To appear in V. Hardcastle, ed., Biology Meets Psychology: Conjectures, Connections, Constraints, MIT Press, forthcoming.

Supple Laws in Psychology and Biology

Mark A. Bedau
Department of Philosopy, Reed College
3203 SE Woodstock Blvd, Portland OR 97202
email: mab@reed.edu
voice: (503) 771-1112, ext. 7337
fax: (503) 777-7769

The nature and status of psychological laws are a long-standing controversy. I will argue that part of the controversy stems from the distinctive nature of an important subset of those laws, which I'll call "supple laws." An emergent-model strategy taken by the new interdisciplinary field of artificial life provides a strikingly successful understanding of analogously supple laws in biology. So, after reviewing the failures of the two evident strategies for understanding supple psychological laws, I'll turn for inspiration to emergent-models explanations of supple laws in biology. I'll conclude by inferring what an emergent model of supple laws in psychology should be like.

Supple Laws in Psychology

It has long been noticed that the regularities and patterns in our mental lives—what I'll call (without attempting to prejudge any questions) "psychological laws"—need ceteris paribus qualifications, i.e., qualifications to the effect that the law holds only provided "everything else is equal." Two typical examples, though extremely simplified, clearly illustrate this phenomenon:

Pure Reason: If A believes P and A believes that P implies Q, then <u>ceteris paribus</u> A will infer Q.

> **Practical Reason**: If A wants G and A believes that M will produce G, then <u>ceteris paribus</u> A will do M.

In general, all psychological laws stand in need of similar <u>ceteris paribus</u> qualifications.

A variety of factors bring about the need for <u>ceteris paribus</u> qualifications in psychological laws. People sometimes fail to infer what is implied by their antecedent beliefs because of inattention or illogic, but some exceptions to the law of Pure Reason reflect attentive logical acumen at its best. For example, if agent A who believes proposition P and believes that P implies proposition Q has good antecedent reason to <u>doubt</u> Q, A might conclude that it is more reasonable to question P or to question whether P entails Q. Similarly, some exceptions to the law of Practical Reason are due to confusion or weakness of will, but other reflect an apt balance of priorities. To use an example from Horgan and Tienson (1989), if some agent A wants a beer and believes that there is one in the kitchen, then A will go get one—unless, as the <u>ceteris paribus</u> clause signals, A does not want to miss any of the conversation, or A does not want to offend the speaker by leaving in mid sentence, or A does not want to drink beer in front of his mother-in-law, or A thinks he should, instead, flee the house since it is on fire, etc.

The status and source of psychological <u>ceteris paribus</u> laws are quite controversial (see, e.g., Hempel 1965, Dennett 1971, Putnam 1973, 1975, Fodor 1981, Cartwright 1983, Horgan & Tienson 1989, 1990, Schiffer 1991, Fodor 1991, Dreyfus 1992, Cartwright 1995, Horgan & Tienson 1996). (I should note that I'm concerned with the controversies about <u>ceteris paribus</u> laws that <u>describe</u> human psychology, not the controversies discussed by Dennett (1984) about how human intelligence <u>employs</u> <u>ceteris paribus</u> reasoning.) Misgivings about the status of psychological <u>ceteris paribus</u> laws are quite varied, ranging from worries about whether they are trivially or logically or analytically true, to worries about whether they are falsifiable or usable in scientific explanations, to worries about whether they can be precisely specified, even in principle, in some algorithm. A similarly bewilderingly variety characterizes the alleged source of the <u>ceteris paribus</u> qualifications; here we find reference to idealizations (Cartwright 1983), implementation-level malfunctions

(Dennett 1971, Putnam 1973, 1975, Fodor 1981, Cartwright 1995), the subtle and complex nature of human cognition (Horgan & Tienson 1989, 1990, 1996), as well as "a background of practices which are the condition of the possibility of all rulelike activity" (Dreyfus 1992, p. 57).

A common thread through the bulk of this diversity of opinion is the assumption that ceteris paribus laws have exceptions when they "go wrong." Now, in a trivial sense this is right—an exception, after all, is an exception—but in a deeper sense this is not right. For note that exceptions to psychological ceteris paribus laws fall into two quite different groups, what I'll call "rule-breaking" and "rule-proving" exceptions. Consider the law of Practical Reason, let A, G, and M be particular agents, goals, and actions, and assume that A wants G and believes that M will produce G. A rule-breaking exception happens when A nevertheless performs some other action M* (M*≠M) even though M is the most reasonable or sensible thing for A to do, given the whole constellation of A's beliefs, desires, capacities, etc. By contrast, a rule-proving exception happens when A performs M* because in this particular situation, given the constellation of the rest of A's beliefs, desires, capacities, etc., M* is more reasonable or sensible for A to do. One could say that, while rule-breaking exceptions involve the "wrong" thing in context happening, rule-proving exceptions involve the "right" thing happening. An exception that proves the rule is appropriate in the context since it achieves the agent's underlying goals better than slavishly following the rule would have and, furthermore, the exception happens because it is appropriate in this way.

Those ceteris paribus laws that have exceptions that prove the rule I will call "supple." All ceteris paribus laws are vague because they describe regularities that hold only for the most part but without delineating what conditions give rise to exceptions. The distinctive feature of supple ceteris paribus laws is that their vagueness has a special source—a certain kind of underlying regularity that explains the supple law. Supple laws have three defining features. The first has to do with how the supple law manifests a deeper, context-sensitive regularity. In typical contexts, the underlying regularity is manifested in the pattern of behavior described by the supple law. But in other contexts the same underlying regularity generates exceptions to the supple law. These exceptions "prove the rule" because they reveal the underlying regularity behind the supple law, they indicate the true "meaning" of the supple law.

