Three paradigms of evolution

Leigh M. Van Valen
Department of Ecology and Evolution
University of Chicago
915 E. 57 St.
Chicago, Ill. 60637

Received 7 April 1989; 13 June 1989

ABSTRACT: The standard paradigm of evolution has marginal difficulties which motivate the development of two alternatives or supplements, one based on information and one on energy. They all agree in most applications, but the informational paradigm has perhaps insuperable practical difficulties. The energetic paradigm is conceptually simple and can in principle unify the several otherwise disparate areas of population biology at a fundamental level. It also gives an easy entry into the study of problems whose very existence is not even apparent in the standard paradigm, such as the evolution of communities and other ecological groups.

We have all probably heard Dobzhansky's aphorism that evolution is change in gene frequencies. I don't know how many of us still regard this as a sufficient description of the fundamental nature of evolution. Some years ago (Van Valen, 1973b, 1974) I proposed a competing, or complementary, aphorism, that evolution is the control of development by ecology. That's not what I will be discussing, though. Because critiques of current evolutionary theory are the refuge of incompetents, ignoramuses, and the malicious, plus a smattering of others who seem to want too much credit for too little originality, I must ask readers to avoid stereotyping this paper and to consider the arguments on their merits, if any.

What I want to do is to indicate why the currently dominant paradigm in evolutionary biology really is flawed in a serious way, mostly at its base, and to sketch two alternatives to it. One alternative is perhaps more a modification than a replacement (this is a matter of terminology which I will ignore), but the other actually does seem to be a different paradigm in Kuhn's original usage. Why this is the case will be apparent later. Being different doesn't mean that the alternative is right; it rather means that neither paradigm can be judged by the criteria of the other. Fortunately all three paradigms agree in most applications. Thus I am not proposing that the Synthesis be scrapped, rather that it be extended a bit and that its conceptual underpinning be modified.

Prologue

There is supposed to be a subject called population biology. I deny its existence. There is ecology, and there is genetics. (There are other elements too: populations exist in time, and in space, and they have components, and they are components of larger groups, and they have phenotypes, and all these of course are also subjects of study.) But so far there is only a surficial interfingerling: things like genotype-environment interaction, or effect of density or age structure on natural selection, or possible genetic effects in population regulation, or the environmental frame for natural selection. All of them leave both ecology and genetics basically intact and disjoint from each other, as even the organization of any book on "population biology" makes evident.

* * *

1Contribution 90. Lothlorien Laboratory of Evolutionary Biology
Evolutionary Theory 9: 1-17 (July, 1989)
The editors thank two referees for help in evaluating this paper.
© 1989, Department of Ecology and Evolution, The University of Chicago
Conversely, the interaction of genetics with biochemistry has transformed both of them, because the nature of their interaction is causally fundamental.

Fitness is the central concept of evolutionary biology, and I will try to sketch in the final alternative how it can contribute to a similarly deep interaction. In order to make sense of what follows, one has to think a bit different from usual. That's what a conceptual revision is.

As an example of thinking differently, one which I won't pursue yet but which may have some promise, consider jointly all the genes in a community (or even in the world, but that has other problems), and ignore complications like promoters, introns, transposons, duplicate genes, conversion, mutation, recombination, and what have you. I don't mean one pair of alleles per locus per individual (whatever an individual may be), although that may be an easy entry into this problem, but all the genes: more cells, more genes. Thus consider all the genes in a community, all the allele copies.

Some of these genes, or their replicas, will be there at some later time; others won't; and frequencies will have changed. This gives a view of absolute fitnesses at the genic level, and species boundaries are irrelevant here even though partly relevant in causing the differences. They are not even causally as predominant as in the received view, though, because whether competition, say, is with members of one's own species or with members of another species, it is still the same process from this perspective, and the same gene can also occur in more than one species.

Some of these genes will have remained because of persistence of single long-lived individuals, some will have increased because an individual has grown, and some will have increased by reproduction or dispersal. Some are in salamanders and some are in birds and some are in grasses and some are dispersed in bacteria. (There is more energy used per gene in birds than in salamanders, and this is mechanistically relevant. The shorter time spent by a unit of energy in birds [Ulanowicz and Hannon, 1987] is another way of looking at the same phenomenon.) Many loci will be taxonomically restricted; their alleles have a frequency of 0 elsewhere.

One can nevertheless ask what the causal controls on community allele frequencies in this sense may be. I don't think this is a trivial question, or one with an obvious answer -- but one has to shift one's thinking even to make sense of it.

Then again, look at the rock called granite. It is composed mostly of grains of feldspars and quartz, with some mica and other minerals inserted among them. When granite weathers, the feldspars and micas become clays but nothing much happens to the quartz grains. They are more resistant and get transported down streams or along shores. Thus most beaches are the result of differentially eroded granite.

This is an example of natural selection in the nonliving world. Quartz grains survive longer than feldspar grains, and there is a progressive increase in the average resistance to weathering, of the set of grains that have still survived. This action of natural selection is even creative, as we see by the formation of a beach.

The lack of reproduction imposes constraints on the flexibility of evolution here, but one shouldn't confuse that with the selection itself. We do have here a common sort of evolution by natural selection, and there are many other nonliving examples. (Try to think of some. Being irreversible, they provide at least local arrows for the direction of time.)

So we should try to relax the constraints that channel our thoughts toward the ways we have thought before.

**Fitness**

Yes, fitness is the central concept of evolutionary biology, but it is an elusive concept. Almost everyone who looks at it seriously comes out in a
different place. There are literally dozens of genuinely different definitions, which I won’t review here. At least two people have called fitness indefinable, a biological primitive. (A primitive is an undefined initial term in logic.) I don’t think that helps. Stearns (1976) once defined it as “something everyone understands but no one can define precisely.” Or is it that we can’t define it because we don’t fully understand it?

