher-scientists or "natural philosophers": Aristotle, Galileo, Descartes, Newton, Leibniz, Hume, Kant, and Darwin. An important lesson for philosophical methodology—with seeds sown by Darwin in his move from essentialistic to populational thinking—is that in complex systems you should expect essentially all of your interesting generalizations to be of the *ceteris paribus* kind. They will have an indefinitely large but unspecified class of exceptions, with a good set of conditions and heuristics for when to expect them. Complex systems have too many degrees of freedom, too many ways both for accomplishing something and for failing to do so to expect simple, exceptionless generalizations to work. However, unless they *characteristically* worked, they wouldn't be here. (That is, *we* wouldn't!) Characteristic behavior provides a basis for generalization, but for such generalizations, exceptions are the rule. Cartwright (1983, 1989) and Wimsatt (Chapter 9; 1972, 1976b) have both written about this from complementary perspectives.

These *ceteris paribus* clauses are not detachable, as they might be for simpler systems where we only worry about proper system closure—the idealized unreal images of economic theory, for example. They are intrinsic to the generalizations and to the systems they describe, which makes them characteristic of any patterns produced by a selection process and operating in a population: not only will individual members of that population differ, but evolution feeds opportunistically on outcomes, manipulating probabilities by diverse paths so that *multiple realizability and multiple exceptions* are inevitable. We hide this with in principle assumptions that if we knew everything, the (complex) unqualified universals would emerge. Thus, our attitudes toward reductionism are flawed by failing to correct appropriately for the incompleteness of our knowledge and the nature of our heuristics (Chapter 10). We don't yet have adequate philosophical and conceptual tools for dealing with uncertain messy situations. I provide a "starter set" (see chapters 4, 5, 6, 7, and 10), but more is required.

My intended audience is philosophers, to get them to change their practice, and also scientists, in the hope that these ideas might be genuinely helpful in their practice. There *is* room for a theory of practice—a genuinely philosophical theory with lessons in how to do philosophy as well as philosophy of science.[3] These are case studies in and of this practice. "But where's the theory?" one might say after reading this collection. "I see some localized conceptual analysis, a lot of methodological exegesis, a little history of science, some thinly disguised theoretical science, and probably a bit too much preaching. There's some pretty applied philosophy in there, but there's no theory of practice." To understand why this complaint is ultimately misplaced, note that philosophers sometimes use *theory* interchangeably for theory and for *meta*-theory. We distinguish the two clearly in principle, but not always very well in practice. Thus the so-called identity theory of mind was not a specific account of what specific kinds of mental things were identical with what specific kinds of physical things. That would have been a scientific theory. Philosophers may be interested in such theories, but don't normally propose them. Rather, it was a philosophical theory—a meta-theory for scientific identity theories of mind—of what kinds of constraints apply to and what kinds of consequences would follow from any specific scientific theory that made such an identification.

Philosophers tend to presume that one could have true and interesting things to say about this in general, even in advance of actually seeing the theory. I don't believe it! This is another kind of slippery inference from in principle thinking, an assumption of a kind of cognitive omniscience or omnipotence. *General meta-theories are particularly vulnerable to discovery of new ways to reframe the question.* Such attitudes occasion justifiable suspicion among scientists! They shouldn't if such judgements were advanced as the quite abstract and theoretical but still empirical hypotheses that they are, and defended accordingly. They are aids in exploring the space of theories of a given kind, and thus productive for both philosophy and science. But historically philosophers have done no better at these kinds of projections than scientists. Neither of us have done very well, at least by standards encouraged by a search for certainty! The history of scientific progress and the evolution of our conceptual categories is littered with one generation's projects and category mistakes that have become the next generation's impossibilities and conceptual truths. What has happened to the supposed fixity of species? The supposed incoherence of infinitesimals? The "sensible" program for the reduction of mathematics to logic? The "natural" affinities between determinism and predictability given the claims of chaos? We should be more careful and modest about meta-theoretical claims.