The second defining feature of supple laws is that their "meaning" is teleological and derives from the telic nature of the underlying regularity. It's not that the underlying regularity <u>has</u> a purpose but that it <u>describes</u> the way in which some purpose or function is achieved. The supple law describes how that purpose is achieved in typical contexts and rule-proving exceptions arise in contexts in which some other means better achieves the same purpose. The teleology in supple laws can be mental but it can also be merely biological, as examples below show. (Further details of my preferred understanding of teleology are developed in Bedau 1990, 1991, 1992a, 1992b, 1993, Bedau & Packard 1992.) Hofstadter (1985) describes mental regularities as "fluid" and Horgan and Tienson (1990) talk of "soft" laws of intentional psychology. But the teleology in supple laws makes them not just "fluid" or "soft" but <u>aptly</u> so. The fluidity or softness of supple laws involves the open-ended context-sensitivity with which some purpose or function is achieved.

The third defining feature is exactly this open-ended context sensitivity of the underlying regularity. For one thing, the purpose in question is achieved in an indefinite number of different ways in an indefinite number of different contexts. But more than this, there may be no rule or algorithm for determining how to achieve the purpose given an arbitrary context; in general, nothing short of trial and error will suffice. Nevertheless, what makes a supple law <u>supple</u> is that, in an indefinite number of different contexts, in one way or another, the purpose captured by the underlying regularity <u>is</u> achieved. In other words, a law is supple only if the law actually has rule-proving exceptions in (enough of) those contexts in which slavishly following the rule would defeat the purpose in question.

Pure Reason and Practical Reason, our two sample psychological <u>ceteris paribus</u> laws, can plausibly be seen as supple laws, for each can be seen as the manifestation of a deeper regularity that concerns how some specific purpose is achieved in an indefinitely open-ended variety of contexts. Consider Pure Reason first. I assume that it's no accident that people tend to infer the consequences of their antecedent beliefs. Presumably, the purpose served by this process is something like having one's beliefs reflect reality as accurately as possible. (I don't mean that individual agent's have this as the conscious intention for fixing their beliefs, of course; moreover I'm not wedded to this particular view of the purpose of forming beliefs.) In addition, underlying the law of Pure Reason is a regularity about reasonable belief formation: ceteris paribus, people infer whatever is most reasonable given what else they believe. In typical contexts, then, when some person A believes both some proposition P and that P implies another proposition Q, then A will infer Q because this is the most reasonable inference to make in the context. In this way, the law of Pure Reason is a manifestation of the deeper regularity about reasonable beliefs. Furthermore, the purpose behind the law of Pure Reason derives from, and is the same as, the purpose behind the reasonable belief regularity. Finally, in those contexts in which it is not most reasonable to infer Q, the reasonable belief regularity will be manifested in some other way, such as inferring that P is false or that P does not imply Q. This will constitute an exception to the law of Pure Reason but one which proves the rule.

Similarly, underlying the law of Practical Reason is a regularity about reasonable action: ceteris paribus, people act in whatever way is most reasonable given their beliefs and desires. In addition, the purpose behind the law of Practical Reason derives from, and is identical to, the purpose captured by the reasonable action regularity, which is presumably something like acting so as to best serve one's needs and desires. The reasonable action regularity manifests itself in typical contexts as the law of Practical Reason, but in other contexts the most reasonable action might break the law of Practical Reason.

Note that ceteris paribus clauses appear in the statements of the reasonable belief regularity and the reasonable action regularities, even though these are the regularities that explain the suppleness of Pure and Practical Reason. This should tip us off that the ceteris paribus clauses in psychological laws cover two quite different kinds of exceptions. Rule-breaking exceptions can still remain after all rule-proving exceptions have been removed. As I mentioned above, not all exceptions to supple laws ar due to their suppleness. Suppleness is an aspect of some ceteris paribus laws but it is not the full explanation of ceteris paribus qualifications in any laws. Since there are at least two quite different kinds of reasons why ceteris can fail to be paribus, there is no such thing as the analysis of ceteris paribus laws.

I have been arguing that the suppleness in psychological laws reflects an open-ended dynamic in a mental system's appropriate adaptation to novel contextual changes. In fact, this supple adaptability in psychological processes is a hallmark of their intelligence. Descartes put his finger on precisely this sign of true intelligence in Part V of the Discourse on Method when he described the difference between human rationality and the behavior of mere machines:

> . . . although [mere machines] perform many tasks very well or perhaps can do them better than any of us, they inevitably fail in other tasks; by this means one would discover that they do not act through knowledge, but only through the disposition of their organs. For while reason is a universal instrument that can be of help in all sorts of circumstances, these organs require a particular disposition for each particular action; consequently, it is morally impossible for there to be enough different devices in a machine to make it act in all of life's situations in the same way as our reason makes us act. (trans. D. A. Cress)

Descartes does not just point out that a hallmark of rational creatures is an open-ended flexibility in their ability to act appropriately as contexts change. He also claims that no mere machine could exhibit this suppleness of behavior. I'll argue in a moment that he is both right and wrong about this: right because no fixed machine can be supple, and wrong because a suitably changing mechanism can be supple. But I whole heartedly agree with Descartes that suppleness is central sign of genuine intelligence. Ceteris paribus laws are often treated as an embarrassing curiosity, to be ignored or excused. The lesson to be learned from Descartes is that attempts to properly describe and explain supple psychological laws should be celebrated and made a central focus of psychology and the philosophy of mind.

A good description and explanation of a supple ceteris paribus law of psychology should have certain virtues. First, it should be precise, specifying when ceteris is not paribus, at least for those exceptions that prove the rule. Second it should be accurate and complete, in the sense that it has no false positives (cases falsely advertised as rule-proving exceptions) and no false negatives (cases falsely advertised as law-conforming). Third, it should be principled, not arbitrary or ad hoc; it should indicate what underlying regularity unifies the law and its rule-proving exceptions. Finally, it should be feasible, consistent with what we know about the natural world, and allowing us to test empirically the account's accuracy and completeness.

There are two evident strategies for describing and explaining supple psychological laws, and neither is very good. The first strategy—what I'll call the common-sense strategy—engages in hand waving by employing ceteris paribus clauses within the account. One version of the common-sense strategy takes the supple law as a brute fact, thus sacrificing a principled explanation of the supple law. Another variant of this strategy avoids this problem by appealing directly to an underlying regularity, such as the reasonable belief and reasonable action regularities I described above.