It does make a difference what concept one uses. Compare, e.g., maximizing I (of one kind or another) with minimizing the probability of extinction, or of loss of an allele. And a general theory shouldn’t apply just to organisms which are like us in relevant ways. Extreme cases test the genfalse. More seriously, natural selection on individuals doesn’t maximize population fitness in general, natural selection over one generation doesn’t maximize fitness over ten years in general, and so on. Maximization can apply only at the precise time scale and the precise level of the selection being considered. Anything else is a byproduct, perhaps one which usually happens but not one to invoke in conceptual analysis.

So. We now have an empirical question rather than one of definitiofalse. More serious selection on individuals doesn’t maximize population fitness in general, natural selection over one generation doesn’t maximize fitness over ten years in general, and so on. Maximization can apply only at the precise time scale and the precise level of the selection being considered. Anything else is a byproduct, perhaps one which usually happens but not one to invoke in conceptual analysis.

So. We now have an empirical question rather than one of definition; what does natural selection actually maximize? (Of course, we then have the question of exactly what natural selection is, but that may be more tractable. And any minimum can be made a maximum by appropriate inversion.) Natural selection is a single process, if wearing different clothes on different occasions, and for any particular definable situation it produces a definable outcome. What is it, then, that these outcomes have in common for organisms?

The standard definition of fitness in population genetics, where there is some degree of uniformity if hardly unanimity, is something like the relative number of individual offspring in the next generation of the population. All seven of the meaningful words here have serious difficulties, some of them recognized and accounted for in careful work but others not. Fitness is the relative number of individual offspring in the next generation of the population.

First, one word not included, which should be: expected. Frequencies and amounts of evolutionary units can change for reasons other than their fitness, as everyone knows when they happen to think of it. Fitness is therefore a propensity (relative to some environment or ordered set of environments), as argued at length by Mills and Beatty (1979) and as commonly realized by biologists (e.g., Dobzhansky, 1951, p. 78). Vrba and Gould (1986, partly misinterpreted by Hoffman, 1987) exemplified the same distinction for processes, using the terms sorting and selection.

Relative: Fitness is obviously relative to the total environment; in the definition it is also relative to the fitnesses of other units (alleles, species, or whatever), and that is where the problem arises. I don’t mean that interaction of selective units causes a problem; they are merely part of each other’s environment. The problem is that fitness is usually regarded as a comparative attribute, not one integral to the unit itself. A behavioral phenotype of starlings might kill the eggs of its neighbors and so increase in relative frequency. But it can simultaneously decrease in absolute frequency and even cause the extinction of the whole population because it doesn’t guard its own nest while it is watching for its neighbors to leave. Of course, some people may think that this would be a good thing to happen to their local population of starlings. Real natural selection, though, occurs in communities, and in the changed behavioral environment both phenotypes are at a disadvantage in this broader context, although the killer is less disadvantaged than the tolerant. Relative fitnesses usually work, but absolute fitnesses always do if one uses the proper context.
Generation: Natural selection occurs on generation length, as is well known, for species which are iteroparous or which otherwise have overlapping generations. (It occurs in other cases too, but a directional response is easier with generation overlap.) Even for univoltine insects, though, the unit of a generation can be inappropriate. Individual natural selection is the same process within species and among individuals of different species, and a bivoltine competitor species has exactly the same potential selective advantage, acting in exactly the same way ecologically, as a bivoltine morph in the first species. Real natural selection occurs in absolute time. Sometimes we can ignore this but sometimes we can’t.

Next: Thoday (1953) once defined fitness as the expectation of being present in $10^6$ years. This interval is of course arbitrary; it merely reflects the occurrence of selection on many time scales. As another example, the evolution of a thicker shell by a snail for protection against a crab may select for stronger chelae in the crab, with the result that the snail is worse off than initially because it is now eaten as readily as it had been but has the additional cost of a heavier shell. Thus, in the absence of other predators long-term selection may actually be for a thinner shell or none at all. (An explanation of this type may be relevant to the remarkably low amount of predator-prey escalation which Vermeij and Dudley (1985; Vermeij, 1987) found for most freshwater mollusks.)

Rain-forest diversity can be nonequilibrium at local and short scales, equilibrium at regional and moderate-term scales, perhaps then having a major extinction (even one caused naturally, since the late Pliocene or so), and perhaps equilibrium again on a scale of $10^6-10^7$ years. In general, for such questions, ask (1) is there any spatiotemporal scale at which (feedback-regulated) balance occurs, (2) what happens at other scales, and (3) what are the causal connections? The very existence of an equilibrium is a severe causal constraint at some nonequilibrium scales.

Population: I have indicated how natural selection is the same overall process among individuals of the same and different species. This is quite a general conclusion: if A is fitter than B, the ratio of A to B will increase whether or not they are in the same species. Sexual reproduction and the nature of the genetic control affect the rate and mechanisms, but organisms live in communities and that is where selection occurs. Species boundaries are important, and so are the characters of species and higher taxa; I am not advocating the bookkeeping view that group selection at such levels collapses into individual selection. However, at the level of individual selection within communities the effect is one of a high correlation among characters.

There are further difficulties with use of populations. It is now reasonably well known that selection can occur among different patches within a population or a community (Wilson, 1980), but it is also true that some populations are continuous over wide areas. There is often no natural spatial scale. This can apply also to colonizers and other fugitive species, where the local population is more or less always an ephemeral part of a fuzzily bounded metapopulation, and in other cases too.

