Living in a physical world

STEVEN VOGEL
Biology Department, Box 90338, Duke University, Durham, NC 27708–0338, USA
(Fax, 919-660 7293*; Email, svogel@duke.edu)*

The diversity of life on earth dazzles all of us – the rich profusion of its designs, the wide size range of its organisms, the complexities of its hierarchical levels, and so forth. Undaunted, we life scientists seek broadly applicable rules, common patterns of organizations, and order beneath the perceptual chaos; we look for alternatives to the easy answers of revealed truth.

Biology, no less than the physical sciences, treads this bumpy path – indeed the overt diversity of life puts especially bad bumps in its way. Perhaps its special difficulty underlies the gradual estrangement of biology from the more obviously successful physics of the post-Newtonian era and its awkward reintegration into the larger world of science in the twentieth century. That process remains incomplete; blame, if leveled, rests on the untidiness and distinctiveness of the subject. The tidy formulas of Newtonian physics work even less well for us than they do for, say, practicing engineers. Life directs its chemistry with sets of governing molecules and carries it out with the aid of catalysts of breathtaking specificity. And biology enjoys a strange organizing principle, evolution by natural selection, barely hinted at elsewhere in science.

No aspect of this reintegration has been (and continues to be) more successful than what we have come to call molecular biology – a statement at once fashionable and incontrovertible, one with which I have no grounds to take issue. What matters here, indeed the entire justification for the essays that begin with the one here, comes down to the following. The very success of this chemically-reductionist biology too easily diverts us from other conjunctions of physical science and biology.

This series will explore aspects of biology that reflect the physical world in which organisms find themselves. Evolution can do wonders, but it cannot escape its earthy context – a certain temperature range, a particular gravitational acceleration, the physical properties of air and water, and so forth. Nor can it tamper with mathematics. The baseline they provide both imposes constraints and affords opportunities. I mean to explore both.

And I will take what other biologists might find an unfamiliar approach – one, by the way, that I have found productive enough to recommend. Instead of asking about the physical science behind a specific biological system, I will consider aspects of the physical world and ask what organisms, any organisms, make of each, both how they might capitalize on them and be in some fashion limited by them. In effect, this will be a search for commonalities and patterns, the only unusual feature being the physical rather than biochemical or phylogenetic bases. If this approach to science were a dart game, I would be thrown out – for throwing darts at a wall first and only subsequently painting targets around the points of impact.

The series will concern itself mainly (but not exclusively) with organisms rather than ecosystems or organelles. It will follow the author’s bias and personal experience toward mechanical matters, doing less than equal justice to radiations and electrical phenomena. It will be speculative, opinionated, and idiosyncratic, aiming to stimulate thought and perhaps even investigation, to open doors rather than just describing them.

When I began to do science, over forty years ago, I wondered first whether and then where I would get ideas worth pursuing. Now, on the cusp of retirement, I wonder what I am going to do with my accumulated head- and notebooks-full of questions. Maybe we need something like a patent expiration date – if one does nothing with a hypothesis for some number of years, it should somehow revert to the public domain. I am not an unequivocal advocate of a strict rule, inasmuch as I have, on occasion, resurrected one of my old ideas, applying some additional insight or new tool in my experimental armamentarium – or just responding to a renewed interest. Still, these essays should, if nothing else, provide an opportunity to air untested ideas with some hope that others might care to pursue them.

Living in a physical world I. Two ways to move material

"No man is an island, entire of itself," said the English poet John Donne. Nor is any other organism, cell, tissue, or organ. We are open systems, continuously exchanging material with our surroundings as do our parts with their surroundings. In all of these exchanges, one physical process inevitably participates. That process, diffusion, represents the net movements of molecules in response to thermal agitation and place-to-place concentration differences. On any biologically-relevant scale, it can be described by exceedingly precise statistical statements, formulas that take advantage of the enormous numbers of individual entities moving around. And it requires no metabolic expenditure, so it is at once dependable and free.

But except over microscopic distances diffusion proceeds at a glacial pace. For most relevant geometries, a doubling of distance drops the rate of transport per unit time by a factor, not of two, but of four. Diffusive transport that would take a millisecond to cover a micrometer would require no less than a thousand seconds (17 min) to cover a millimeter and all of a thousand million seconds (3 y) for a meter. Diffusion coefficients, the analogs of conventional speeds, have dimensions of length squared per time rather than length per time.

Organisms that rely exclusively on diffusion for internal transport and exchange with their surroundings, not surprisingly, are either very small or very thin or (as in many coelenterates and trees) bulked up with metabolically inert cores. Those living in air (as with many arthropods) can get somewhat larger since diffusion coefficients in air run about 10,000 times higher than in water, which translates into a hundred-fold distance advantage. Beyond such evasions, macroscopic organisms inevitably augment diffusion with an additional physical agency, convection, the mass flow of fluids. Circulatory systems as conventionally recognized represent only one version of a ubiquitous scheme.

One might expect that good design balances the two physical processes. Excessive reliance on diffusion would limit size, slow the pace of life, or require excessively surface-rich geometries. Excessive reliance on flow would impose an unnecessary cost of pumping or require an unnecessarily large fraction of body volume for pipes, pumps, and fluid. A ratio of convective transport to diffusive transport ought, in other words, to have values around one for proper biological systems. Such a ratio represents nothing novel; one has long been used by chemical engineers. This so-called Péclet number, *Pe*, is a straightforward dimensionless expression:

$$Pe = \frac{vl}{D} , \tag{1}$$

where v is flow speed, l is transport distance, and D is the diffusion coefficient. (Confusingly, a heat-transfer version of the Péclet number may be more common than this mass-transport form; it puts thermal diffusivity rather than molecular diffusion coefficient in its denominator.)

Calculating values of the Péclet number can do more than just give a way to check the performance of the evolutionary process. In particular, it can provide a test for our hypotheses about the primary function of various features of organisms. I think that justification can be put best as a series of examples, which will follow after a few words about the origin of this simple ratio.

One can view the ratio as a simple numerator, mv, for bulk flow, with a denominator representing a simplified form of Fick's first law for transport (mass times distance divided by time) for diffusion, DSm/V, where S is cross section and V is volume. Using l^2 for area and l^3 for volume, one gets expression (1). Of course, the way we have swept aside all geometrical details puts severe limits on what we can reasonably expect of values of *Pe*. Only for comparisons among geometrically similar systems can we have real confidence in specific numbers. Still, living systems vary so widely in size that even order-of-magnitude values ought to be instructive.

From a slightly different viewpoint, the Péclet number represents the product of the Reynolds number (*Re*) and the Schmidt number (*Sc*). The first,

$$Re = \frac{r\,l\,v}{m} , \tag{2}$$

where r and m are fluid density and viscosity respectively, gives the ratio of inertial to viscous forces in a flow. At high values bits of fluid retain a lot of individuality, milling turbulently as in a disorderly crowd; at low values bits of fluid have common aspirations and tend to march in lock-step formation. In short, it characterizes the flow. The second,

$$Sc = \frac{m}{r\,D} , \tag{3}$$

is the ratio of the fluid's kinematic viscosity (viscosity over density) to the diffusion coefficient of the material diffusing through it. It gives the relative magnitudes of the diffusivities of bulk momentum and molecular mass. In short, it characterizes the material combination, solute with solvent, that does the flowing.

A few cases where calculating a Péclet number might prove instructive.

(i) *The sizes of our capillaries and kidney tubules*

Consider our own circulatory systems, in particular the size of the vessels, capillaries, where function depends on both diffusion and flow. Do we make capillaries of proper size? After all, we devote about 6·5% of our body volume to blood and expend about 11% of our resting metabolic power pushing it around – so it ought to be important. And it appears that we do size them properly. For a capillary radius of 3 μm, a flow of 0·7 mm s^{-1}, and a diffusion coefficient (assuming oxygen matters most) of 18×10^{-10} m^2 s^{-1}, the Péclet number comes to 1·2. If anything, the value turns out a bit better than one expects, given the approximations behind it (Middleman 1972).

Of course nature might pick different combinations of radius and flow speed without offending Péclet. (We ignore the side issue of fit of red blood cells, tacitly assuming that their size is evolutionarily negotiable.) Smaller vessels would permit faster flow and lower blood volume, but the combination would, following the Hagen-Poiseuille equation, greatly increase pumping cost. Larger vessels require greater blood volume, the latter already fairly high, and slower flow, which would make the system less responsive to changes in demand. One suspects something other than coincidence for the similar blood volume (5·8%) in an octopus (Martin *et al* 1958).

Quite likely this choice of capillary size, based on Péclet number and some compromise of volume versus cost, sets the sizes of much of the rest of our circulatory systems in an effective cascade of consequences. According to Murray's law (LaBarbera 1990) the costs of construction and operation set the relative diameters of all vessels; thus if something sets diameter at one level in their hierarchy, it ends up determining the diameters of all the rest. The rule is a simple one – branching conserves the cubes of the radii of vessels, so the cube of the radius of a given vessel equals the sum of the cubes of the vessels at some finer level of branching that connect with it.