The common-sense strategy can achieve a sort of descriptive adequacy, as far as it goes, for its descriptions appeal to the supple law or its underlying regularity and these do obtain. But the strategy lacks all the virtues sought in an account of suppleness. First, no serious attempt is made to indicate under what conditions rule-proving exceptions occur. Appealing to an underlying regularity does provide a minimal explanation of what causes rule-proving exceptions. But if the underlying regularities are like those I sketched above (e.g., ceteris paribus, people infer whatever is most reasonable given what else they believe), they will invoke vague phrases like "whatever is most reasonable" or "as appropriate" and thus will not precisely indicate when to expect a rule-proving exception. This leads to the other flaws with the common-sense strategy. For one thing, a common-sense description and explanation will be too imprecise to be accurate or complete. It might not be able to be convicted of advertising, but that is because it makes no precise and testable advertisements. Furthermore, its imprecision blocks us from testing whether it fits with our understanding of the natural world, so the common-sense strategy is unfeasible.

An alternative strategy for accounting for the suppleness of mental processes is, in effect, to predigest the circumstances that give rise to rule-proving exceptions and then specify (either explicitly or through heuristics) how and when they arise in a manner that is precise enough to be expressed as an algorithm. This is exactly the strategy followed by so-called "expert systems" in artificial intelligence. This expert-systems strategy yields models of supple psychological laws that are explicit and precise enough to be implemented as a computer model, and this gives the strategy three important virtues. First, it obviously is precise, since an expert system will indicate precisely when ceteris is not paribus. Second, an expert systems account is principled, provided the experts' information is. Third, an expert-systems account is feasible (barring some problem with the algorithm), so the account's dynamic behavior can be directly observed and tested for plausibility.

There are well-known problem with expert-systems, though. (See, e.g., Dreyfus 1992, Hofstadter 1985, Holland 1986, Langton 1989a, Horgan & Tienson 1989, 1990, Chalmers, French, & Hofstadter 1992.) They sometimes work well in precisely circumscribed domains, but they systematically fail to produce the kind of supple behavior that is characteristic of intelligent response to an open-ended variety of circumstances. Their behavior is brittle—they lack the context sensitivity that is distinctive of intelligence—so they are inaccurate and incomplete. This brittleness shows no evidence of being merely a limitation with present expert systems; attempts to improve matters by amplifying the knowledge base only invite combinatorial explosion. Although precise and feasible and principled, their brittleness makes expert-systems accounts of supple mental dynamics inaccurate and incomplete. Our experience with expert systems suggest that Descartes was right that the aptness of supple psychological processes is too open-ended to be embodied in any fixed mechanism or algorithm.

I want to promote a third strategy for describing and explaining supple psychological laws. I'm not in a position today to give a concrete illustration of this strategy, so I will do the next best thing and give a concrete illustration of an analogous strategy in an analogous context: emergent artificial life models of supple biological laws. I'll call this approach the emergent-model strategy since its leading idea is that supple macro-level laws implicitly emerge from an evolving population of explicitly interacting micro-level entities. The result is to view the supple law as produced by a mechanism—a micro-level population of interacting agents—but a mechanism with a build-in capacity to make adaptive changes as the environmental context alters. Emergent models provide a distinctive, indirect but constructive kind of explanation of supple phenomena. My overall conclusion will be that the emergent-model strategy is an adequate way, and the only evident adequate way, to understand supple laws, and my central argument will be an appeal to the analogy with artificial life.

Emergent Models of Supple Biological Laws

Evolving populations display various macro-level patterns on an evolutionary time scale. For example, adaptive innovations that arise through genetic changes tend, <u>ceteris paribus</u>, to persist and spread through the population so as to maximize the population's adaptive fit with its environment. Of course, these patterns are not precise and exceptionless universal generalizations; they hold only for the most part. When their vagueness is due to context-dependent fluctuations in what is appropriate, the macro-level evolutionary dynamics are supple in the sense intended here. These sorts of supple dynamics of adaptation result not from any explicit macro-level control (e.g., God does not adjust allele frequencies so as to make creatures well adapted to their environment); rather, they emerge statistically from the micro-level contingencies of natural selection.

The new interdisciplinary field of artificial life is exploring a certain characteristic kind of computer model of evolutionary dynamics. These emergent models (as I'll call them) consist of a micro-level and a macro-level. (I should stress that I am using "micro" and "macro" in a generalized sense. Micro-level entities need not be literally microscopic; individual organisms are not. "Micro" and "macro" are relative terms; an entity exists at a micro-level relative to a macro-level population of similar micro-level entities. These levels can be nested. Relative to a population, an individual organism is a micro-level entity; but an individual organisms is a macro-level object relative to the micro-level genetic elements (say) that determine the organism's behavioral strategy.) Emergent models generate complex macro-level dynamics from simple micro-level mechanisms in a characteristic way. This form of emergence arises in contexts in which there is a system, call it <u>S</u>, composed out of micro-level parts. The number and identity of these parts might change over time. <u>S</u> has various macro-level states (macrostates) and various micro-level states (microstates). <u>S</u>'s microstates are the states of its parts. <u>S</u>'s macrostates are structural properties constituted wholly out of microstates; macrostates typically are various kinds of statistical averages over microstates. Further, there is a relatively simple and implementable micro-level mechanism, call it <u>M</u>, which governs the time evolution of <u>S</u>'s microstates. In general, the microstate of a given part of the system at a given time is a result of the microstates of nearby parts of the system at preceding times. Given these assumptions, I will say that a macrostate P of system S with micro-level mechanism M is emergent if and only if P (of system S) can be explained from M, given complete knowledge of external conditions, but P can be predicted (with complete certainty) from M *only* by simulating M, even given complete knowledge of external conditions. So, we can say that a model is emergent if and only if its macrostates are emergent in the sense just defined.

Although this is not the occasion to develop and defend this concept of emergence (see Bedau 1997), I should clarify three things. First, "external conditions" are conditions affecting the system's microstates that are extraneous to the system itself and its micro-level mechanism. One kind of external condition is the system's initial condition. If the system is open, then another kind of external condition is the contingencies of the flux of parts and states into S. If the micro-level mechanism is nondeterministic, then each nondeterministic effect is another external condition.