Individual: Sure and natural selection occurs on individuals. Some individuals have a propensity in the current total environment to die sooner than others, to have more offspring, etc. At any level of selection the phenotype is what is selected. Every individual is unique, though. The phenotype, and even the genotype in sexual reproduction (usually a bit in asexual reproduction too, if one remembers mutation and genome churning), of each offspring differs from that of each parent. An individual perpetuates not itself but components of itself. This isn’t playing with words. It forces us to deal explicitly with components of the phenotype and genotype, such as genetic characters and alleles or loci. That is even how we recognize different propensities for specific total phenotypes or genotypes, as distinct from looking at what actually happens, which is an integration of fitness and chance. Such an epistemic problem (recognition) doesn’t affect reality, but there is a similar need for decomposition on the ontic or
Three paradigms of evolution

real-world level when we consider response to selection, even though phenotypes do approximately reappear with different genotypes. For fitness itself, however, all this doesn't really matter at the ontic level. An individual in some total environment does have a specific fitness at each time scale. That its offspring will differ from itself is no different from the fact that half the offspring of a heterozygote, in any mating, will be homozygous. We still assign a fitness to the heterozygote itself, and this has proved useful. I do not know any corresponding usefulness for individual fitnesses as such, though. (This is a matter of degree; I am not implying that they are useless. I have in fact estimated some, for perhaps the first time, and used them, in unpublished work.) A more serious difficulty with individuals will appear in the discussion of numbers, below.

Offspring: It is just as good, and maybe better, for a massive coral or a tree to stay alive, occupying the same good site, as it is for it to reproduce into an uncertain world. Of course this depends on the time scale one is considering, and it emphasizes in another way how important it is to keep clearly in mind the continuum of time scales of selection. It is just as hard for a reef coral to senesce as it is for our own lineages of germ cells to senesce; pleiotropy has no place to act, and moreover the reproductive value of such a coral remains high. Persistence is an important component of fitness and is intimately related to the spatiotemporal heterogeneity of the total environment.

Finally, numbers: Here we come to a greater divergence from the received viewpoint, from the implicit assumptions that one individual is equivalent to another and that there are natural units identifiable as individuals in the first place.

For individual selection in Drosophila or mammals, it is a reasonable approximation. It is not reasonable for group selection at any level, unless things average out appropriately, as they sometimes do. It is not reasonable in most cases of selection among species and even some genotypes (Connelly, 1988). It is not reasonable, again, for reef corals or for trees or for many other organisms. These are all what I have called somatogens (Van Valen, 1987), organisms which do not have a single germ line from which all reproduction occurs sexually. There are two kinds of somatogens, clonates (which form clones) and polygernates (which have more than one germ line, like most plants and colonoid animals). How they undermine a reliance on numbers is the subject of the next section.

Expansion

An important means of competition for many somatogens is individual expansion. (Contraction is negative expansion.) Size is important in natural selection, and not just for its consequences of greater reproduction, lower predation, and such. The expansion of a sponge at the expense of an adjacent bryozoan is natural selection, but it occurs with no death and no reproduction. If a branch with completely green leaves grows more than one with variegated leaves, that too is natural selection. A plant or a graptolite may grow out, or it may put the same energy into sending forth a new asexual ramet or into sexual reproduction. In each case its goal, and it is a nonconscious goal (Van Valen, 1984b), is to acquire more of what it needs and to make more of itself. Yes, Bertrand Russell (1927, p. 27) said it too: "Every living thing is a sort of imperialist, seeking to transform as much as possible of its environment into itself and its seed." Our theories can make us forget the obvious.

Such non-Weismannian evolution applies to groups too. Somatogens seem to resemble groups in the relative importance of their fitness components. A major mechanism of group selection is expansion of particular groups. That is, e.g., why blood group A is decreasing in our own species at the astonishing rate of about 1 percent (5 percent of its actual frequency) in 35 years (Thompson, 1972; Van Valen, 1976): this is a correlated effect of selection on cultural characters of groups. No extinction or founding of new groups is of any importance here. Similarly,
Wright's intergroup selection (1932) operates primarily by differential expansion. A population can, at least in principle, produce more populations just by subdividing itself. There are then more populations but even jointly they probably play a smaller role in the community. Our theories would nevertheless say that the lineage now has greater fitness. Vrba (1980) in fact made about this interpretation for two clades of antelopes, one of which has many species branches and the other only one. (There is a seriously misleading attribute of her diagrams, however. Like most cladists she refuses to recognize the ancestral position of any species to any other. This results in falsely early times of apparent divergence of lineages and too many apparent cases of terminal extinction. Other people have actually used diagrams like these for calculations, which are thereby strongly biased.) Yet today the latter clade, the impala, seems to have about as many individuals as the former; as Vrba herself (1983, 1984) later noted, it doesn't seem less fit. Yet again, an individual impala is less than half the size of an average member of the other clade; the same number of individuals would use rather less energy. Treasure your exceptions: like an old tree, deeper penetration may reveal less substance.

Another example: A group A can gradually overwhelm group B by an increase in the size of its component parts, while group B increases in numbers of parts until near the end. This is partly what happened in the 30-million-year process of the extinction of the multituberculates, an ancient mammalian order, by their placental competitors: the maximum diversity of multituberculates occurred several million years after the effective beginning of the competition (Van Valen and Sloan, 1966; Krause, 1986). The use of numbers of species gives a non-monotone curve for what is really a steady process.