What about the reabsorptive tubules of our kidneys, in particular those just downstream from the glomerular ultrafiltration apparatus? Again, the system represents a far-from-insignificant aspect of our physiology; 20 to 25% of the output of the heart passes through this one pair of organs. About 20% of the plasma volume squeezes out of the blood in the process, in absolute terms around 60 ml min^{-1} per kidney. Each kidney consists of about 2,000,000 individual units, the nephrons. Thus each glomerulus sends on for selective reabsorption about $0{\cdot}5 \times 10^{-12}$ m^3 s^{-1}.

The sites of the initial phase of reabsorption are the proximal tubules, each about 40 micrometers in inside diameter. Combined with the earlier figure for volume flow, that means a flow speed of 0·40 mm s^{-1}. So we have speed and size. Diffusion coefficient can be assigned no single number, since the tubules reabsorb molecules spanning a wide size range, from small organic molecules and ions to small proteins with molecular weights of around 40,000. So coefficients most likely range from about $0{\cdot}75 \times 10^{-10}$ to 40×10^{-10} m^2 s^{-1}. That produces Péclet numbers from 2 to 100. At first glance these seem a bit high, but the story has an additional aspect. Those tubules reabsorb at least 80% of the volume of the filtrate, so by the time fluid leaves them, its speed has dropped by at least a factor of 5. That gives exit Péclet numbers a range of 0·4 to 20, with an average number in between – quite reasonable values, indicative (to be presumptuous) of good design. Flow in the tubules comes at a relatively low cost, at least relative to the power requirements of filtration and the kidney's chemical activities. So one might speculate that the system contrives to bias its Péclet numbers so for most molecules over most of the length of the tubule values exceed one, albeit not by much.

(ii) *The size of plant cells*

One can argue that the boundary between the cellular and the super- (or multi-) cellular world reflects the upper size limit of practical, diffusion-based systems, that getting above cell size takes some form of convective augmentation of transport. I like that view, which tickles my particular biases. But I have to admit that the notion cannot apply to plant cells. On average, the cells of vascular plants run about ten times the size of animal cells, with "size" taken as typical length. They are of the order of 100 μm in length but somewhat less in width; 25 μm should be typical of the distance from membrane to center. That increased size might have devastating consequences for transport were it not for the internal convection common to such cells. Put another way, the size scale at which convective transport comes in does not correspond to the size of plant cells.

That bulk flow system within plant cells goes by the name "cyclosis." We know quite a lot – but far from all – about how microfilaments of actin (a key component of muscle) power it; only its speed matters here. That speed is around 5 μm s^{-1} (Vallee 1998). Focusing on oxygen penetration and using a penetration distance of 25 μm gives a Péclet number of 0·07. That tells us that the system remains diffusion dominated, that cyclosis does not reach a significant speed. Looking at carbon dioxide penetration, with a diffusion coefficient of 0·14 m^2 s^{-1}, raises that number too little to change the conclusion.

Perhap we should take a different view. Size, speed, and a presumptive Péclet number around one permit calculating a diffusion coefficient, which comes to $1{\cdot}25 \times 10^{-10}$ m^2 s^{-1}. That corresponds to a non-ionized molecule with a molecular weight of about 6000. Thus the system appears convection-dominated for proteins and other macromolecules and diffusion dominated for dissolved gases, amino acids, sugars, and the like.

(iii) *Sinking speeds of phytoplankton*

Diatoms plus some other kinds of small algae account for nearly all the photosynthetic activity of open oceans.

Paradoxically for light-dependent organisms, most of the time most of these phytoplankters are negatively buoyant. Not that they sink rapidly; 4 μm s^{-1} (a foot a day, in the antediluvian units used where I live) is typical. According to one commonly cited view, that sinking improves access to carbon dioxide by minimizing the depletion around a cell caused by its own photosynthetic activity. In effect, the cell walks away from its personal environmental degradation. Still better, it walks away with no cost of locomotion. Of course it (or its progeny) may eventually suffer, as the sinking brings it down to depths at which net photosynthesis cannot be achieved. Somehow (and wave-induced water mixing comes into the picture) the cost-benefit analysis favours this slight negative buoyancy.

Calculating a Péclet number casts serious doubt on this view, doubt first raised (with an equivalent argument) by Munk and Riley (1952). For a diatom about 10 μm in diameter, that sinking rate of 4 μm s^{-1}, and the diffusion coefficient of CO_2, 14×10^{-10} m^2 s^{-1}, we get a value of 0·03. Diffusion, in short, rules; convection, here due to sinking, will not significantly improve access to carbon dioxide. We might have chosen a slightly larger distance over which CO_2 had to be transported to be available at adequate concentration, but even if a distance ten times longer were chosen, the conclusion would not be altered.

Why, then, should a phytoplankter sink at all? The calculation tacitly assumed uniform concentration of dissolved gas except where affected by the organism's activity, so it might be seeking regions of greater concentration, lowering sinking rate wherever life went better. In a world mixed by the action of waves that seems unlikely, even if (as appears the case) buoyancy does vary with the physiological state of a cell. Perhaps phytoplankters bias their buoyancy toward sinking so they are not likely to rise in the water column and get trapped by surface tension at the surface. If perfect neutrality can not be assured, then sinking may be preferable, as long as the speed of sinking can be kept quite low. Surface tension may be a minor matter for us, but it looms large for the small. In the millimeter to centimeter range a creature can walk on it – the Bond number, the ratio of gravitational force to surface tension force is low. Below that a creature may not be able to get loose once gripped by it – the Weber number, the ratio of inertial force to surface tension force drops too far (Vogel 1994). But that argument presumes that diatoms have hydrophobic surfaces, which, I am told, may not be the case. So another hypothesis would be handy.

(iv) *Swimming by microorganisms and growing roots*

More often we think of movement by active swimming than by passive sinking. Some years ago, the physicist Edward Purcell (1977) wrote a stimulating essay about the physical world of the small and the slow, looking in particular at bacteria. Among other things, he asked whether swimming, by, say, *Escherichia coli*, would improve access to nutrients. By his calculation, a bacterium one μm long, swimming at 20 μm s^{-1} (see Berg 1993), would only negligibly increase its food supply, assumed to be sugar. To augment its supply by a mere 10%, it would have to go fully 700 μm s^{-1}. Purcell's answer to why swim at all turned on the heterogeneity of ordinary environments and the advantage of seeking the bacterial equivalent of greener pastures, as suggested above for diatoms. Otherwise the bacterium resembles a cow that eats the surrounding grass and then finds it most efficient to stand and wait for the grass to grow again.

The Péclet number permits us to cast the issue in more general terms. Sucrose has a diffusion coefficient of $5{\cdot}2 \times 10^{-10}$ m^2 s^{-1}; together with the data above we get a Péclet number of about 0·04. Swimming, as Purcell found, should make no significant difference. But the conclusion should not be general for microorganisms. Consider a ciliated protozoan, say *Tetrahymena*, which is 40 μm long and can swim at 450 μm s^{-1}. If oxygen access is at issue, the Péclet number comes to 10, indicating that swimming helps a lot. Indeed it might just be going unnecessarily fast, prompting the thought that getting enough of some larger molecule might underlie its frantic pace. Or it might swim for yet another reason.

Growing roots provide a case as counterintuitive as the result for swimming bacteria but in just the opposite direction (Kim *et al* 1999). A root can affect nutrient uptake by altering local soil pH. Root elongation speed runs around 0·5 μm s^{-1}, slower than the most sluggish tortoise. But it turns out to constitute a significant velocity, enough so that (at least in sandy soil) the Péclet number gets well above one. Values for the rapidly diffusing H^+ ions for typical growing roots may exceed 30, using root diameter as length. That means motion affects the pH distribution in the so-called rhizosphere more than does diffusion.

(v) *Flow over sessile organisms*

For sinking diatoms and swimming microorganisms we evaluated hypotheses about why creatures did what they did. In some loosely analogous situations we can test claims about their physical situations, in particular about flows. How fast must air or water flow over an organism to affect exchange processes significantly? To put the matter in sharper terms, can the Péclet number help us evaluate a claim that extremely slow flow matters? After all, neither producing nor measuring very low speed flows is the most commonplace of experimental procedures.

For instance, consider the claim that a flow of 0·2 to 0·3 mm s^{-1}, around a meter per hour, significantly increases photosynthesis in an aquatic dicot, *Ranunculus pseudofluitans* (Westlake 1977). The finely dissected, almost filamentous leaves are about 0·5 mm across. Inserting the diffusion coefficient of CO_2 gives a Péclet number around 100, which certainly gives credibility to the report. One guesses that even slower flows should be significant.

Another paper (Schumacher and Whitford 1965) reports that a flow of $10\ \mathrm{mm\ s^{-1}}$ significantly increases photosynthesis in a green alga, *Spirogyra*, made up of filaments about 50 µm in diameter. A Péclet number of about 300 provides emphatic support, again suggesting that far slower flows should also matter. Conversely, it prompts one to ask whether so-called still water, the control in such comparisons, was still enough so flow was truly negligible. My own experience suggests that thermal convection and persistence of filling currents can complicate attempts to prevent water from flowing – still water does not just happen.

A third paper (Booth and Feder 1991) looked at the influence of water flow on the partial pressure of oxygen at the skin of a salamander, *Desmognathus*. It found that currents as low as $5\ \mathrm{mm\ s^{-1}}$ increased that partial pressure, facilitating cutaneous respiration. With a diameter of 20 mm, that flow produces a Péclet number of 50,000. A sessile *Desmognathus* may need flow, but it does not need much. Again, the quality of any still-water control becomes important.