Second, given the system's initial condition and other external conditions, the micro-level mechanism completely determines each successive microstate of the system. The macrostate P is a structural property constituted out of the system's microstates. Thus, the external conditions and the micro-level mechanism completely determine whether or not P obtains. In this specific sense, the micro-level mechanism plus the external conditions "explain" P. One must not expect too much from these explanations. For one thing, the explanation depends on the massive contingencies under the initial conditions. It is awash with accidental information about S's parts. Furthermore, the explanation might be too detailed for anyone to survey or grasp. It might even obscure a simpler, macro-level explanation that unifies systems with different external conditions and different micro-level mechanisms. Nevertheless, since the micro-level mechanism and external conditions determine P, they explain P.

Third, in principle we can always predict S's behavior with complete certainty, for given the micro-level mechanism and external conditions we can always simulate S as accurately as we want. Thus, the issue is not whether S's behavior is predictable—it is, trivially—but whether we can predict S's behavior only by simulating S. When trying to predict a system's emergent behavior, in general one has no choice but simulation. This notion of predictability only through simulation is not anthropocentric; nor is it a product of some specifically human cognitive limitation. Even a Laplacian supercalculator would need to observe simulations to discover a system's emergent macrostates.

Norman Packard devised a simple emergent model of evolving sensory-motor agents which demonstrates how supple, macro-level evolutionary dynamics can emerge implicitly from an explicit micro-level model (Packard 1989, Bedau & Packard 1992, Bedau, Ronneburg, & Zwick 1992, Bedau & Bahm 1994, Bedau 1994, Bedau & Seymour 1994, Bedau 1995). What motivates this model is the view that evolving life is typified by a population of agents whose continued existence depends on their sensory-motor functionality, i.e., their success at using local sensory information to direct their actions in such a way that they can find and process the resources they need to survive and flourish. Thus, information processing and resource processing are the two internal processes that dominate the agents' lives, and their primary goal—whether or not they know this—is to enhance their sensory-motor functionality by coordinating these internal processes. Since the requirements of sensory-motor functionality may well alter as the context of evolution changes, continued viability and vitality requires that sensory-motor functionality can adapt in an open-ended, autonomous fashion. Packard's model attempts to capture an especially simple form of this open-ended, autonomous evolutionary adaptation.

The model consists of a finite two-dimensional world with a periodically replenished resource distribution and a population of agents. An agent's survival and reproduction are determined by the extent to which it finds enough resources to stay alive and reproduce, and an agent's ability to find resources depends on its sensory-motor functionality—that is, the way in which the agent's perception of its contingent local environment affects its behavior in that environment. An agent's sensory-motor functionality is encoded in a set of genes, and these genes can mutate when an agent reproduces. Thus, on an evolutionary time scale, the process of natural selection implicitly adapts the population's sensory-motor strategies to the environment. Furthermore, the agents' actions change the environment because agents consume resources and compete with each other for space. This entails that the mixture of sensory-motor strategies in the population at a given moment is a significant component of the environment that affects the subsequent evolution of those strategies. Thus, the fitness function in

Packard's model—what it takes to survive and reproduce—is constantly buffeted by the contingencies of natural selection and unpredictably changes (Packard 1989).

All macro-level evolutionary dynamics produced by this model ultimately are the result of explicit micro-level mechanisms acting on external conditions. The model starts with a population of agents with randomly chosen sensory-motor strategies, and the subsequent model dynamics explicitly controls only local micro-level states: resources are locally replenished, an agent's genetically encoded sensory-motor strategy determines its local behavior, an agent's behavior in its local environment determines its internal resource level, an agent's internal resource level determines whether it survives and reproduces, and genes randomly mutate during reproduction. Each agent is autonomous in the sense that its behavior is determined solely by the environmentally sensitive dictates of its own sensory-motor strategy. On an evolutionary time scale these sensory-motor strategies are continually refashioned by the historical contingencies of natural selection. The model generates macro-level evolutionary dynamics only as the indirect product of an unpredictably shifting agglomeration of directly controlled micro-level events (individual actions, births, deaths, mutations). The model has no provisions for explicit control of macro-level dynamics. Moreover, macro-level evolutionary dynamics are typically emergent in the sense that, although constituted and generated solely by the micro-level dynamic, they can be derived only through simulations.

It should be noted that Packard's model is not intended as a realistic simulation of some actual biological population. Rather, it is an "idea" model, aiming to capture the key abstract principles at work in evolving systems generally. Packard's model is in effect a thought experiment—but an emergent thought experiment (Bedau 1998). As with the armchair thought experiments familiar to philosophers, Packard's model attempt to answer "What if X?" questions, but what is distinctive about emergent thought experiments is that what they reveal can be discerned only by simulation.

I will illustrate the emergent supple dynamics in Packard's model in some recent work concerning the evolution of evolvability. The ability to successfully adapt depends on the availability of viable evolutionary alternatives. An appropriate quantity of alternatives can make evolution easy; too many or too few can make evolution difficult or even impossible.

For example, in Packard's model, the population can evolve better sensory-motor strategies only if it can test sufficiently many sufficiently novel strategies; in short, the system needs a capacity for evolutionary "creativity." At the same time, the population's sensory-motor strategies can adapt to a given environment only if strategies that prove beneficial can persist in the gene pool; in short, the system needs a capacity for evolutionary "memory."

Perhaps the simplest mechanism that simultaneously affects both memory and creativity is the mutation rate. The lower the mutation rate, the greater the number of genetic strategies remembered from parents. At the same time, the higher the mutation rate, the greater the number of creative genetic strategies introduced with children. Successful adaptability requires that these competing demands for memory and creativity be suitably balanced. Too much mutation (not enough memory) will continually flood the population with new random strategies; too little mutation (not enough creativity) will tend to freeze the population at arbitrary strategies. Successful evolutionary adaptation requires a mutation rate suitably intermediate between these extremes. Furthermore, a suitably balanced mutation rate might not remain fixed, for the balance point could shift as the context of evolution changes. One would think, then, that any evolutionary process that could continually support evolving life must have the capacity to adapt automatically to this shifting balance of memory and creativity. So, in the context of Packard's model, it is natural to ask whether the mutation rate that governs first-order evolution could adapt appropriately by means of a second-order process of evolution. If the mutation rate can adapt in this way, then this model would yield a simple form of the evolution of evolvability and, thus, might illuminate one of life's fundamental prerequisites.