Is a grass which is colonizing a new field by ramets automatically much less fit than one which is colonizing a field by seeds? In the first case, one individual, in the sense of a genet (Harper, 1977), is expanding, even though it may eventually lose connections between clumps, while in the second there is lots of reproduction and new individual genets. Or compare the fitness of the parents of a tree and of 10^5 or 10^6 herbs covering the same area: are the parental trees 10^2 or 10^3 times as fit?

It is also now realized that for many plants, corals, and other somatogens, body size is more important than age. And further, in population regulation the standard theory counts individuals. We know, however, that the several different criteria of individuality, which normally coincide when applied to the animals most familiar to us, do not coincide for most somatogens (Van Valen, 1978, 1987). What, exactly, is an individual of a reef coral or an aphid or a slime mold or a tillering grass? But it can hardly be the case that natural processes depend on how we define terms, and in addition a large "individual" is not equivalent to a small one, nor an active one to one in dormancy. In this theory also, it is essential to eliminate counts of individuals from its basic formulation. In this case energy flow provides the only universal or natural replacement; whether that is also the case for fitness remains to be determined.

Somatogens themselves bridge levels of selection, because the boundary between an individual and a group becomes fuzzy as our several criteria no longer coincide with each other. For some somatogens, notably obligate clonates and obligate mutualists, the boundary between a deme and a species becomes fuzzy. How, exactly, does individual selection differ from group selection for a clonate? Or from somatic selection? Or deme selection from species selection? Clonates interact with nonclonates too; these questions are primarily of processes rather than of terminology. We must therefore recognize that levels of selection intergrade with each other at short time scales as well as at long ones. Even if there were no other reason to do so (and there are, both up and down), this conclusion forces us to look skeptically at the usual view of individual selection as the overwhelmingly most powerful level.

As Westoby (1983) put it, "The first imperative of success for a piece of
Three paradigms of evolution

Genetic information is that tissue containing copies of it be made at least as fast as tissue containing it is lost. Whether a net profit of growth is accumulated within the individuals or as multiplication of individuals all the same size is a second-order question."

Expansion is an important component of fitness in the real world, and we ignore it at our peril (or at our theories' peril). In fact the other components of fitness can easily be subsumed under it, as the rest often are under reproduction. Thus death is complete contraction and reproduction is a form of growth, if often combined with that of another parent so that only half can count for each (as in careful work from the standard approach). Rather than saying that natural selection is expected differential reproduction, we should say that it is expected differential expansion.

Information

There are two complementary ways to look at life, related to its two most fundamental aspects. These aspects are energy and information. Energy is straightforward, but information is not. Even the concept of biological information, above the level of single molecules like DNA per se, is another slippery surface with a variety of traps (cf. Van Valen, 1986). I don't have an adequate answer to this problem, but fortunately one isn't necessary here because the information to be considered is that of DNA itself. This has its own difficulties, although they are less severe because they are less fundamental.

Information and energy each provide the basis for a paradigm of evolutionary theory. Neither paradigm is quite like what we have thought of, and I haven't yet found a way to distinguish empirically between them. Both resolve the empirical and conceptual difficulties which we have noticed in existing theory.

In the informational paradigm, which is an extension of that of Williams (1966), Dawkins (1976), and Sibly (1989, who repeated a classic argument of Birch [1960] without realizing it), DNA is all-important. (Yes, substitute RNA with respect to some viruses, and even protein with respect to prions if the latter exist.) I don't mean just genes, whatever their boundaries may be, but the DNA as such: each nucleotide pair, each adjacent two nucleotide pairs, each adjacent three pairs, and so on in all possible overlapping contiguous sets. Methylation or other modification may or may not be relevant, depending empirically on how long it lasts and how it gets removed. Interchromosomal effects are relevant but are more awkward to incorporate, as are organelles, maternal transcripts, and autonomous inheritance of the cell cortex. Each nucleotide pair and larger set is relevant only to a particular location in the genome except when within-genome processes such as conversion and transposition are relevant. Location change by translocation, etc., is relevant only for those sets spanning the breakpoints; those sets then vanish.

The realized fitness of any one of these nucleotide sets, then, is its total absolute frequency in the universe being considered. (Its actual fitness, being a propensity, can only be approximated.) Recall the earlier approach: the more cells or nuclei, the more DNA; the greater ploidy, the more DNA. Expansion thereby has a natural place in this paradigm. So do selfish deletions, Ghiselin's bête noire (1987). Mutation can be regarded as extinguishing an example of one set of nucleotide sets and creating an example of another. Alternatively, the sets can be regarded as fuzzy and the transition can be gradual, as may be more appropriate for pseudogenes and other situations. The real diversity of levels and time scales of evolution, and of natural selection, can be incorporated directly, just as in the other paradigms. The interactions among different DNA sets are important and in this way they can be treated in a natural manner. The emphasis on DNA here is meant to be causal, not a bookkeeping device, and therefore to incorporate all relevant interactions.

In order to incorporate realistically the DNA around introns and transposons,
it seems necessary to consider sets of nucleotides separated by others which are excluded from the set. Is there an objective way to determine just which groups of nucleotides should be used in this manner? For instance, what happens when a transposon gets incorporated permanently into a site? Similarly, long-range interactions, not to mention interchromosomal ones, need incorporation, and they may be evolutionarily very transitory. Trifonov (1989) has distinguished two spatially more inclusive levels of "coding" (for nucleosome placement and interactive protein recognition) in DNA sequences, above the level of codon triplets. A third, for putative looped-DNA interactions such as long-distance promoters, also probably exists.