(vi) *Two functions for gills*

Most swimming animals use gills to extract oxygen from the surrounding water. Whatever their particulars, gills have lots of surface areas relative to their sizes. Many aquatic animals suspension feed, extracting tiny edible particles from the surrounding water. Whatever their particulars, such suspension feeding structures have lots of surface areas relative to their sizes. While most suspension-feeding appendages look nothing like gills, some not only look like gills but share both name and functions. No easy argument implies that such dual function gills should balance those two functions. Quick calculations of Péclet number can tell us which function dominates their design and help us to distinguish respiratory gills from dual-function gills.

Consider a limpet, *Diodora aspera*, a gastropod that uses its gills for respiration. With gill filaments about 10 µm apart, a flow rate of $0{\cdot}3\ \mathrm{mm\ s^{-1}}$ (J Voltzow, personal communication), and the diffusion coefficient for oxygen, the Péclet number comes to about 2. A bivalve mollusk, the mussel *Mytilus edulis*, with dual function gills presents a sharp contrast. The effective distance here is about 200 µm and the speed about $2\ \mathrm{mm\ s^{-1}}$ (Nielsen *et al* 1993). That gives a Péclet number around 100 for oxygen access. Clearly the system pumps far more water than necessary were respiration the design-limiting function.

One can do analogous calculations for fish, where a few kinds use gills for suspension feeding as well as respiration. A typical teleost fish has sieving units 20 µm apart (Stevens and Lightfoot 1986) with a flow between their lamellae of about $1\ \mathrm{mm\ s^{-1}}$ (calculated from data of Hughes 1966). For oxygen transport, the resulting Péclet number is 5·5, not an unreasonable value for an oxygenating organ. One gets quite a different result for a fish that uses its gills for suspension feeding. A somewhat higher 80 µm separates adjacent filtering elements. but the main difference is in flow speeds. These run around $0{\cdot}15\ \mathrm{m\ s^{-1}}$ for passive ("ram") ventilators (Cheer *et al* 2001), and $0{\cdot}55\ \mathrm{m\ s^{-1}}$ for pumped ventilators (Sanderson *et al* 1991). The resulting Péclet numbers, 6,500 and 20,000 (again using oxygen diffusion) exceed anything reasonable for a respiratory organ.

(vii) *Air movement and stomatal exchange*

All of the previous examples involve diffusion and convection in liquids. The same reasoning ought to apply to gaseous systems as well – fluids are fluids, and diffusion and convection occur in all.

Leaves lose, or "transpire," water as vapour diffuses out though their stomata and disperses in the external air. Transpiration rates depend on a host of variables, among them wind speed and stomatal aperture, the latter under physiological control. Immediately adjacent to a leaf's surface, the process depends, as does any diffusive process, on concentration gradient, from the saturated air at the stomata to whatever might be the environmental humidity. The stronger the wind, the steeper the concentration gradient as the so-called boundary layer gets thinner.

Consider a bit of leaf 20 mm downstream from the leaf's edge, with downstream indicating the local wind direction. And assume a wind about as low as air appears to move for appreciable lengths of time, as a guess, $0{\cdot}1\ \mathrm{m\ s^{-1}}$. The effective thickness of the velocity gradient outward from the leaf's surface can be calculated (Vogel 1994) as

$$d = 3{\cdot}5\sqrt{\frac{xm}{vr}}\ , \tag{4}$$

where x is the distance downstream, and m and r are the air's viscosity and density, respectively, 18×10^{-6} Pa s and $1{\cdot}2\ \mathrm{kg\ m^{-3}}$. The thickness comes to 6 mm. (This must be regarded as a very crude approximation; among other things, the formula assumes a thickness that is much less than the distance downstream.) With that thickness, that wind speed, and the diffusion coefficient of water vapour in air, $0{\cdot}24 \times 10^{-4}\ \mathrm{m^2\ s^{-1}}$, the Péclet number is 25. So even that low speed suffices to produce a convection-dominated system.

What might that tell us? It implies, for one thing, that changes in wind speed should have little or no direct effect on water loss by transpiration. If water loss does vary with wind speed, one should look for something other than a direct physical effect, something such as changes in stomatal aperture. For another thing, it implies that a leaf in nature would not have adjacent to its surface very much of a layer of higher-than-ambient humidity. So-called "vapour caps" are not likely to mean much with even the most minimal of environmental winds.

(viii) *The sizes of morphogenetic fields and synaptic clefts*

A variant of the Péclet number may give some insight into such things as the development of animals. Much of pattern formation depends on the diffusion of substances, morphogens, whose concentration gradients establish embryonic fields. Establishing larger fields not only means lower gradients (or higher concentrations of morphogens) but would take more time, a non-negligible resource in a competitive world. Breaking up velocity into length over time we get:

$$\frac{l^2}{Dt} \, . \tag{5}$$

(The reciprocal of this expression is sometimes called the mass transfer Fourier number.)

To get a situation in which diffusion is not relied on excessively, we might assume a value of one. A typical morphogen has a molecular weight of 1000; its diffusion coefficient when moving through cells (a little lower than in water) ought to be around 1×10^{-10} m^2 s^{-1}. A reasonable time for embryonic processes should be a few hours, say 10^4 s. The numbers and the equation imply embryonic fields of around 1 mm, about what one does indeed find. The argument for the size of embryonic fields (put somewhat differently) was first made by Crick (1970).

In effect, the calculation produces what we might consider a characteristic time for a diffusive process. Consider ordinary synaptic transmission in a nervous system. The most common transmitter substance, acetylcholine, has a molecular weight of 146 and a diffusion coefficient around 7×10^{-10} m^2 s^{-1}. With a 20 nm synaptic cleft, the corresponding time comes to 0·6 ms. That value is at most slightly below most cited values for overall synaptic delay, which run between about 0·5 and 2·0 ms, implying that much or most of the delay can be attributed to transmitter diffusion.

Where else might calculations of Péclet numbers provide useful insight? We have not considered, for instance, olfactory systems, either aerial or aquatic. Are the dimensions and flow speeds appropriate in general; are they appropriate for the specific kinds of molecules of interest to particular animals? What of the speeds and distances of movement of auxins and other plant hormones? Might we learn anything from comparing systems in which oxygen diffuses within a moving gas with ones in which it diffuses in a flowing liquid, systems such as, on the one hand, the tubular lungs of birds and the pumped tracheal pipes of insects and, on the other, the gills of fish, crustaceans, and the like?

In fields such as fluid mechanics and chemical engineering, dimensionless numbers pervade have amply proven their utility. I argue here, as I did on a previous occasion (Vogel 1998) that they can help us see the relevance of physical phenomena to biological systems. Péclet number may be an especially underappreciated one, but (as I hope to illustrate in further pieces) far from the only one worth our consideration.

Who, incidentally, was this person Péclet? One does not normally name a number after oneself. Someone may propose a dimensionless index and then the next person who finds it useful names it after the first. Or the first to use one may name it for some notable scientist who worked in the same general area. Péclet number is a case of the latter. Jean Claude Eugène Péclet (1793–1857) was part of the flowering of French science just after the revolution. He was a student of the physical chemists (as we would now call them) Gay-Lussac and Dulong – names yet remembered for their laws – and a teacher of physical science. He did noteworthy experimental work on thermal problems and wrote an influential book, *Treatise on Heat and its Applications to Crafts and Industries* (Paris 1829).

Putting his name on a dimensionless number was done a century later, by Heinrich Gröber, in 1921, in another important book, *Fundamental Laws of Heat Conduction and Heat Transfer*. That thermal version of the Péclet number antedates the mass-transfer version used here. The latter, as far as I can determine, first appears in a paper on flow and diffusion through packed solid particles, by Bernard and Wilhelm, in 1950. They note its similarity to the dimensionless number used in heat-transfer work and call their version a "modified Peclet group, symbolized Pe' ". They shift, confusingly and deplorably, from an acute accent in "Péclet" to a prime ('), now usually omitted, at the end. Analogous indices for thermal and material processes is not unusual, but ordinarily the two carry different names – such as Prandtl number and (as earlier) Schmidt number. Amusingly, most sources mention one of the versions of the Péclet number with no acknowledgement that there is any other.

Acknowledgements

I thank Fred Nijhout, Howard Riesner, Janice Voltzow, and Peter Jumars for assistance in gathering data and imposing coherence on this rather disparate material.