Previous work (Bedau & Bahm 1994) with fixed mutation rates in Packard's model revealed two robust effects. The first effect was that the mutation rate governs a phase transition between genetically ordered and genetically disordered systems. When the mutation rate is too far below the phase transition, the whole gene pool tends to remain frozen at a given strategy; when the mutation rate is significantly above the phase transition, the gene pool tends to be a continually changing plethora of randomly related strategies. The phase transition itself occurs at a characteristic mutation rate. The second effect was that evolution produces maximal population fitness when mutation rates are around values just below this transition. The upshot of these two effects is that evolutionary adaptation tends to be maximized when the gene pool is "at the edge of disorder."

In the light of our earlier suppositions about balancing the demands for memory and creativity, this work suggest that evolutionary memory and creativity are balanced at the edge of genetic disorder. To test this balance hypothesis, Packard's model was modified so that each agent has an additional gene encoding its personal mutation rate (Bedau & Seymour 1994). In this case, two kinds of mutation play a role when an agent reproduces: the child inherits its parents' sensory-motor genes, which mutate at a rate controlled by the parent's personal (genetically encoded) mutation rate; and the child inherits its parent's mutation rate gene, which mutates at a rate controlled by a population-wide meta-mutation rate. Thus, first-order (sensory-motor) and second-order (mutation rate) evolution happen simultaneously. So, if the balance hypothesis is right and mutation rates at the critical transition produce optimal conditions for sensory-motor evolution because they optimally balance memory and creativity, then we would expect second-order evolution to drive mutation rates into the critical transition. It turns out that this is exactly what happens.

Examination of many, many simulations confirms the pattern predicted by the balance hypothesis: Second-order evolution tends to drive mutation rates to the edge of disorder, increasing population fitness in the process. If natural selection is prevented from shaping the distribution of mutation rates in the population, the mutation rates wander aimlessly due to random genetic drift. But the mutation dynamics are quite different when natural selection operates. Although the population is initialized with quite high mutation rates well into the disordered side of the spectrum, as the population becomes more fit (i.e., more efficiently gathers resources) the mutation rates in the population drop into the ordered side of the mutation spectrum.

If the balance hypothesis is the correct explanation of this second-order evolution of mutation rates into the critical transition, then we should be able to change the mean mutation rate by dramatically changing where memory and creativity are balanced. In fact, the mutation rate does rise and fall along with the demands for evolutionary creativity. For example, when we randomize the values of all the sensory-motor genes in the entire population so that every agent immediately forgets all the genetically stored information learned by its genetic lineage over its entire evolutionary history, the population must restart its evolutionary learning job from scratch. It has no immediate need for memory (the gene pool contains no information of proven value); instead, the need for creativity is paramount. Under these conditions, we regularly observe a striking sequence of events: (a) the residual resource in the environment sharply rises, showing that the population has become much less fit; (b) immediately after the fitness drop the mean mutation rate dramatically rises as the mutation rate distribution shifts upward; (c) by the time that the mean mutation rate has risen to its highest point the population's fitness has substantially improved; (d) the fitness levels and mutation rates eventually return to their previous equilibrium levels.

These results show that the mutation rate distribution shifts up and down as the balance hypothesis would predict. A change in the context for evolution can increase the need for rapid exploration of a wide variety of sensory-motor strategies and thus dramatically shift the balance toward the need for creativity. Then, subsequent sensory-motor evolution can reshape the context for evolution in such a way that the balance shifts back toward the need for memory.

This all provides evidence for the following supple law of second-order evolution (at least in the modified Packard model):

> **Edge of Disorder**: Mutation rates evolve ceteris paribus to the edge of disorder.

The Edge of Disorder law is supple because it is the manifestation of a deeper regularity that concerns how some specific purpose is achieved in an indefinitely open-ended variety of contexts. The underlying regularity here is that, ceteris paribus, mutation rates evolve in such a way that evolutionary memory and creativity are optimally balanced for successful adaptability. This point at which the mutation rate balances evolutionary memory and creativity is typically at the edge of genetic disorder, but an indefinite variety of environmental contingencies can shift the point of balance. In other words, the Edge of Disorder law is vulnerable to exceptions that prove the rule. Not only are there rule-breaking exceptions in which evolutionary memory and creativity are not balanced (e.g., micro-level stochasticity, a break down in the micro-level mechanism implementing the process of second-order evolution, etc.); there are also rule-proving exceptions caused by memory and creativity being balanced somewhere other than the edge of disorder.

As with all ceteris paribus laws, the Edge of Disorder law is not precise and exceptionless. Furthermore, when ceteris is not paribus, it is generally unpredictable how contextual contingencies will relocate the balance between evolutionary memory and creativity. No evident fixed mechanism could guaranteed to find the appropriate mutation rate. Yet it remains a lawful regularity that mutation rates evolve to the point of balance, ceteris paribus. The explanation for this is that the process of evolution is continually changing the system's micro-level mechanism. The micro-level mechanism in this case is the population of agents with their genetically encoded sensory-motor strategies and their genetically encoded mutation rates. As the environmental context of this mechanism changes, the mechanism itself is subject to continual alteration by first- and second-order evolution. There is no algorithm for determining what adaptation will prevail in which context, which alteration in the mutation rate genes will emerge, but trial and error and natural selection can be counted on—ceteris paribus—to continually create appropriate mutation rates as novel situations unfold. This open-ended supple flexibility in the dynamics of the evolution of evolvability is the deep reason why the memory/creativity balance regularity resists any precise and exceptionless formulation.

Even though its suppleness means that the Edge of Disorder law and its underlying memory/creativity balance regularity cannot be explicitly stated precisely, accurately, and completely, the emergent model which produces them is itself an implicit description and explanation of the law and its underlying regularity—and one with all the virtues such accounts should have. We must remember to distinguish the model's micro- and macro-levels. At any given time the model's micro-level mechanism is completely precise and fully explicit. The Edge of Disorder law and the balance regularity, by contrast, are emergent macro-level phenomena which lack the micro-level mechanism's precision and explicitness. Nevertheless, since the model generates the law and the regularity, it implicitly describes and explains them, and it does so in a way that is fully consistent with all that we know about the natural world. Thus the emergent model is a feasible account of the supple macro-level law.