Such considerations suggest that we define our sets by functional criteria. This is, of course, what is usually done, when we deal with genes and the like. A more detailed consideration of genes will indicate the sorts of difficulties which arise when functional criteria of any sort are used. Consider introns. Should we ignore them, treat them separately, or just consider them integral parts of the gene? One can argue reasonably for each alternative. Introns usually evolve relatively rapidly, suggesting a nonspecific function if any, but some are now known to function even beyond catalyzing their own splicing. Introns can come and go as units in evolution, as well as change in sequence and length. Or consider a 3n-long transposition into an exon. The reading frame is the same but the protein has n more amino acids. A deletion of one nucleotide, on the other hand, changes all amino acids downstream as well as the protein's length. The structurally small change is likely to be functionally more severe than the structurally larger change; function isn't directly related to structure. Because of the sorts of higher-level interactions mentioned above, the ends of a gene don't correspond to sharp breaks in function.

For these and related reasons it is not possible to specify sharply bounded functioning units in the genome. Such idealized units can be used as approximations, and the value of such use is not in question. For an exact theory, though, they won't do, and we are left with the strings of nucleotides themselves. Dawkins (1982) argued against selfish nucleotides, but he didn't appreciate the problems arising from using genes. It may be that different sets of nucleotides should be weighted differently, depending on their function, but this opens an entire new level of complexity. How does one evaluate importance of function, exactly? For what? Over what time interval(s)? And so on. These questions may have answers, but the answers will not be simple.

Genes, however construed, are not the only functioning units in the genome, and some units (as completely selfish DNA) quite probably exist only for their own benefit. They can be created and eliminated as functioning units by mutation even more easily than genes (via pseudogenes or simple loss of function) can, because they are shorter.

The selective influences on DNA change include, in an important way, interactions which aren't visible (except as a bookkeeping artifact) at the haploid level, especially when generations overlap (Nagylaki and Crow, 1974). The haploid level is where gene frequencies and other nucleotide-set frequencies reside. Various people (e.g., Gould, 1977; Wagner, 1988) have taken this ignoring of interactions as a serious or even devastating argument against the selfish-gene approach. While I agree with the argument, the conclusion may be too narrowly focused. If we expand the selfish-gene approach to incorporate such interactions (as is perhaps more obvious in the informational paradigm), which is clearly possible in principle if not in practice, the mechanistic argument evaporates and we are again left with not-so-merely practical difficulties.

It is probably possible in principle to treat information as the basis of a paradigm of evolution. I don't know how to do it, though, in an adequately detailed way, and existing approaches are approximations which gloss over the real difficulties. All the immense number of nucleotide sets have to be considered jointly, in some manner which itself is not adequately specifiable yet. Unlike the
conceptual simplicity of energy, the nucleotide sets give a remarkably complex mess. In this general paradigm natural selection is indeed expected differential reproduction, but at the level of sets of nucleotides.

Energy

In any real case we can see what the course of events will be if we have adequate information. There is always a single outcome in real natural selection, except to the extent that stochasticity and chaos intervene, and the difficulty in such cases is with respect to knowledge rather than to causality.

What is the universal currency that is used? What is maximized? Consider the expansion component of natural selection by itself. Expansion of what, exactly? What is causally involved?

Biomass has been suggested, and it is, more or less, what is relevant for the food of the next trophic level up. But mass is not for the most part causal within a population, or in competition. More mass per se is not important except for situations like the ramming of rams.

We know that energy drives the physiological dynamics of life just as it drives inorganic processes. (Technically, both enthalpy and entropy are involved.) I conclude that free energy, normally in the form of reduced carbon or its redox derivatives, is in some way the universal currency. It provides a uniformly available measure which, moreover, acts causally in the relevant biological interactions.

Various objections need consideration. I do not mean that the availability of free energy regulates populations. Sometimes it does, and probably it usually does for the soil biota. Other sorts of constraints exist, though, which organisms can't overcome in the available time (say too little phosphorus, or water, or a high predation risk) and which therefore often regulate the populations subjected to them. Nevertheless, under such locally fixed constraints organisms do seem to maximize their energy control. This is not, e.g., the contrast of (immediate) energy maximization with time minimization; both are involved in a more basic maximization. We can specify the aspect of the environment which each organism tries to incorporate into itself and its seed. Lotka (1922) had an idea close to this, about the time of the quotation from Russell. Surrogate resources, like phosphorus or water above, are resources whose acquisition is selected for as a way of getting more energy.

As Egbert Leigh has commented to me, if life evolved elsewhere it would not necessarily need any of the specific chemical materials our life does, but it would need free energy.

Maintenance energy, including reproductive energy not put into offspring, is not maximized; in fact it is apparently minimized under the constraints applicable in any particular case. Nevertheless, this sometimes results in apparently wasteful maintenance, as with endotherms. But if they get more expansive energy by wasting more, that is what is selected for. Efficiency per se is not maximized.

What does seem to be maximized is the rate of acquisition of expansive energy, the energy used for growth and successful reproduction. This is the surplus energy above maintenance. Several years ago there was a graffito in the men's room in the Zoology Building at Chicago: "Fatness is fitness -- Van Valen." This is actually an interesting caricature. Over a very short interval fatness is indeed fitness, but we are usually interested in intervals longer than that, where excess fat or other stored energy may sometimes interfere with survival or reproduction or even with further expansion. It is quite essential for us to consider the time scales of selection carefully and explicitly. Similarly, a cancer is selected within the individual on a short time scale, with different cell sublineages being selected successively on the basis of different properties as the cancer progresses. As far as we know, such cell clones always die out (under natural conditions) on longer time scales. Or a grass which grows more and then is cropped down to the same
level as others is no better off on the usual time scale.