I have no complete and systematic meta-theory of practice to offer here, though these chapters point often in the direction of one. There are specific theories of practice for different important practices in different circumstances; there are many more to be studied. As normative theories of reasoning—of problem solving in various kinds of scientific contexts—they are still of philosophical interest. I believe that these theories—offered piecemeal, but articulating well together—are both descriptively and normatively more adequate to the practice of science than anything you will find in your philosophy. Often, there is no competing philosophical theory of them (or a scientific one either) since I urge a conception of philosophy reaching beyond its current boundaries. There is also meta-theory, offered in justification of specific types of practices, but not yet a *systematic* meta-theory. When it does arrive, it should be as full of *ceteris paribus* qualifiers as its subject matter. Meanwhile, I have neither aimed for completeness nor marked all meta-theoretic remarks as such. Adequate meta-theory should flow from such specific theories of practice after we have made them ours—to realize what they can do so we can generalize appropriately. But that is another book for another time and perhaps for another person.

My motivation here is different. I am by choice a conceptual engineer, not a pure theoretician. If what I have to say is not useful to scientists I have not yet done my job right. This additional constraint makes a harder task than philosophers usually set for themselves. I may not have succeeded. I may be wrong. More often, despite my best efforts, what I offer will still be too abstract to be immediately useful for practicing scientists. This is not yet a tried-and-true recipe book of laboratory formulas for conceptualization, model-building, criticism, and experimental design. For concrete paradigms of analogous texts within the sciences, see the illuminating and entertaining but useful laboratory preparations and formulas in the venerable yearly editions of the *Handbook of Chemistry and Physics,* or the classic *Procedures in Experimental Physics* of John Strong (already in its seventeenth edition since 1938 when I bought my copy in 1959). These authors recognized well the need for piecemeal elaboration. To move in that direction is a worthy aim for any philosopher of science. The last section of Robert Nozick's (1993) *The Nature of Rationality* had some of this character for philosophical investigations.

I have worked mostly in the biological sciences, so my examples are closer to home there. But I am also moved by experience and continuing interests in mathematical modeling in engineering and the less foundational of the physical sciences, and in psychology and the social sciences. (I usually describe myself as a "philosopher of the inexact sciences"—see chapters 13 and 10.) Biology is often a sufficiently variegated paradigm to provide lessons for those areas too. The sciences I seek are the "everything in between" of the frontispiece.

Our real world is complex, and we are faced—as biological and social beings—with the need to make an ongoing stream of decisions that serve us well *here.* We surround ourselves with idealizations. We think we understand these idealized laboratory and conceptual worlds of our own construction, and it is all-too-tempting to refer any questions to one of them. But I am deeply impressed with the often successful strategies we have for dealing with complexity in the real worlds we inhabit. To study these strategies, we must study them *there,* not in one of the idealized constructions we have made for other purposes. We certainly will need new idealized constructions to study this cluster of problems, but idealizations made specifically for these purposes! Controlled and simplified laboratory worlds and idealizations are designed by cognitive scientists or decision theorists to reduce the response of an agent to a few—ideally to one—degrees of freedom in order to simplify theory testing. Such designs try to eliminate interaction effects so we rarely see any. They give few clues, often misleading ones, for how we do or should respond in worlds with many degrees of freedom and interacting constraints. Similarly, we get too easily enamoured of the tidy conceptualizations we model and claim logically necessary consequences from (when practicing "possible worlds semantics") as philosophers. For these problems, such model-worlds do not make enough discriminations among relevant kinds to be adequate models of our world.