References

Berg H C 1993 *Random walks in biology* (Princeton: Princeton University Press)

Bernard R A and Wilhelm R H 1950 Turbulent diffusion in fixed beds of packed solids; *Chem*. *Eng*. *Progr*. **46** 233–244

Booth D T and Feder M E 1991 Formation of hypoxic boundary layers and their biological implications in a skin-breathing aquatic salamander, *Desmognathus quadramaculatus*; *Physiol*. *Zool*. **64** 1307–1321

Cheer A Y, Ogami Y and Sanderson S L 2001 Computational fluid dynamics in the oral cavity of ram suspension-feeding fishes; *J*. *Theor*. *Biol*. **210** 463–474

Crick F 1970 Diffusion in embryogenesis; *Nature* (*London*) **255** 420–422

Hughes G M 1966 The dimensions of fish gills in relation to their function; *J*. *Exp*. *Biol*. **45** 177–195

Kim T K, Silk W K and Cheer A Y 1999 A mathematical model for pH patterns in the rhizospheres of growth zones; *Plant*, *Cell Environ*. **22** 1527–1538

LaBarbera M 1990 Principles of design of fluid transport systems in zoology; *Science* **249** 992–1000

Martin A W, Harrison F M, Huston M J and Stewart D W 1958 The blood volume of some representative molluscs; *J*. *Exp*. *Biol*. **35** 260–279

Middleman S 1972 *Transport phenomena in the cardiovascular system* (New York: John Wiley)

Munk W H and Riley G A 1952 Absorption of nutrients by aquatic plants; *J*. *Mar*. *Res*. **11** 215–240

Nielsen N F, Larson P S, Riisgård H U and Jørgensen C B 1993 Fluid motion and particle retention in the gill of *Mytilus edulis*: Video recordings and numerical modelling; *Mar*. *Biol*. **116** 61–71

Purcell E M 1977 Life at low Reynolds number; *Am*. *J*. *Phys*. **45** 3–11

Sanderson S L, Cheer A Y, Goodrich J S, Graziano J D and Callan W T 1991 Crossflow filtration in suspension-feeding fishes; *Nature* (*London*) **412** 439–441

Schumacher G J and Whitford H A 1965 Respiration and P^{32} uptake in various species of freshwater algae as affected by a current; *J*. *Phycol*. **1** 78–80

Stevens E D and Lightfoot E N 1986 Hydrodynamics of water flow in front of and through the gills of skipjack tuna; *Comp*. *Biochem*. *Physiol*. **83A** 255–259

Vallee R B 1998 *Molecular motors and the cytoskeleton* (San Diego: Academic Press)

Vogel S 1994 *Life in moving fluids* (Princeton: Princeton University Press)

Vogel S 1998 Exposing life's limits with dimensionless numbers; *Phys*. *Today* **51** 22–27

Westlake D R 1977 Some effects of low-velocity currents on the metabolism of aquatic macrophytes; *J*. *Exp*. *Bot*. **18** 187–205

Series

Living in a physical world
II. The bio-ballistics of small projectiles

STEVEN VOGEL
Department of Biology, Duke University, Durham, NC 27708-0338, USA
(Fax, 919-660-7293; Email, svogel@duke.edu)

1. Introduction

Many animals jump; many plants shoot their seeds. While ‘many’ may not imply ‘most’, terrestrial life is rife with examples of ballistic motion, motion in which a projectile gets all of its impetus prior to launch.

For most of us, the trajectories of projectiles appeared briefly early in a basic physics course. Some tidy equations emerged in unambiguous fashion from just two facts. A projectile moves horizontally at constant speed; only the downward acceleration of gravity (g) alters its initial vertical speed. Where launch and landing heights are the same, a simple formula links range (d) with launch speed (v_o) and projection angle (Q_0) above horizontal:

$$d = \frac{v_o^2 \sin 2q_o}{g} \quad . \tag{1}$$

So, for a given initial speed, a projectile achieves its greatest horizontal range when launched at an upward angle of 45°. That maximal range is simply

$$d_{\max} = \frac{v_o^2}{g} \quad . \tag{2}$$

Thus an initial speed of 40 m s^{-1} (144 km h^{-1}) could take a projectile 163 m. Enroute, the projectile reaches a maximum height, $h_{\max}$, of a quarter of that best range, or

$$h_{\max} = \frac{v_o^2}{4g} \quad . \tag{3}$$

The trajectory forms a nicely symmetrical parabola, and the loss of range at angles above 45° exactly mirrors the loss at lower angles – as shown in figure 1. Such tidiness gives (as once said) the biologist severe physics-envy.

In promoting these expressions, text or teacher may mutter, *sotto voce*, something about an assumed absence of air resistance, about the presumption that drag exerts a negligible effect.

Nevertheless the scheme generates significant errors even for a cannon ball. It gives still worse errors for golf balls – drag can halve the range of a well-driven golf ball (Brancazio 1984). The errors are tolerable only because golfers, however fanatic, rarely turn for help to physics. What keeps a projectile going is inertia; whether we view its consequences in terms of momentum or kinetic energy, mass provides the key element. Ignoring, to take a broad-brush view, variation in both density and shape, mass follows volume. What slows a projectile are two factors, gravity and drag. The standard equations deal with the downward force of gravity and produce their nice parabolas. Drag, the force that acts opposite the direction of motion, manifests itself in deviations from such simple trajectories; its magnitude varies in proportion either to surface area or diameter, depending on the circumstances. The smaller the projectile, the greater are both surface area or diameter relative to volume. So the smaller the projectile the less adequately that idealized, dragless trajectory should describe its motion. Since gravitational force, kinetic energy, and momentum all depend on mass, the less dense the projectile, the greater will be the relative influence of drag.

The upshot is that biological projectiles will be poorly served by these simple equations. Few are very large and none very dense, so their performances pale besides those of long-travelling and damage-inducing chunks of rock or iron. Still, life’s projectiles are diverse in ancestry, size, and function. Sports, hunting, and warfare, uses that come first to mind, matter least often to species other than our own. Instead, two functions dominate. Some organisms jump, forming single, whole-body projectiles; others shoot propagules – fruits, seeds, spore clusters, even individual spores.

2. Dealing with drag

In short, to look with any degree of realism at the trajectories of biological projectiles, we must, so to speak, put drag into the equation. As it happens, that turns out to be trickier than one might expect. We biologists imagine a physical world run according to straightforward (if sophisticated) rules, at least when compared with the messy scene that evolution generates. The drag, at least of a simple object such as a sphere, ought to behave with predictable lawfulness rather than with our eccentric awfulness; one should be able to look up a basic equation for drag versus speed or drag versus size. Not so! Within the range of speeds and sizes that might matter to organisms, these are distinctly ill-tempered functions. The trouble traces to changes, sometimes abrupt, in how fluids flow over objects, whether laminar or turbulent, whether surface-following or separated, and so forth. For a large object going at a fairly high speed, drag varies with the square of speed and the area of the object. For a small object going slowly, drag varies with speed itself and the length of the object. In between, the relationship bears no resemblance to anything that might tempt use of our customary regressions and power laws.

Fortunately, two twentieth-century accomplishments save the day. First, from direct measurements we know how drag varies with speed and size for ordinary objects such as spheres moving through ordinary fluids such as air and water. And, second, even the most minimal desktop computer now makes short work of calculating draggy trajectories by an iterative approach. One starts with a projectile of a given size, density, launch speed (‘muzzle velocity’ in the common parlance of these violent times), and launch angle. After a short interval, the computer informs us of the projectile’s slightly different speed and path, the two altered by gravity, acting downward, and drag, acting opposite the projectile’s direction. The computer then takes the new speed and path as inputs and repeats the calculation to get yet another speed and path. In the simplest case, the computer stops iterating when the projectile’s height has returned to that of its launch – when it has returned to the ground.

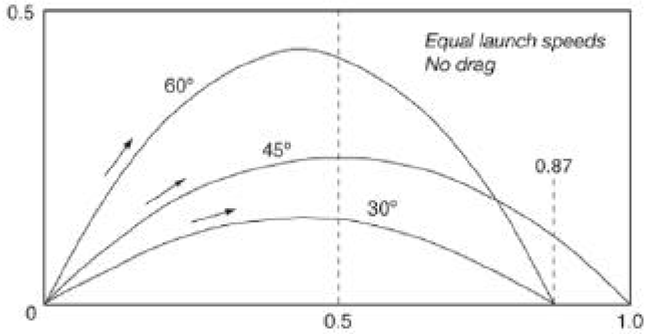

Figure 1. Without drag, trajectories are perfectly parabolic, with descent speeds and angles equal to ascent speeds and angles. For a given initial speed, maximum range occurs with a launch angle of 45º; ranges after either 30º or 60º launches are 87% of that maximum.

The way drag gets into the picture, though, takes a little explanation. We normally express drag in dimensionless form, as the so-called drag coefficient, C_d. It amounts to drag (D) relative to area (S) divided by a kind of idealized pressure, that which would push on something were the fluid coming directly at it to effect a perfect transfer of momentum and then obligingly (and quite unrealistically) disappear from the scene. Specifically,

$$C_d = \frac{D/S}{rv^2/2} \ , \tag{4}$$

where r is fluid density and v is the speed of the object through the fluid. The commonest reference area is the maximum cross section of the object normal to flow, the area facing the oncoming fluid. Unfortunately, the relationship between speed and drag coefficient behaves no better than that between speed and drag itself – the equation just dedimensionalizes drag. If drag were simply proportional to area, fluid density, and the square of velocity, then C_d would be constant (and unnecessary). So variation in C_d exposes the eccentricities of drag. And C_d depends, not only on shape, but on the object’s size and on the fluid’s viscosity (m here, but often h) and density.