The time scale and level of selection which we are considering determine whether growth or reproduction is the more important, as the above classes of cases exemplify. Only reproduction, however, usually leads to new adaptations. In addition, in many somatogens there is no sharp distinction between growth and reproduction; think of a tillering grass or a budding bryozoan or a clone of cladocerans. It is therefore dangerous to use a basic theoretical framework which depends on such a sharp distinction and which must create one arbitrarily even when it doesn’t exist.

Likewise, populations are often at densities above that which maximizes their expansive energy, but this is again a conflict between group fitness and individual fitness. Each individual maximizes its own fitness, and group selection can’t in this case counteract such individual selection. Maximization is specific to each level and time scale of selection. That some are stronger than others leads to a minor "tragedy of the commons" in this case and to diverse but more or less obvious conflicts in other cases, as with meliotic drive or selective extinctions.

The Arizona collective (Bernstein et al., 1983) has characterized biological adaptations as being resolvable into three aspects: capacity to increase in number, capacity to avoid decay, and capacity to acquire and process resources. Although I don’t think that this taxonomy is quite adequate, all three categories are easily incorporated into the maximization of expansive energy on appropriate time scales.

It is gratifying that known forms of natural selection all fall naturally out of this view when appropriate restrictions are made on it. The details are perfectly straightforward, even trivial in most cases.

Various people (e.g., Williams, 1966, 1986; Nunney, 1985) have regarded selection, particularly group selection, as a way to explain biological adaptations rather than as a process to follow wherever it may lead. So if process A can explain the adaptation, we don’t need to consider process B. This is a particular case of the fallacy of null hypotheses (e.g., Van Valen, 1985a), where one hypothesis is given an initially privileged position. Alternatively, by treating possible causes together we are in the framework of estimation rather than that of hypothesis-testing (not that the latter is always unimportant) and we have an initial predilection to ask, e.g., how much each process contributes, if anything.

Comparison of paradigms

That the standard paradigm is poorly founded is, I take it, made at least plausible by the arguments I have given. Both the informational and the energetic paradigms escape from these problems, and neither seems to have comparable difficulties of its own. It is odd, although perhaps reassuring, that all three paradigms agree with each other in applications. This situation resembles that of mathematics, the content of which is not thrown into doubt by serious lack of agreement on its foundations. (Some views of mathematical foundations, especially but not only constructivism, prohibit aspects of mathematics otherwise thought valuable. Such arguments can be only heuristic, though, because the contested domains can be regarded as structures built on weak or even false conjectures.)

In the energetic paradigm, the fitness of a unit such as a nucleotide pair or an allele is its expected rate of change in abundance, suitably scaled. This is the same as that in realistic versions of the other two paradigms, because such units control energy only by being part of larger units. I discuss scaling more generally below, but here I note that it is arbitrary (and has no evolutionary or dynamical significance) whether, say, an allele of a long gene is weighted more than an allele of a short gene.

The energetic and informational paradigms usually coincide in their applications, exactly or approximately, but they don’t always do so. Cell size and so on provide classes of cases, at one level of evolution or another, where they
Three paradigms of evolution

Differ on some time scale. Cells do differ in size. Eggs are big cells, even those without yolk, and cleavage progressively reduces the size of embryonic cells until feeding starts. Over this interval informational copies greatly increase but energy control decreases per cell and probably often for the entire embryo. Siphonous green algae have the largest cells known, up to eight meters long. Most are syncytial, but Acetabularia is a well-known example with a single nucleus. A short-term energetic advantage, with no corresponding informational advantage, is for such a cell to enlarge. A larger cell controls more energy but need have no more information. The size-determining program of the cell has, however, like that of an embryo, been selected over longer time intervals, at which the two paradigms coincide in their optima.

Despite the existence of divergent predictions like this, I haven't found a way to distinguish between the two paradigms empirically. Field observations are of results of selection predominantly over time scales where the two coincide, and experiments must use one paradigm or both, rather than distinguishing their effects in nature. Each paradigm has its own version of natural selection, and one can't criticize the results of one by appealing to the process of the other. In this sense fitness, or natural selection, is indeed a primitive (an undefined initial term in logic), although this is an empirical judgment rather than an analytical one and I hope it can be falsified. The situation resembles the traditional problem of values: one must start somewhere.

One might think that an individual endotherm is automatically not equivalent to an individual ectotherm of the same size in the energetic paradigm, while they are in the informational paradigm. However, the extra energy of endotherms is used in maintenance, not in expansion. Endotherms can have an advantage from faster life cycles, less relative difficulty in providing for young, etc., but such effects are included in both paradigms.

"From time to time it is suggested that evolution maximizes a variety of parameters . . . , but the kinetic equations of genetic change do not allow any of these maximization principles" (Lewontin, 1984). That is a true statement within the standard domain, as far as I know, but we have seen that this domain is too narrow to encompass all of evolution. For instance, expansion as such is quite excluded. The informational and energetic paradigms each have a real maximization principle which exists because the expanded domain, as well as large domains within the standard paradigm, cancel the exceptions. Thus, frequency-dependent selection obviously increases the energy control of the positively selected unit (although not necessarily that of the population of all units) to a local maximum. Other cases are similarly trivial.

Can one then merge the two paradigms? They do seem in a way to be complementary rather than competing. Energy is part of the environment of DNA, and DNA specifies (yes, incompletely) what the energy is acting on. One doesn't have separate theories for an enzyme and its substrate. The contrast is somewhat like that between Aristotle's efficient and formal causes. However, I don't think that such a reconciliation is possible without subordinating one approach to the other. They do make some different predictions, even though at time scales which aren't easily, if at all, amenable to empirical resolution. But maybe this doesn't matter.