In addition, the model's description of the law can be made as precise, accurate and complete as desired. On the one hand, the law is a statistical pattern in the model's macro-level behavior, one with an indefinite complexity, so no finite description of the pattern can capture all of it. On the other hand, running the model again and again from different places in parameter space allows us to observe as much of the pattern's structure as we want. Furthermore, the operation of the model shows you exactly when ceteris is not paribus, so you can simply observe under what conditions the system does or does not evolve to the edge of genetic disorder. In this way you can fill out your understanding of both rule-breaking and rule-proving exceptions as much as desired.

The model is somewhat like a recursive grammar for an infinite language. A grammar is a compact, precise, complete and accurate description and explanation of the structure of a language, and generating more and more sentences wth the grammar provides a more and more accurate picture of the language. There is no guarantee one's proposed grammar will accurately capture the intented language. Similarly, it is an empirical question how well a given emergent model captures a specific macro-level phenomenon. But the model's feasibility enables us to test the account's accuracy and completeness using ordinary empirical methods, and we must continue revising the model if we observe false positives or false negatives in the model's macro-level patterns. We obviously should not accept a model until we are confident that it accurately and completely generates the desired macro-level pattern. So, when the empirical evidence supports a model, it ipso facto supports the model's accuracy and completeness. Finally, the emergent model provides a principled explanation of the Edge of Disorder law, exceptions and all. Since one and the same model generates both the law and its exceptions, the model is a concrete embodiment of what unifies the manifold instances of the law as well as its various exceptions. In addition, as the evidence reviewed above indicates, one and the same memory/creativity regularity produces both the law and its rule-proving exceptions, so the Edge of Disorder law is not arbitrary or ad hoc.

The emergent-models strategy for explaining supple laws also avoids the familiar controversies about ceteris paribus laws. First, I take it to be obvious that the law is neither trivially true, nor tautologous, nor analytic. The law is rather surprising, in fact, and could well have been false. Second, the law is clearly falsifiable. For example, it would be falsified if increasing the demands for memory (or creativity) did not typically cause mutation rates to evolve down (or up). Third, the law has explanatory power: it explains the typical second-order evolutionary changes in the mutation rate distribution in the population. Finally, although the law is non-algorithmic in the sense that it cannot be generated by any fixed algorithm, we have seen that it can be generated by a process involving the continual modification of micro-level mechanisms through the process of natural selection.

The Edge of Disorder law is just one example of a supple law described and explained by an emergent model. Many other characteristic features of living systems can be captured in a similar fashion as emergent phenomena in artificial life models; see, e.g., Farmer, Lapedes, Packard, & Wendroff (1986), Langton (1989b), Langton, Taylor, Farmer, & Rasmussen (1992), Varela & Bourgine (1992), and Brooks & Maes (1994). In every case, supple macro-level dynamics emerge from, and are explained by, an evolving micro-level mechanism consisting of a parallel, distributed network of communicating agents deciding how to behave in their local environment based on selective information from their local environment. This growing empirical evidence from artificial life continually reinforces the conclusion that emergent models can provide a good description and explanation of supple dynamics.

Toward Emergent Models of Supple Psychological Laws

The main conclusion of the previous section is that only artificial life's emergent models provide good descriptions and explanations of supple biological laws; all other evident alternative strategies lack some of the virtues we should seek in such accounts. The central claim I will defend in this concluding section is that only analogous emergent models provide good descriptions and explanations of supple laws in psychology. Although this claim is backed by evidence, it is just a conjecture. The macro-level behavior of emergent models is especially difficult to predict a priori, for by definition an emergent model's macro-level behavior can be settled only through the process of simulation. So prognosticating about what emergent models can or cannot do is risky. The acid test of such claims is to "put your model where your mouth is" (Bedau 1998). Although I have no such a models to offer today, the advantages and distinctive features of artificial-life-like emergent models can be illuminated by comparisons with recent work in cognitive science and neuroscience.

My central argument has two main premises: Emergent artificial life models are the only evident way to understand supple laws in biology, and they do an impressively good job. Furthermore, supple biological laws are strikingly analogous to supple psychological laws. I conclude that analogous emergent models promise to be the only possible way to understand supple laws in psychology. Let me be the first to admit two inherent weaknesses of this argument: it combines an argument from analogy with an argument about "the only straw afloat," and both are inconclusive. A significantly stronger argument must wait until we actually have a concrete emergent model to examine. In the meantime, though, we should note that the two premises do have significant support and thus do transfer some significant support to my conclusion.

What would an artificial-life-like emergent model of supple psychology look like? Its central move would be to shatter the seemingly indivisible Cartesian ego and construe it as an emergent macro-level effect of an adapting micro-level population of interacting proto-mental agents. It is unclear what the micro-level agents should be (although almost certainly not neurons) or how they should interact. It is also unclear by exactly what process they should adapt; perhaps Lamarkian selection should replace the natural selection in artificial life models, and presumably multiple levels of selection should interact simultaneously on different time scales. What is clear is that the familiar ceteris paribus laws of psychology should be recognizable as the emergent effect of an adapting population of interacting micro-level agents competing for influence in a context-dependent manner.

Artificial-life-like emergent models would have some similarity with certain existing models, such as those of Hofstadter and his students (Hofstadter 1985, Mitchell 1993, French, 1995), classifier systems (Holland 1986), and connectionist (neural network, parallel distributed processing) models (Rumelhart & McClelland 1986; Anderson & Rosenfeld 1988), so the important features of what I'm calling emergent models can be highlighted by comparing them with these other models. I'll focus on connectionist models since they are especially well-known.