Fortunately, these last three variables operate as a particular combination, the dimensionless Reynolds number, mentioned in the previous essay,

$$Re = \frac{rlv}{m} \ , \tag{5}$$

where flow-wise length provides the commonest reference, l (Vogel 2004). Again, Re represents the ratio of inertial to viscous forces as fluid crosses an immersed object. Untidy still, but now one needs to know only how drag coefficient varies with Reynolds number and all the other relationships follow, at least for a given shape.

For present purposes, this last function, $C_d = f(Re)$, breaks into three separate domains. When Re exceeds 100,000, (again assuming a sphere) $C_d = 0{\cdot}1$. For Re’s between 1,000 and 100,000, $C_d = 0{\cdot}5$. Thus for both domains, drag varies with the square of speed, but with different constants of proportionality. For Reynolds numbers below 1,000, the best encapsulation I have seen comes from White (1974):

$$C_d = \frac{24}{Re} + \frac{6}{1 + Re^{1/2}} + 0{\cdot}4 \ . \tag{6}$$

(The first term on the right represents Stokes’ law, trustworthy at Reynolds numbers below about one.) The computer need only decide, for each iteration, which of the approximations to apply.

Such a program gives all the important characteristics of a realistically draggy trajectory, starting with a projectile's size, density, launch angle, and launch speed – range, maximum height, impact angle, and impact speed. Looking at the computation point by point gives the shape of the trajectory. With only a little playing around one can work back from an observed range to a launch speed. Of course, the scheme assumes spherical projectiles, but most non-streamlined objects can be reasonably approximated as spheres of the same (or a little greater) volume. A version of such a program can be found in Appendix 2 of Vogel (1988).

For an initial example, consider a cannon and its projectile – a particular one whose barrel (on a new carriage) graces Edinburgh Castle for the delectation of tourists. James II of Scotland took delivery in 1457 of the weapon, called 'Mons Meg' after Mons, Belgium, where it was produced and an anonymous Meg (or Margaret). While too heavy to be of much use as a transportable siege weapon, apparently it could throw stone spheres half a meter in diameter about 3000 m. Assuming a typical density for stone (2700 kg m^{-3}), the computer program yields a launch speed of 180 m s^{-1} and a launch angle for maximum range (for that speed) of about 43º, a shade lower than the dragless 45º. No longer do the ranges at 30º and 60º match; now the 30º range wins by about 4%. Drag drops the best range of the projectile to 85·9% of the dragless, 45º calculation – we might say that it incurs a 'drag tax' of 14·1%. The difference would certainly have mattered – if the cannon could have been accurately aimed and ranged. So the simple formulas we were taught fall short (long, really) even where we thought they applied.

3. Playing games with balls

As suggested earlier, drag bothers a well-driven golf ball. If dragless, an initial speed of 60 m s^{-1} (216 km h^{-1}) would take it 365 m. Drag reduces that to 243 m, a tax of no less than 35·3%; that maximal range occurs with a launch angle of 41·5º. That noticeably distorts the standard parabola, with a descent a bit steeper than the preceding ascent and with a landing speed a little below launch speed. Is this result general for the balls we use in our various sports?

One might guess that a basketball, larger and less dense, would suffer relatively more from drag. But in practice its lower speed and thus relatively low drag (D, of course, not C_d) mitigates the problem. For a launch speed of 20 m s^{-1} (72 km h^{-1}) it goes 35·7 instead of 40·7 m, losing only 12·3%. And the best launch drops only a little below 45º, to 43·5º. A well-kicked football (in North America, a soccer ball) goes faster than a thrown basketball; unsurprisingly, its susceptibility to drag lies between those of golf and basketballs. A launch speed of 30 m s^{-1} (108 km h^{-1}) takes it 67·6 m instead of 91·6 m, a tax of 26·1%, with the best distance achieved with a launch angle of about 43º. Basketballs and footballs have about the same sizes and overall densities – launch speed determines the difference. For none of these, though, does drag amount to more than a secondary factor.

Since so much in fluid mechanics depends on the Reynolds number, we might examine the present values for projectiles at launch. For the cannonball, Re = 6,000,000; for the football, 440,000; for the basketball, 320,000; for the golf ball, 170,000. Clearly Reynolds number alone provides no easy key to the importance of drag. Nor does what we have called the drag tax depend in any direct fashion on launch speed. We will revisit the way the effect of drag on trajectories might be predicted in a few pages.

4. Where drag matters little for organisms . . .

First, though, we should examine existing data for biological projectiles, taking advantage of the computer to estimate launch speed from range and vice versa. Such data exist for a wide variety of systems – the present account will be selective rather than exhaustive. Together with those for the preceding cases, input data and results are summarized in table 1. Consider, to start with, a small, jumping mammal, a species of kangaroo rat (*Dipodomys spectabilis*) native to western North America and similar to the jerboa of North Africa and the marsupial kowari of Australia. It can be approximated as a sphere about 0·1 m in diameter with a density of about 750 kg m^{-3}. According to Biewener *et al* (1981), it can hop along bipedally at up to 3·1 m s^{-1} (11·2 km h^{-1}), which implies a launch speed (above the horizontal, of course) of about 3·1/sin 45º or 4·4 m s^{-1} (15·8 km h^{-1}). It achieves its best performance at a launch angle indistinguishable from 45º and incurs a drag tax of only about 1·1%. Why so little effect? Mainly, as we saw for the basketball, its decent size and thus fairly high mass together with its low launch speed and thus relatively low drag.

Among mammals that make haste with bipedal hopping, kangaroo rats are among the smallest. Simple consideration of surface-to-volume ratio – or, in effect, drag-to-gravity ratio – tells us that larger mammals will suffer even less from drag. So we anticipate that neither control of body posture, streamlining, nor altered piloerection will make much difference either to range, best launch angle, or speed. Where shape and postural changes do matter are among animals that glide, where lift-to-drag ratio plays a crucial role, and among animals that 'parachute', deliberately increasing drag to lower falling speeds.

Similarly, drag should not be a significant factor for any fair-sized animal that locomotes with a sequence of short ballistic trajectories – one that goes arm-over-arm, brachiating from hand-hold to hand-hold (see, for instance, Usherwood and Bertram 2003). Nor will it matter for those amphibians that throw their prehensile tongues forward as prey-capturing devices, despite their impressive performances – the tongue of *Bufo marinus*, a large toad, accelerates at over 30 times gravity to launch at nearly 3 m s^{-1} (10·8 km h^{-1}) (Nishikawa and Gans 1996); that of the salamander *Hydromantes imperialis* extends by 80% of its body length (Deban *et al* 1997). Nor does drag make a great difference for a yet odder practitioner of ballistics. At least one insect lineage shoots fecal pellets, apparently to minimize their potential predator-directing role (Weiss 2003). The pellets of a skipper caterpillar, *Calpodes ethlius*, average 2·8 mm in diameter and about 930 kg m^{-3} in density. After launch at 1·6 m s^{-1} (5·8 km h^{-1}) they go about 0·246 m, only 5·4% below their dragless range, achieved at 1º below the dragless 45º angle (input data from Caveney *et al* 1998). The pellets may be on the small side, but they do not go fast.

5. Smaller jumpers

As the size of jumpers drops, drag becomes increasingly important, as one can see from table 1. A desert locust (*Schistocerca gregaria*) can be approximated as a 10 mm sphere of 500 kg m^{-3} density; a launch speed to 3·0 m s^{-1} (10·8 km h^{-1}) takes it about 0·85 m downrange, 6·1% less than its dragless range (data from Bennet-Clark 1975). So it does only a bit worse than a skipper's pellet.

A particular froghopper or spittle bug (*Philaenus spumarius*), smaller than a locust (about 4 mm in diameter), takes advantage of a slightly faster launch, 4·0 m s^{-1} (14·4 km h^{-1}), to go farther, about 1·22 m, in the process, though, suffering a worse loss of range, 25·0% to drag and doing best at 42º (data from Burrows 2003).

A flea beetle (*Psylliodes affinis*), still smaller (about 1·6 mm), has a similar, if a bit lower, initial speed, 2·93 m s^{-1} (10·5 km h^{-1}); the latter takes it less far, 0·543 m, but with a worse drag tax, 37·9%. It gets its best range at a launch angle of 40º (data from Brackenbury and Wang 1995).

Fleas, smaller yet, encounter far greater trouble with drag. According to Bennet-Clark and Lucey (1967), a rabbit flea (*Spilopsyllus cuniculatus*) about 0·5 mm in diameter takes flight at 4·0 m s^{-1} (14·4 km h^{-1}). Drag reduces its range from 1·61 m to a mere 0·3 m, a loss of no less than 80·8%. And that best range (still assuming the game consists of long jumps across horizontal surfaces) happens with a launch angle of 30º. It lands at a speed no longer equal to launch speed but fully four times slower. (Bossard 2002 measured similar launch speeds for cat fleas.)

These insects launch at similar speeds; with smaller size their worlds become draggier and their trajectories less parabolic. Whatever their direction, they jump into the teeth of a sudden, severe windstorm. One has the sense

Table 1. Input data and simulation results for the various projectiles. Landing speeds assume launch at the angles that maximize horizontal range and equal launch and landing elevation.