Our Present and Future Naturalistic Philosophical Methods

Consider an analogy: suppose our current philosophical methods are like the ideal gas law and phenomenological thermodynamics—rooted in deeper theories, phenomena, and mechanisms of which we are currently unaware, and can normally ignore while confidently and reliably producing useful knowledge using these methods. They constitute idealized phenomenological theories for reasoning abstractly to and from first principles—based in logic, to be sure (what could be deeper than that?), but humor me. Suppose they are *sui generis,* and have no deeper foundation in other things. We take them as foundational to our methodology. Inspired by logic and sometimes by obsolete psychology, these idealizations are at best first-order approximations: *exceptionless generalizations; the search (always and exclusively) for logically valid arguments; analyses in terms of necessary and sufficient conditions; knowledge or belief structures that are supposedly free of inconsistency and closed under entailment; crisp inner-outer and subjective/objective distinctions and other structures from a view of natural order as a compositional hierarchy of logical schemata or computational algorithms; crisp fact-value distinctions, and the separate operation of modular and non-interacting cognitive, conative, and affective processes; the perceived gulf between issues of meaning and empirical fact (and the low status accorded the latter in argument!); the special attention paid to claims that hold in principle but not in practice. . . .* This list could go on for a long time. We have used these tools well when they work well, but now we push them too widely. They have become entrenched, masking their status in a naturalistic philosophy as merely tools.

For some parts of the world these tools and techniques can produce compact, simple, and elegant answers. There, world-pegs fit category-holes so closely that we don't notice the small shavings left over. There, philosophical analysis with present tools can and has made real progress. Like phenomenological thermodynamics, these answers are immensely useful, practically irreplaceable, and work well most of the time. But the world is also very complex, and our concepts and language are constantly changing. The orders that philosophy produces in this way are local approximations to temporal strobe-pictures of changing conceptual relations, end-directed cognitions, and social norms—modulated by affect, and differently seen and constituted by networks of embodied and socialized cognitive beings. Too many problem areas are not well illuminated by our philosophical lampposts—at least where they are currently placed.

This sounds problematic. Is this attacking logic? Logic doesn't break down by being approximately right. It is a reflection of the standing of logic in our scheme that we wouldn't think of violations in this way. I'm not attacking logic, but some of the tools we have crafted with it and other time-honored idealizations that seem almost part of it. These tools "break down" by failing to organize areas compactly and with closure when less exact methods do so much more simply.

The alternative methods I propose aren't necessarily empirical because they are "messy." Insisting on an exact, precise, complete, exceptionless description can hide important order that is there. Sometimes, if we are willing to tolerate a few exceptions, context-dependencies, and approximate truths—if we're willing to "defocus" a little—we can get nice, compact, *ceteris paribus* qualified but robust generalizations. That's still worth doing in the vast regions of our conceptual world—most of it—where that's the best we *can* do (Cartwright, 1983). It's also worth doing because it opens our world to a deeper explanation, for constrained fuzzy order is the general case: *on our time and size scale, and for several orders of magnitude in each direction, it is exceptionless crisp precision that is the degenerate special case. Nor are we—detecting and assessing this order—disembodied cognitive engines: we yoke cognition and affect quasi-independently in a noise-tolerant manner in the service of local and more global ends.*[4]

Ceteris paribus clauses are nice. They allow us the regularities and modularities we know are there while reminding us of the exceptions—fluctuations or deviations from a macroscopic order that point the way to a deeper understanding. This cognitive form (general pattern + exceptions) and its relatives (broad similarities + attendant differences, models + qualifications, etc.) are deeply anchored in the structure of our case-based organization of knowledge: "This is like that (which you already know) but with the following differences." It wears the microstructure of cumulative learning on its face. Similar schemata appear again and again in the chapters that follow. This form is intrinsic to multi-level or multi-perspectival theories (Chapter 8), to the use of models (Chapter 6), and to the broader robust generalizations that form the architecture of our real science (Chapter 4).

Brownian motion was the pivotal imperfection in the picture of macroscopic continuous inert fluids that let us see the inner workings of liquids and gases. The new statistical micro-theory didn't eliminate these macroscopic continua, despite what some have said! It explained their properties, and revealed other dimensions that enriched our view of them immensely. The micro-theory in condensed matter physics in which my kitchen table is mostly empty space between electron orbitals did not make it any less solid, real, or impenetrable to my finger. It can be both, once the different qualities of the *detectors* appropriate to these two levels are taken into account (Wimsatt, 1976a). Physical continua behave as continuous to macroscopic detectors in a *scale-dependent* way that changes and breaks down at lower levels. There is a conflict only if we try to see them on the analogy of the mathematical continuum of the real numbers, which is *scale-independent,* showing the properties of continua on all size scales. This point is nicely made for physical fractal properties by Mandelbrot (1982). This bi-level perspective on the table or other macro-level objects enriches our view of the world too. It shows how multiple legitimate (macro- and micro-) views of our objects are deeply imbedded in our physical world—not just a product of our subjectivities. (Brownian motion and a broader realist multi-perspectivalism both reappear in Chapter 10.)