Emergent models of mental phenomena and connectionist models have some striking similarities. First, both tend to produce fluid macro-level dynamics as the implicit emergent effect of micro-level architecture. In addition, both employ the architecture of a parallel population of autonomous agents following simple local rules. For one thing, the agents in an emergent model bear some analogy to the units in a connectionist net. Furthermore, the agents in many artificial life models are themselves controlled by internal connectionist nets (e.g., Todd & Miller 1991; Ackley & Littman 1992; Belew, McInerney, & Schraudolph 1992; Cliff, Harvey, & Husbands 1993; Parisi, Nolfi, & Cecconi 1992; Werner & Dyer 1992). In addition, for decades connectionism has explored recurrent architectures and unsupervised adaptive learning algorithms, both of which are echoed in a general manner in much artificial life modeling.

But there are important differences between typical artificial life models and many of the connectionist models that have attracted the most attention, such as feed-forward networks which learn by the back-propagation algorithm. First, the micro-level architecture of artificial life models is much more general, not necessarily involving multiple layers of nodes with weighted connections adjusted by learning algorithms. Second, emergent models employ forms of learning and adaptation that are more general than supervised learning algorithms like backpropagation. This frees artificial life models from certain common criticisms of connectionism, such as the unnaturalness of the distinction between training and application phases and the unnatural appeal to an omniscient teacher. Third, typical connectionist models passively receive prepackaged sensory information produced by a human designer. In addition, they typically produce output representations that have meaning only when properly interpreted by the human designer. The sort of emergent models characteristic of artificial life, by contrast, remove the human from the sensory-motor loop. A micro-level agent's sensory input comes directly from the environment in which the agent lives, the agent's output causes actions in that same environment, and those actions have an intrinsic meaning for the agent (e.g., its bearing on the agent's survival) in the context of its life. Through their actions, the agents play an active role in controlling their own sensory input and reconstructing the own environment (Bedau 1994, 1996b). Finally, the concern in the bulk of existing connectionist modeling is with equilibrium behavior that settles onto stable attractors. By contrast, partly because the micro-level entities are typically always reconstructing the environment to which they are adapting, the behavior of the emergent models I have in mind would be characterized by a continual, open-ended evolutionary dynamic that never settles onto an attractor in any interesting sense.

Neuroscientists sometimes claim that macro-level mental phenomena cannot be understood without seeing them as emerging from micro-level activity. Churchland and Sejnowski (1992), for example, argue that the brain's complexity forces us to study macro-level mental phenomena by means of manipulating micro-level brain activity. Their position has a superficial similarity to the emergent models perspective, but there is an important difference between the two. For Churchland and Sejnowski, manipulating the mind's underlying micro-level activity is merely a temporary practical expedient, a means for coming to grasp the mind's macro-level dynamics. Once the micro-level tool has illuminated the macro-level patterns, it has outlived its usefulness and can be abandoned. No permanent, intrinsic connection binds our understanding of micro- and macro-levels. By contrast, my thesis is that the mind's macro-level dynamics can be adequately described or explained only by making essential reference to the micro-level activity from which it emerges. The micro-level mechanism in the emergent model is a complete and compact description and explanation of the macro-level dynamics. Since these global patterns are supple, they inevitably have rule-proving exceptions. Thus, to get a precise and detailed description of the macro-level laws, there is no alternative to simulating the model. In this way, the micro-level model is ineliminably bound to our understanding of the emergent supple laws.

At this stage of development the emergent-model strategy for understanding supple psychological processes raises at least as many questions as it answers. Our final judgment of it must await the time when we have concrete models to explore. But we can conclude today that this strategy shows some striking promise. This is encouraging, for the suppleness of psychological processes is at once both enigmatic and essential to the intelligence of life and mind.

Acknowledgments

For valuable discussion, thanks to Terry Horgan, Karen Neander, Norman Packard, Kim Sterelny, and audiences at my 1987 ISHPSSB presentation in Seattle, at my Cognitive Science seminar at the University of California at Los Angeles, and at my Philosophy colloquia at the University of California at San Diego and the University of Newcastle.

References

Ackley, D., & Littman, M. 1992. Interactions between evolution and learning. In C. Langton, C. Taylor, D. Farmer, & S. Rasmussen (Eds.), Artificial life II. Reading, MA: Addison-Wesley.

Anderson, J. A., & Rosenfeld, E. (Eds.). 1988. Neurocomputing: foundations of research. Cambridge, MA: Bradford Books/MIT Press.

Bedau, M. A. 1990. Against mentalism in teleology. American Philosophical Quarterly, 27, 61–70.

Bedau, M. A. 1991. Can biological teleology be naturalized? The Journal of Philosophy, 88, 647-655.

Bedau, M. A. 1992a. Where's the good in teleology? Philosophy and Phenomenological Research, 52, 781–805.

Bedau, M. A. 1992b. Goal-directed systems and the good. The Monist, 75, 34–49.

Bedau, M. A. 1993. Naturalism and teleology. In S. J. Wagner & R. Warner (Eds.), Naturalism, a critical appraisal. Notre Dame: University of Notre Dame Press.

Bedau, M. A. 1994. The evolution of sensory-motor functionality. In P. Gaussier & J.-D. Nicoud (Eds.), From perception to action. Los Alamitos, CA: IEEE Computer Society Press.

Bedau, M. A. 1995. Three illustrations of artificial life's working hypothesis. In W. Banzhaf & F. Eeckman (Eds.), Evolution and biocomputation: computational models of evolution. Berlin: Springer.

Bedau, M. A. 1996a. The nature of life. In M. Boden (Ed.), The philosophy of artificial life. New York: Oxford University Press.

Bedau, M. A. 1996b. The extent to which organisms construct their environments. Adaptive Behavior, 4, 476-482.

Bedau, M. A. 1997. Weak emergence. In J. Tomberlin (Ed.), Philosophical perspectives: mind, causation, and world, Vol. 11. New York: Blackwell.

Bedau, M. A. 1998. Philosophical content and method in artificial life. In T. W. Bynam and J. H. Moor (Eds.), The digital phoenix: how computers are changing philosophy. New York: Blackwell.

Bedau, M. A. & Bahm, A. 1994. Bifurcation structure in diversity dynamics. In R. Brooks & P. Maes (Eds.), Artificial life IV. Cambridge, MA: Bradford Books/MIT Press.