Projectile	Effective diameter (mm)	Launch speed (m s^{-1})	Landing speed (m s^{-1})	Launch Reynolds number	Best launch angle	Maximum range (m)	Range loss from drag (%)
Cannonball	500·0	180·0	155·0	6,000,000	43º	3,000·0	14·1
Golf ball	42·9	60·0	42·0	170,000	41·5º	243·0	35·3
Basketball	240·5	20·0	18·0	320,000	43·5º	35·7	12·3
Football	220·0	30·0	23·0	440,000	43º	67·6	26·1
Kangaroo rat	100·0	4·4	4·3	29,000	44·5º	1·84	1·1
Skipper pellet	2·8	1·6	1·5	300	44º	0·246	5·4
Desert locust	10·0	3·0	2·8	2,000	44º	0·85	6·1
Froghopper	4·0	4·0	3·0	1,100	42º	1·22	25·0
Flea beetle	1·6	2·93	1·89	310	40º	0·543	37·9
Rabbit flea	0·5	4·0	1·0	130	30º	0·3	80·8
Hura seed	16·0	70·0	10·0	75,000	28º	30·0	94·0
Croton seed	3·5	8·5	5·6	2,000	41º	4·6	37·5
Vicia seed	2·7	9·0	4·8	1,600	38º	4·1	49·9
Ruellia seed	2·2	12·0	4·7	1,800	35º	4·9	66·5
Pilobolus sporangium	0·3	20·0	1·1	400	17º	0·82	98·0
Sordaria spores	0·04	30·0	0·05	80	7º	0·06	99·96
Gibberella spore	0·01	35·0	0·003	23	1º	0·0046	99·997

that fleas have explored the lower limit of jumping for practical animal locomotion.

6. Explosively launched seeds

Plants and fungi may lack equipment for continuous propulsion, but dispersal of their propagules must be as important as is travel for animals. They certainly have ways to give seeds and spores high-speed launches, ways that represent more biological diversity and span a greater range of sizes and initial speeds than in jumping insects. They also make much greater use (with, again, lots of diversity) of elevated launch sites. Still, the same physical imperatives apply. Drag gets relatively worse as size decreases, but so fast are the better among these projectiles that drag can be a major factor even for fairly large ones – much as we saw for golf balls.

Among large ballistic seeds, the current champion appears to be a tropical tree, sometimes planted as an ornamental, *Hura crepitans* (Swaine and Beer 1977; Swaine *et al* 1979). Its disk-shaped seeds (sometimes used as wheels for children's toys) are about 16 mm across and 350 kg m^{-3} in density. They launch with quite an audible pop at prodigious speeds, as high as 70 m s^{-1} (250 km h^{-1}). That speed (using horizontal range from ground level as benchmark gets a little artificial for a tree that grows to 60 m) can take them nearly 30 m. Impressive as that distance sounds, it is a small fraction of the 500 m that a *Hura* seed would go in a vacuum – range loss exceeds 94%. Curiously, this fastest speed known in the plant kingdom is indistinguishable from the maximum in the animals, the dive (largely passive and thus comparable) of a falcon (Tucker *et al* 1998).

Smaller seeds that lift off at more modest speeds fall into the same pattern we saw in jumping insects – the smaller the draggier. The 3·5 mm seeds of *Croton capitatus* (Euphorbiaceae), launched at 8·5 m s^{-1} (30·6 km h^{-1}) and 41°, go 4·6 m and pay a drag tax of 37·5%. The 2·7 mm seeds of *Vicia sativa* (Fabaceae), launched at 9 m s^{-1} (32 km h^{-1}) and 38°, go 4·1 m and pay a drag tax of 49·9% (Garrison *et al* 2000). The 2·2 mm seeds of *Ruellia brittoniana* (Acanthaceae), launched at 12 m s^{-1} (43 km h^{-1}) and 35°, go 4·9 m and pay a drag tax of 66·5% (Witztum and Schulgasser 1995).

Explosive seed expulsion occurs less often in still smaller seeds almost certainly because the increased surface-to-volume ratio will result in a further increase in relative drag, whatever the specific aerodynamic regime. Lurking behind the adaptational pattern are the inevitably conflicting demands of ballistic versus wind-borne travel – in effect drag minimization versus drag maximization. For ballistics large size, high density, and compact shape are preferable; for wind carriage small size, low density, and ramose shapes work better. (Stamp and Lucas 1983, among others, discuss such matters.)

7. Explosively launched spores

Small size, though, has proven less discouraging for ballistic spore dispersal by fungi. Most likely, the short stature of most fungi reduces their ability to put spores into the kinds of air movements particularly effective for passive travel. And with truly tiny propagules, even fairly dense spheres will have an agreeably low terminal velocity in free fall, making them better at staying up once aloft.

The most famous fungal projectile is the sporangium of the ascomycete, *Pilobolus*. *Pilobolus* erects its hypha (stalk) a few mm above piles of bovid and equid dung; the sporangium atop the hypha shoots off, with a bit of cell sap, at an initial speed of 20 m s^{-1} (72 km h^{-1}). A sporangium (of the density of water), 0·3 mm in diameter, should go 0·82 m at a best angle of 17°, paying a drag tax of 98%. In fact, sporangia go two or three times that far, almost certainly because they carry that cell sap. It adds mass without much increase in diameter, and it may even provide a slightly streamlined tail. Early in its travel, when going fastest and thus covering most territory, Reynolds numbers of up to 400 are high enough for such shaping to help.

That speed of 20 m s^{-1}, incidentally, comes neither from my back-calculation nor stroboscopic measurement. Long ago, A H Reginald Buller (1934) adopted a technique first used (as he says) by Napoleon's technicians when they measured bullet velocities. After firing through two rotating disks of paper they measured the offset of the second hole; that, together with the distance between the disks and their rotation rate, gave bullet speed. Buller used a perforated disk in front and an unperforated one behind, taking advantage of the sporangium's habit of sticking to whatever it hit.

Pilobolus, oddly, may make little use of wind. It fires shortly after dawn, not a windy time of day, taking aim at the sun, at that time low in the sky. Perhaps it aims to launch at about that 17° angle that maximizes windless range – no one, I think, has looked into the matter. The objective is a bit of grass or other forage far enough from its own pat of dung to be attractive to another grazer – completion of its life cycle requires passing through a herbivore's gut, and large herbivores (parasite-privy consumers) prefer not to graze too closely to what we used to call horse-apples and cow-pies.

A higher launch speed produces a much lower range in a still smaller projectile. Another ascomycete fungus, one once favoured by geneticists, *Sordaria*, shoots eight-spore clusters, about 40 μm across. Ingold and Hadland (1959) give it a typical range of about 60 mm, from which I calculate an initial speed of 30 m s^{-1} (108 km h^{-1}) and a drag tax of 99·96%. If horizontal range were the objective, its best launch angle would be a mere 7º – in fact, it seems to shoot upward. Why shoot at all? Further above a surface implies greater ambient air movement. With a terminal velocity below 50 mm s^{-1}, an upward shot would expose it to moving air for nearly a second, enough time for even the most modest wind to move it laterally.

Fungal guns come even smaller. Recently Frances Trail and Iffa Gaffoor measured a range of 4·6 mm for individually ejected ascospores of *Gibberella zeae*, a corn pathogen. Initial speed became important in justifying the high pressure used for launching and in identifying the responsible osmolytes, so I was drawn into the project. From spore density, about that of water, and size, about 10 μm, I calculated the remarkable launch speed of 35 m s^{-1} (125 km h^{-1}). That speed, without drag, would take a spore 125 m, so drag costs it no less than 99·997% of its potential range. It reaches its best range at a 1º launch angle; with a terminal velocity under 3 mm s^{-1}, it gets nearly as far (albeit briefly) from the launch site if shot vertically. With a vertical shot from the ground it will be exposed to moving air for about a second and a half and to far longer if launched from the surface of a plant. So that low terminal velocity should not limit its displacement. Rather, shooting will get it out beyond most or all of the low-speed air near the launch site. At this point, specific data exist for no ballistic projectiles smaller than *Gibberella* spores.

8. Generalizing

Why does a higher drag tax inevitably come with a lower optimal launch angle? With greater relative drag, descents are both steeper and slower than ascents. The rule for maximizing range in a draggy world comes down to getting one's distance while one still has decent speed – not wasting that fine launch speed going in a minimally useful direction, in particular, upward. The point becomes clear when one looks at some draggy trajectories that have been marked at uniform time intervals, as in figure 2.