Alternatively, one might think of leaving each paradigm intact and combining their results in some way. I think this is impossible. It is impossible to maximize two criteria at once unless they are specifically weighted so that they form one criterion, and any weighting is necessarily arbitrary in this case.

However, I do want to make a strong claim: the informational and energetic paradigms are the only two paradigms consistent with what we know about evolution. Both are unconventional. Unlike the usual paradigm, they each incorporate naturally the various gradations of the real biological world. And each invokes causal processes which we know are fundamental. In ontogeny information is primary, energy secondary. In evolution, though, I think the energetic approach is
both much more tractable and more natural at the level of the phenotype and above, where different sets of alleles (not to mention other nucleotide sets) commonly give the same result and it is the result which is selected as such. If the differences between the paradigms really don't matter, the energetic paradigm is also the only one simple enough to use in a precise way. But these judgments are based on esthetics and practicality, not on evidence.

In each paradigm there is a technical question about the scaling of fitness. It is appropriate to keep the standard convention that fitnesses are always non-negative. A fitness of 0 in some interval and environment is the expectation of elimination in that interval. A relative fitness of 1 can be given to the situation of exact replacement or persistence of the same unit at the same size (energy equivalent). The scale is linear, so that an expected doubling of expansive energy or number of copies of a set of nucleotides will give a relative fitness of 2. Multiplication of relative fitnesses by the constant appropriate in a particular case will give absolute fitnesses. All this is just as in standard theory, and the numerical relations and theory of the latter carry over directly although they may be interpreted a bit differently. As we have seen, though, standard theory isn't broad enough for the multifarious real biological world.

Energy for growth is not necessarily equivalent to that for reproduction. Even apart from different efficiencies, a dormant individual persists with negative or at most (in cryptobiosis) zero expansive energy while a reproducing individual must use energy to replace itself. The net result can be the same, though. On the other hand, in each case death results in loss of control of the expansive energy previously incorporated into growth and so it gives a reduction in momentary fitness, which can be scaled relatively as going from 1 to 0. Formally, then, in the energetic paradigm

$$w_t = F + \Delta F \quad \text{(or, in relative terms, } \frac{F + \Delta F}{F})$$

where $w_t$ is fitness in the environment or set of environments after time-interval $t$, $F$ is the realized and potential expansive free energy of the selective unit, and $\Delta F$ is its expected change in $t$. The unit may of course be fuzzily bounded. I will deal with the simple and almost conventional mathematics elsewhere; it has no conceptual surprises.

There is an important sense, though, in which a static large group seems to be more fit than a static small one, ceteris paribus: it is using more free energy in a productive way. "Large" must obviously be interpreted here energetically rather than in terms of number of subordinate units like individuals or species or copies of a nucleotide set. Change rather than possession is what natural selection acts on, so rate of increase (at the same level and time scale) rather than size per se is maximized. However, they are closely related and I don't think it would violate our sensibilities much to use the term realized fitness to refer to the expansive energy controlled by a selective unit.

An allele's energy control is its frequency, so genetic equations from the energetic viewpoint are the same as those in received theory if the units are appropriate and if absolute fitnesses are used. The same is true for population ecology, with its imprecise foundations.

Applications

So what can we do with all this, besides having the comfort of a coherent basis for our science? We can compare energy used for growth with that used for reproduction and note that we get a contrast rather like that between $K$-selection and $r$-selection respectively. (The differences between the comparisons may be of interest, though.) We can look at natural selection within entire communities in a natural and causal way. We can compare the fitness of a bird with that of a fungus. (Should one become the other if it could? -- a question of some interest
in cases where evolutionary change is more realistic.) We can be motivated to incorporate energy into life tables to get a useful result I have called energy tables (Van Valen, 1975).

We can note that phagiers, i.e. predators, herbivores, and such, remove expansive energy, with the same result but in a different way from ordinary competitors. So phagy (Van Valen, in press) and competition are aspects of a single more general process which can perhaps best be subsumed under the rubric of generalized competition: all organisms try to get more free energy, for its acquisition is normally in short supply relative to their particular needs and constraints. They just do so in different ways. The differences are important, but so is the similarity.

I note here a mistake in my earlier analysis (Van Valen, 1973a), of the Red Queen's Hypothesis. The zero-sum assumption, which I regard as central, relates to total use of free energy, not just to expansive energy. Because, e.g., endotherms are more than an order of magnitude less efficient than ectotherms, this can make a difference in some cases. So, strictly speaking it is not realized fitness which has a zero-sum constraint, but overall energy use. Wastrels reduce the total realized fitness of their community, probably on all time scales because wastrels tend to be evolutionarily volatile, the getters-and-spenders of evolution. Even today, though, wastrels are rarely an energetically major component of communities. And, to the extent that their portion is itself constant and wastrels and non-wastrels are homogeneous groups in their efficiency, the zero-sum constraint applies to realized fitness directly. Of course these groups aren't strictly homogeneous, but then the zero-sum assumption itself doesn't need to be exact.

Energy and community evolution

It is probably in the study of community evolution that the energetic paradigm is heuristically most promising. Community evolution is more than species replacement. It tends to be ignored; even its existence is denied. It is supposed to be an epiphenomenon of the changes in distribution and evolution of species and lineages. These behave individualistically, each in response to its own opportunities and constraints. In spatially gradational environments their boundaries do not even approximately coincide. Over time communities are assembled and break up. Some species now widely separated from each other were sympatric several thousand years ago because of the different climate then.