Bedau, M. A., & Packard, N. 1992. Measurement of evolutionary activity, teleology, and life. In C. Langton, C. Taylor, D. Farmer, & S. Rasmussen (Eds.), Artificial life II. Reading, MA: Addison-Wesley.

Bedau, M. A., Ronneburg, F., & Zwick, M. 1992. Dynamics of diversity in a simple model of evolution. In R. Männer & B. Manderik (Eds.), Parallel problem solving from nature 2. Amsterdam: Elsevier.

Bedau, M. A. & Seymour, R. 1994. Adaptation of mutation rates in a simple model of evolution. In R. Stonier & X. H. Yu (Eds.), Complex systems: mechanisms of adaptation. Amsterdam: IOS Press.

Belew, R. K., McInerney, J., & Schraudolph, N. N. 1992. Evolving networks: Using the genetic algorithm with connectionist learning. In C. Langton, C. Taylor, D. Farmer, & S. Rasmussen (Eds.), Artificial life II. Reading, MA: Addison-Wesley.

Brooks, R. & Maes, P. (Eds.). 1994. Artificial life IV. Cambridge, MA: Bradford Books/MIT Press.

Cartwright, N. 1983. How the laws of physics lie. New York: Oxford University Press.

Cartwright, N. 1995. Ceteris paribus laws and socio-economic machines. The Monist, 78, 276–294.

Chalmers, D. J., French, R. M., & Hofstadter, D. R. 1992. High-level perception, representation, and analogy. Journal of Experimental and Theoretical Artificial Intelligence, 4, 185-211.

Churchland, P. S. & Sejnowski, T. J. 1992. The computational brain. Cambridge, MA: Bradford Books/MIT Press.

Cliff, D., Harvey, I., & Husbands, P. 1993. Explorations in evolutionary robotics. Adaptive Behavior, 2, 73-110.

Dennett, D. C. 1971. Intentional systems. The Journal of Philosophy, 68, 87–106.

Dennett, D. C. 1984. Cognitive wheels: the frame problem of AI. In C. Hookway (Ed.), Minds, machines, and evolution: philosophical studies. Cambridge: Cambridge University Press.

Dreyfus, H. 1992. What computers still cannot do (Rev. ed.). Cambridge, MA: MIT Press.

Farmer, J. D., Lapedes, A., Packard, N., & Wendroff, B. (Eds.). 1986. Evolution, games, and learning: models for adaptation for machines and nature. Amsterdam: North Holland.

Fodor, J. A. 1981. Special sciences. In J. A. Fodor (Ed.), Representations. Cambridge, MA: Bradford Books/MIT Press.

Fodor, J. A. 1991. You can fool some of the people all of the time, everything else being equal: Hedged laws and psychological explanations. Mind, 100, 19–34.

French, R. M. 1995. The subtlety of sameness: a theory and computer model of analogy-making. Cambridge, MA: Bradford Books/MIT Press.

Hempel, C. 1965. Aspects of scientific explanation. In C. Hempel (Ed.), Aspects of scientific explanation and other essays in the philosophy of science. New York: Free Press.

Hofstadter, D. R. 1985. Waking up from the boolean dream, or, subcognition as computation. In D. R. Hofstadter (Ed.), Metamagical themas: questing for the essence of mind and pattern. New York: Basic Books.

Holland, J. H. 1986. Escaping brittleness: The possibilities of general-purpose learning algorithms applied to parallel rule-based systems. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), Machine learning II. Los Altos, CA: Morgan Kaufmann.

Horgan, T. & Tienson, J. 1989. Representation without rules. Philosophical Topics, 17, 147-174.

Horgan, T. & Tienson, J. 1990. Soft laws. Midwest Studies in Philosophy, 15, 256-279.

Horgan, T. & Tienson, J. 1996. Connectionism and the philosophy of psychology. Cambridge, MA: Bradford Books/MIT Press.

Langton, C. 1989a. Artificial life. In C. Langton (Ed.), Artificial life. Reading, MA: Addison-Wesley.

Langton, C. (Ed.). 1989b. Artificial life. Reading, MA: Addison-Wesley.

Langton, C., Taylor, C. E., Farmer, J. D., Rasmussen, S. (Eds.). 1992. Artificial Life II. Reading, MA: Addison-Wesley.

Mitchell, M. 1993. Analogy-making as perception. Cambridge, MA: Bradford Books/MIT Press.

Packard, N. 1989. Intrinsic adaptation in a simple model of evolution. In C. Langton, (Ed.), Artificial Life. Reading, MA: Addison-Wesley

Parisi, D., Nolfi, N., & Cecconi, F. 1992. Learning, behavior, and evolution. In F. Varela & P. Bourgine (Eds.), Towards a practice of autonomous systems. Cambridge, MA: Bradford Books/MIT Press.

Putnam, H. 1973. Reductionism and the nature of psychology. Cognition, 2, 131– 146.

Putnam, H. 1975. The nature of mental states. In H. Putnam (Ed.), Mind, language, and reality. Cambridge, England: Cambridge University Press.

Putnam, H. 1991. Representation and reality. Cambridge, MA: Bradford Books/MIT Press.

Rumelhart, D. E., & McClelland, J. L. 1986. Parallel distributed processing: explorations in the microstructure of cognition, 2 Vols. Cambridge, MA: Bradford Books/MIT Press.

Schiffer, S. 1991. Ceteris paribus laws. Mind, 100, 1–17.

Todd, P. M., & Miller, G. F. 1991. Exploring adaptive agency II: Simulating the evolution of associative learning. In J. -A. Meyer & S. W. Wilson (Eds.), From animals to animats, proceedings of the first international conference on the simulation of adaptive behavior. Cambridge, MA: Bradford Books/MIT Press.

Varela, F., & Bourgine, P. (Eds.). 1992. Towards a practice of autonomous systems. Cambridge, MA: Bradford Books/MIT Press.

Werner, G. M., & Dyer, M. G. 1992. Evolution of communication in artificial organisms. In C. Langton, C. Taylor, D. Farmer, & S. Rasmussen (Eds.), Artificial life II. Reading, MA: Addison-Wesley.