Can aerodynamic lift be used to extend the range of a ballistic projectile? True airfoil-based gliding, used in many lineages of both animals and plants, requires a fairly specialized shape. Another possibility, though, consists of Magnus-effect lift – spinning or tumbling in such a way that the top of the projectile moves in the opposite direction of its overall flight and thus moves with (or at least less rapidly against) the oncoming air. The effect goes by various names in our various sports – slicing,

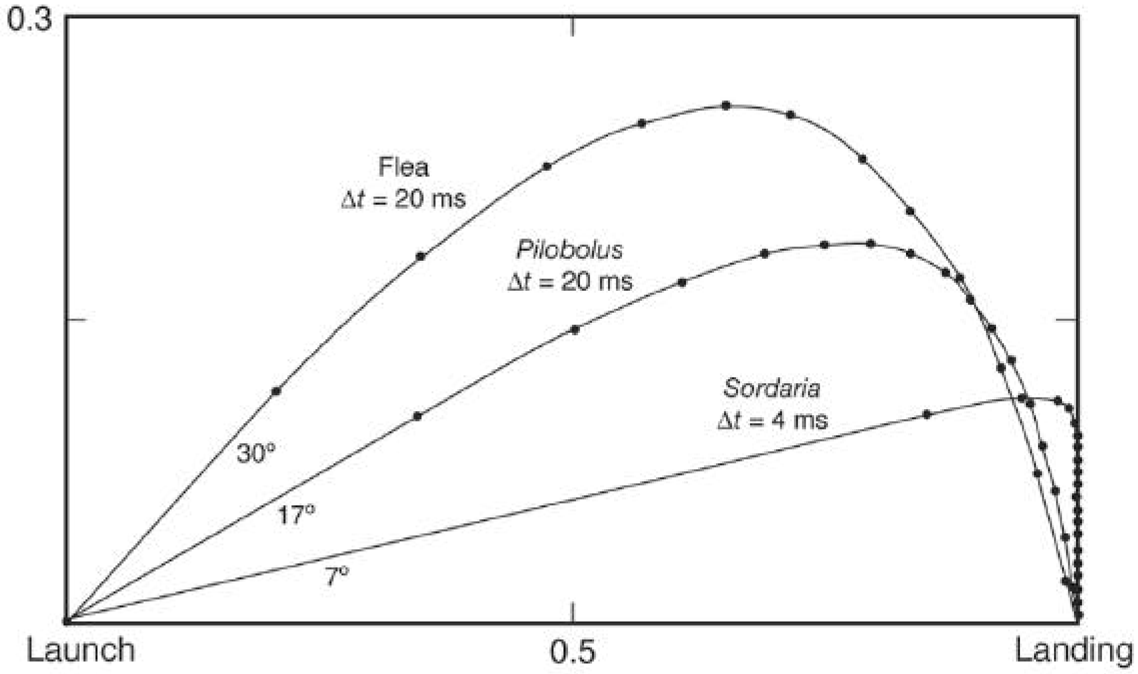

Figure 2. Maximum-range trajectories for three quite drag-afflicted projectiles, a jumping flea, a *Pilobolus* sporangium, and a *Sordaria* eight-spore cluster. Marked points on each curve give distances after equal time increments. Note that axes give horizontal distance as a fraction of maximum range and that the y-axis has been expanded two-fold relative to the x-axis.

top-spin, and so forth; but whatever the name and whether desired or counterproductive, it causes a projectile to deviate from the trajectory that gravity and drag would otherwise determine. But any gain will be small, probably less than 10% in a large seed or jumper and quite a lot less in small forms. Springtails, small flightless insects (order Collembola), appear to use the device (Brackenbury and Hunt 1993), spinning with their upward surfaces moving with the wind at about 16 revolutions per second. Other suggestive cases await investigation.

How might an organism project tiny propagules with less severe limitations than those experienced by *Sordaria* and *Gibberella*? Neither jet nor rocket propulsion occurs in aerial systems, but both the requirement for very high prelaunch accelerations and the disability imposed by drag can be ameliorated. A widely distributed moss, *Sphagnum*, may do so, although I have found no specific investigation of the matter. *Sphagnum* makes a nearly spherical capsule on a stalk well above its green gametophyte body. Prior to launch the walls of the capsule squeeze it down to a more cylindrical form; the increase in air pressure blows off its the lid and the spores go out in the blast of air (Ingold 1939). As a result of that brief tail-wind, they do not immediately encounter the full oncoming wind determined by their speed. And they go off in a cloud-like group. That should permit drag reduction by what in our vehicular world is called ‘drafting’ and which works far better in the very viscous regime of tiny particles – in effect, pooling mass and reducing effective surface area.

The present essay, like its predecessor (Vogel 2004) and, I hope, its successors, is intentionally eclectic, deliberately bringing together material made heterogenous by our traditional disciplinary divisions. Contriving effective comparisons all too often entails looking at how something performs under circumstances that may be adaptively irrelevant. Thus, as noted, *Pilobolus* may pick a launch angle that gives greatest range; *Sordaria* and *Gibberella* almost certainly do not, with both using ballistics in combination with wind dispersal. Seeds and skipper pellets land at lower heights than those from which they were launched; *Hura* trees, for instance, grow quite tall. So real best ranges and optimal angles require further input data and adjustments of the basic computer program. But beyond exposing underlying commonalities, bringing disparate material together can direct attention to gaps in understanding and to investigational opportunities.

9. Predicting and modelling

Finally, we might explore the utility of an index to the degree to which drag will alter (or fail to alter) a ballistic trajectory – a ‘range index’. I have found none that gives a precise prescription, in part because drag cannot be reduced to a simple proportionality. Still, one can produce an order-of-magnitude index with little difficulty from the ratio of the two forces that contribute to the form of a trajectory, gravity and drag. Gravitational force is proportional to *mg* or the product of density, length cubed, and gravitational acceleration. For drag we might use the product of pressure drag and viscous drag. Pressure drag is proportional to the product of density, speed squared, and length squared; viscous drag to the product of viscosity, speed, and length. Squaring gravitational force keeps the index dimensionless; taking the square root of the ratio of gravitational force, squared, to the product of the two forms of drag keeps values from getting unwieldy. Thus we have

$$RI = \left[\frac{r_p^{\,2}\, l_p^{\,3}\, g^2}{r_m m_m v_o^3}\right]^{1/2} = 2110\left[\frac{r_p^{\,2}\, l_p^{\,3}}{v_o^3}\right]^{1/2}, \qquad (7)$$

where subscript *m* refers to the medium, air, and subscript *p* to the projectile; v_o is launch speed. The version on the right includes the SI values of gravitational acceleration and the room temperature density and viscosity of air.

This range index suggests two things for biological projectiles, among which density varies by little more than a factor of two. First, high values, meaning minimal effects of drag, will characterize large objects travelling slowly – such as jumping mammals. Conversely, low values, meaning substantial drag effects, will occur with small objects going rapidly. Reynolds number, our usual index for the nature of a flow, includes the product of length and speed; this index uses their ratio. Second, for many fluid problems, what matters is the ratio of viscosity to density, the so-called kinematic viscosity (as in the Reynolds number). Air and water differ only about 15-fold. Here we have the product of viscosity and density; air and water differ by almost 50,000-fold. So shifting to water will cause the index to plunge, and buoyancy will decrease effective *g* as well. That rationalizes the scarcity of underwater ballistic devices in either nature or human technology. What about specific values of the index? Figure 3 plots index values for the cases discussed earlier against real range relative to dragless range. One sees that major effects of drag occur when projectiles are very fast (the *Hura* seed) or very small (the three fungal cases), although substantial effects (note the logarithmic scales) happen, as expected, for more ordinary items. Crude as it is, the range index may prove useful in anticipating the performance of yet other biological projectiles without recourse to a recondite computer program.

The range index also serves as a loose rule for making scale models. It enables a person to get a feel, through a