I seek for philosophical methods the naturally rooted analogues to statistical mechanics; the "deeper" principles that yield and explain them in the limits, and explain when, how, and where fluctuations and deviations from those idealized pictures are possible and to be expected. From them, a naturalistic epistemology and a naturalistic methodology, extending to metaphysics and the valuational dimensions of human experience, should be both possible and inevitable. Naturalism need not be fundamentally eliminative or destructive of traditional views and methods. The essentials of philosophical methods can and must survive a thoroughgoing naturalism, but the naturalism includes societies, cultures, and ecosystems, with essentially all of our cognitive and cultural structures and regularities—descriptive, affective, and normative—intact as "phenomenological" laws and objects: reference groups, ideologies, and markets are as real as neurons, genes, and quarks. This is the vision of the aims of science and its objects exploited in Chapter 8. The tools we already have will look and act only a little differently than we thought—and mostly under unusual circumstances—but we will have a richer understanding of them. We will also have to learn how to use some powerful new tools, but for this minor inconvenience we get rich, deep, and robust connections with the world.

Some older tools of lesser importance should achieve a new centrality, and we may need to recognize or create some brand new ones. At least two topics should have a more central importance in this new epistemology: The first would be heuristics (their general nature, their variety, how we calibrate and debug them, the warnings appropriate to their use in various areas, their relations to counterfactuals and *ceteris paribus* clauses, and how more complex heuristics are articulated into methodologies or customized for specific applications). In many ways, heuristics could come to occupy the methodological position for theories of the middle-range of laws for fundamental theory; these theories will no longer be less important because they are less "fundamental."

The second topic would fit a metaphysics course, an epistemology course, or even an ethics course, depending upon how it was developed. We used to talk blithely about what objects to allow, and on what grounds—the topic of ontology—in traditional metaphysics courses. But objects have boundaries—why not study them?[5] We need to understand how we place or locate boundaries, and how they can be generated or changed over time, particularly with the kind of overlapping real-world complexities discussed in chapters 9 and 10. We need a metaphysics seminar on the nature, detection, and consequences—or as I like to say, the "care and feeding"—of *functional localization fallacies*. This term is common in methodological discussions in neurophysiology; close to the surface in the units of selection controversy in evolutionary biology; and crying out for recognition in the methodological individualism/methodological holism dispute in the social sciences, and in discussions of reductionism generally. It is such a general problem that it infects all areas of philosophy. What else is G. E. Moore's "Naturalistic fallacy" (if it is one) but a kind of *mis*-localization claim? (Moore should have been half as careful elsewhere: his "isolation test" for determining "intrinsic goodness" is one of the more glaring functional localization fallacies in the history of philosophy, since it assumes that intrinsic goodness cannot inhere in the relation of an entity to other entities.) To generalize: *when a system can be described at a variety of levels of organization, or from a variety of perspectives, how do you recognize when a property is attributed to it at the wrong level, or from the wrong perspective? And how do inferences go wrong when this happens?* This study is most naturally pursued through examples, which should lead to deeper characterizations of the problem.

I haven't provided all of the necessary tools in this work—or completed drawings of the new edifice. Neither have the eliminativists, and they want you to make do with nothing at your level in the meantime! Or deconstructionists, who in their haste to leave no stone unturned have left nothing more than word games. Other systematic philosophies—neo-Wittgensteinian, neo-Kantian, or others—have done no better. I have more faith in this multifaceted, multi-level world of intersecting appearances and real interactions than in any of theirs. Here are fairly finished drawings of some important rooms, and still incomplete sketches of the whole thing. I hope that this doesn't strain too many current conventions for architectural rendering for you to be able to see the inferential and presumptive modesty, and the power, coherence, and integrative promise of what I've got so far. A rain forest is a rich place after all, still far richer than we know.