All this is true, at least as a good approximation, but to think that there is then no basis for real evolution of communities is to cut down the corn to get at the weeds. Communities aren't superorganisms, but they have structure and components and processes and controls of their own (Ricklefs, 1979). These attributes change over time, and community evolution consists of these changes. Succession is one kind of short-term community evolution: that it is more or less cyclic and predictable doesn't remove its similarities to longer-term changes. The shifts in abundance of insect species from year to year, usually asynchronously in places even tens of kilometers apart (Taylor and Taylor, 1979; Taylor, 1986), represent another kind of short-term community evolution. Expansion predominates in the latter evolution, while local extinction and immigration, together with expansion from dormancy for plants and microorganisms, predominate in succession. Seasonal cycles in abundance of short-lived species, like insects (all life stages) and plankton, are often even more predictable. That evolution can show such regularities shouldn't delude us into thinking it nonevolutionary; genetic microevolution and long-term evolution within clades, the kinds usually thought of, are indeed irregular in important ways, but they aren't all of evolution.

We often need to refer to ecologically integrated regions larger than local communities. Some of these have names, but there is no simple way to refer to all spatial scales jointly. I therefore propose that the term "community" have no upper bound on size, and that local communities be referred to as local communities
whenever confusion is possible. We implicitly use different spatial scales even now when we focus on organisms of quite different sizes, such as birds and bacteria.

Evolution within species and lineages also contributes to community evolution; this is perhaps too obvious to mention. Less obvious is the community control on much or even most of this within-taxon evolution (e.g., Van Valen, 1983; Wilson, 1980; Vermeij, 1987). And community evolution extends over geological time also, where there are aspects not apparent at shorter time scales (e.g., Boucot, 1983; Zambach, 1977, 1983; Van Valen, 1984a).

A major aspect of natural selection in communities is usually not recognized as such because it is called interspecific competition, involving what Damuth (1985) termed avatars of the whole species. Supraspecific natural selection in communities can also occur, like other levels both as direct competition and as differential response to the physical environment. I have discussed this matter elsewhere (Van Valen, submitted). Supraspecific taxa aren’t adaptively arbitrary units, except to cladists, and there are various examples of selection involving such taxa and other community subgroups (e.g., Van Valen and Sloan, 1966; Ashton, 1988; Van Valen, submitted).

Our first concern is with life itself, with the flow of free energy which powers it. This can be in a local community, or worldwide, or at some intermediate breadth. Thus we view the biotic world as a system of energy flow. If this is taken seriously we can ask in what ways the flow is partitioned, what causes the flow and its partitions and regulates the various rates of flow. This sounds suspiciously like systems ecology, and the resemblance is real although the categories and methods of systems ecology may often be inappropriate. (For instance, Odum’s central concept of energy quality is not, as far as I know, relevant to evolution beyond community assembly. Free energy is free energy, however many links it may have gone through before reaching its present location).

The study of food webs and other topics are also relevant, and the approach has ramifications into a variety of other subdisciplines not commonly regarded as related, such as aspects of physiology and biogeography. Community evolution is potentially a subdiscipline itself, particularly if we think of communities as being of different degrees of homogeneity on different spatiotemporal scales. We can ask how the partitions and flows change over time, at very different time scales, and what processes cause these changes. I have elsewhere (Van Valen, submitted) looked at such questions explicitly, using the Paleocene mammals of North America, and have found surprisingly large differences during different phases of community evolution. Evolution is not just what happens to lineages or clades.

To the extent that the energetic paradigm has merit, it replaces genetics by ecology as the central causal subject of the evolutionary half of biology. Furthermore, it is probably appropriate to think of supramolecular evolution as being directed mostly by movement of the phenotype x environment adaptive landscape (on which see Dodson, 1975). The evidence from the bloom phase of adaptive radiation (Van Valen, 1985b), as well as other considerations, suggests that genetic (when distinct from developmental) constraints are rarely important in real evolution. From this perspective too, then, ecology replaces genetics as the central causal subject.

There is a lot more that could be said, but we may see here a way to a deep causal unification of population biology itself and its neighbors. Think about it sometime.

Acknowledgments

I thank V.C. Maiorana and the referees for improvements.
Literature cited


Sepkoski, J.J., Jr., and A.I. Miller. 1985. Evolutionary faunas and the
distribution of Paleozoic benthic communities in space and time. In:
Phanerzoic Diversity Patterns (J.W. Valentine, ed.), pp. 181-190. Princeton:
Princeton Univ. Press.

Sepkoski, J.J., Jr., and P.M. Sheehan. 1983. Diversification, faunal change, and
community replacement during the Ordovician radiations. In: Biotic Interactions
in Recent and Fossil Benthic Communities (M.J.S. Tevesz and P.L. McCall, ed.),


Review of Biology, 51: 3-47.

Taylor, L.R. 1986. Synoptic dynamics, migration and the Rothamsted Insect Survey


Thayer, C.W. 1983. Sediment-mediated biological disturbance and the evolution of
marine benthos. In: Biotic Interactions in Recent and Fossil Benthic
Press.

Thoday, J.M. 1953. Components of fitness. Symposia of the Society for
Experimental Biology, 7: 96-113.

Thompson, E.S. 1972. Rates of change of world ABO blood-group frequencies.

Trifonov, E.N. 1989. The multiple codes of nucleotide sequences. Bulletin of
Mathematical Biology, 51: 417-432.


----. 1974. A natural model for the origin of some higher taxa. Journal of
Herpetology, 8: 109-121.
----. Submitted. Levels of selection in the early Cenozoic radiation of mammals.
In: Evolutionary Innovations (M.H. Nitecki, ed.)

Systematic Zoology, 15: 261-278.

527 pp.

between the functional shell morphology of freshwater and marine pelecypods.
Three paradigms of evolution