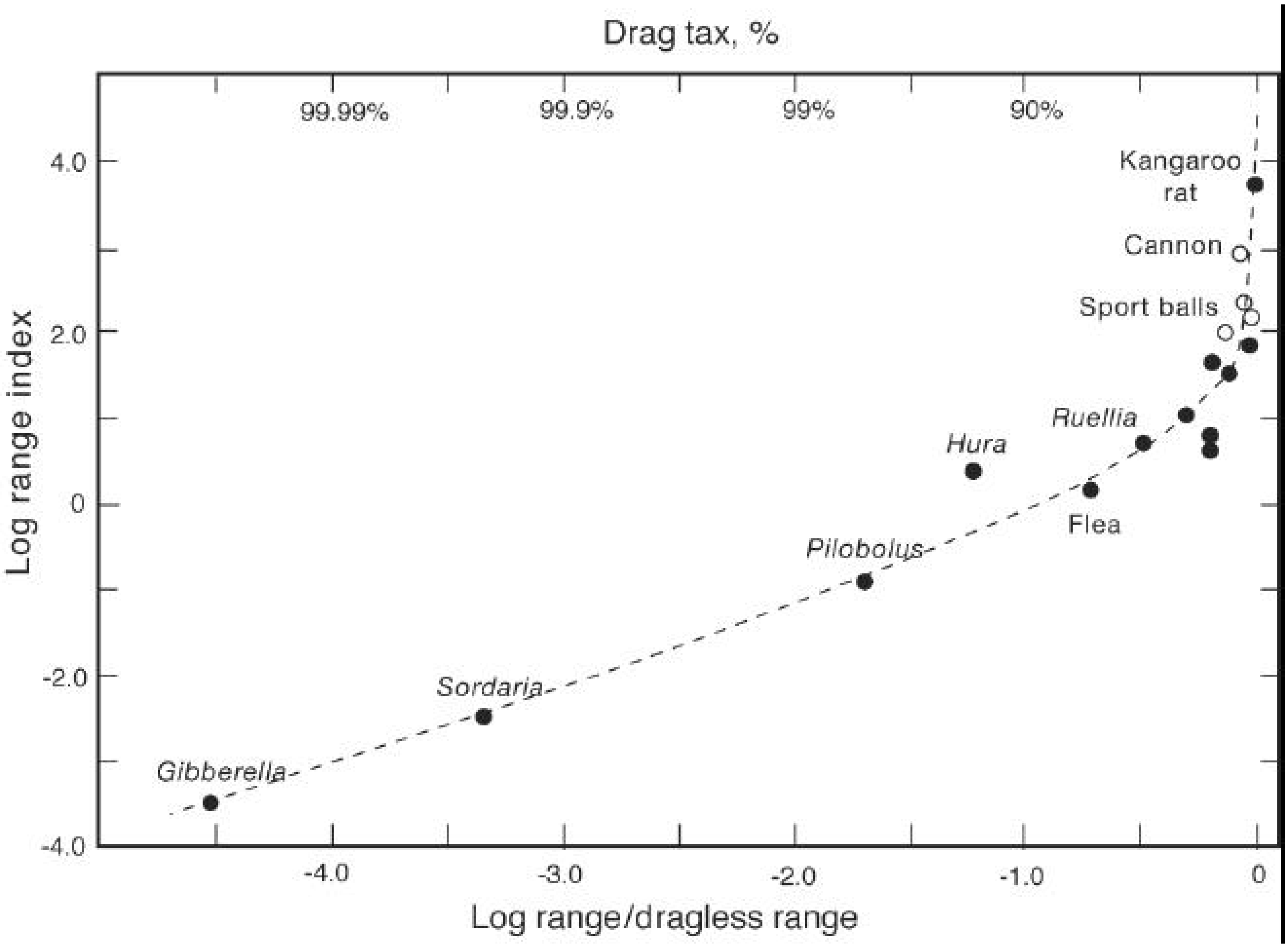

Figure 3. The relationship between the range index (equation 7) and real maximum range relative to dragless maximum range (bottom scale) and 'drag tax' (top scale). Open circles are projectiles used by humans.

bit of hands-on activity, for the world of very draggy projectiles. Just weigh a balloon, inflate it, and throw it as far as possible. Estimate launch speed from equation (2) and the range of a thrown projectile of minimal drag. From mass, size, and speed, you can then calculate a trajectory index. For a 150 mm, 0·66 g balloon, I got a value of 0·36, putting it between a jumping flea and a *Pilobolus* sporangium, and suggesting a range loss around 95% – about what happens when I throw the balloon.

Acknowledgements

I am indebted to Lewis Anderson, Frances Trail, Peter Wainwright, Martha Weiss, and the late Robert Page for introducing me to various of these biological projectiles and to Anne Moore and Molly McMullen for suggestions for clarifying the present presentation.

References

Bennet-Clark H C 1975 The energetics of the jump of the locust *Schistocerca gregaria*; *J. Exp. Biol.* **63** 53–83

Bennet-Clark H C and Lucey E C A 1967 The jump of the flea: a study of the energetics and a model of the mechanism; *J. Exp. Biol.* **47** 59–76

Biewener A, Alexander R McN and Heglund N C 1981 Elastic energy storage in the hopping of kangaroo rats (*Dipodomys spectabilis*); *J. Zool. London* **195** 369–383

Bossard R L 2002 Speed and Reynolds number of jumping cat fleas (Siphonaptera: Pulicidae); *J. Kans. Entom. Soc.* **75** 52–54

Brackenbury J and Hunt H 1993 Jumping in springtails: mechanism and dynamics; *J. Zool. London* **229** 217–236

Brackenbury J and Wang R 1995 Ballistics and visual targeting in flea-beetles (Alticinae); *J. Exp. Biol.* **198** 1931–1942

Brancazio P J 1984 *Sport science* (New York: Simon and Schuster)

Buller A H R 1934 *Researches on fungi* Volume VI (London: Longmans, Green)

Burrows M 2003 Froghopper insects leap to new heights; *Nature* (*London*) **424** 509

Caveney S, McLean H and Surry D 1998 Faecal firing in a skipper caterpillar is pressure-driven; *J. Exp. Biol.* **201** 121–133

Deban S M, Wake D B and Roth G 1997 Salamander with a ballistic tongue; *Nature* (*London*) **389** 27–28

Garrison W J, Miller G L and Raspet R 2000 Ballistic seed projection in two herbaceous species; *Am. J. Bot.* **87** 1257–1264

Ingold C T 1939 *Spore discharge in land plants* (Oxford, UK: Clarendon Press)

Ingold C T and Hadland S A 1959 The ballistics of *Sordaria*; *New Phytol.* **58** 46–57

Nishikawa K C and Gans C 1996 Mechanisms of tongue protraction and narial closure in the marine toad *Bufo marinus*; *J. Exp. Biol.* **199** 2511–2529

Stamp N E and Lucas J R 1983 Ecological correlates of explosive seed dispersal; *Oecologia Berlin* **59** 272–278

Swaine M D and Beer T 1977 Explosive seed dispersal in *Hura crepitans* L. (Euphorbiaceae); *New Phytol.* **78** 695–708

Swaine M D, Dakubu T and Beer T 1979 On the theory of explosively dispersed seeds: a correction; *New Phytol.* **82** 777–781

Tucker V A, Cade T J and Tucker A 1998 Diving speeds and angles of a gyrfalcon (*Falco rusticoles*); *J. Exp. Biol.* **201** 2061–2070

Usherwood J R and Bertram J E H 2003 Understanding brachiation: insight from a collisional perspective; *J. Exp. Biol.* **206** 1631–1642

Vogel S 1988 *Life's devices: The physical world of animals and plants* (Princeton, NJ: Princeton University Press)

Vogel S 2004 Living in a physical world I. Two ways to move material; *J. Biosci.* **29** 391–397

Weiss M R 2003 Good housekeeping: why do shelter-dwelling caterpillars fling their frass?; *Ecol. Lett.* **6** 361–370

White F M 1974 *Viscous fluid flow* (New York: McGraw-Hill)

Witztum A and Schulgasser K 1995 The mechanics of seed expulsion in Acanthaceae; *J. Theor. Biol.* **176** 531–542

*e*Publication: 21 February 2005

Living in a physical world
III. Getting up to speed

STEVEN VOGEL
Department of Biology, Duke University, Durham, NC 27708-0338, USA
(Fax, 919-660-7293; Email, svogel@duke.edu)

1. Introduction

Generalizations in biology come hard, so we treasure any that cut through life's overwhelming diversity. In his famous essay, "On Being the Right Size," J B S Haldane (1926) notes that jumping animals of whatever size should reach the same maximum height; Haldane attributes the insight to Galileo. Other iconic figures make the same assertion—Giovanni Alphonso Borelli (1680), grandfather of biomechanics; D'Arcy Thompson (1942), godfather of biomechanics; and then A V Hill (1938), father-figure for muscle physiologists.

The basic reasoning is straightforward. The force of a muscle varies with its cross-sectional area. The distance a muscle can shorten varies with its length. So the work a muscle can do will vary with the product of the two, in effect with its mass. All mammals have about the same mass of muscle relative to mass, about 45%, and other jumping animals differ only a little more. Thus the work available for a jump should be proportional to body mass. At the same time, the energy, mgh, required to achieve a given height, h, should also be proportional to body mass, m (gravitational acceleration, g, of course, stays constant).

Put in slightly different terms, launch speed, v_o, sets height for a projectile shot upward, and kinetic energy at launch is $^1/_2\ mv_0^2$. So the energy required to achieve a given launch velocity, like the work available, will be proportional to body mass. Either way, height should not depend on body mass.

As Borelli (1680), in the first great treatise on biomechanics, put it "…if the weight and mass of a dog is a fiftieth of those of a horse [ ] the motive force of the dog would be a fiftieth of that of the horse. Therefore, if the other conditions are equal [ ], the dog will jump as far as the horse." ('Force' for Borelli meant something close to what we recognize as work or energy.)

The last essay (Vogel 2005) focused on the behaviour of ballistic projectiles after launch. This one fleshes out the story by looking at what happens prior to launch, how projectiles of diverse sizes and functions reach their similar launch speeds.

2. The scaling of acceleration

What does constant jump height imply about prelaunch acceleration? The smaller the creature, the shorter the distance over which it can accelerate to that standard launch speed and the higher its acceleration. That, though, raises no apparent problem for muscle-driven launches. If force, F, scales with length squared (muscle cross section) and mass scales with length cubed (muscle volume), then by Newton's second law, $a = F/m$, acceleration, a, should scale inversely with body length: $a \propto l^{-1}$ – small jumpers should naturally achieve higher accelerations. Consider two adept mammalian jumpers. A lesser galago (or bushbaby) with a leg extension of about 0·16 m accelerates at 140 m s^{-2} while an antelope with a leg extension of 1·5 m accelerates at only 16 m s^{-2} (Bennet-Clark 1977). The comparison comes close to the predicted inverse proportionality – so far, so good.

How does acceleration scale when we look beyond such muscle-powered animal systems and include other projectiles such as those whose trajectories were examined in the previous essay? Table 1 compares body (or projectile, for non-jumpers) size with data or estimates for prelaunch acceleration. Bear in mind its limitations. (i) Its selection of systems makes no claim to be representative, although it does span the whole size range for which we have data. (ii) For want of any ready alternative, the entries assume steady prelaunch acceleration. (iii) Accelerations not reported in the literature have been calculated from launch speeds and estimates (from body proportions) of prelaunch travel distance. For mammals, with negligible drag, launch speeds come from jump heights; for smaller projectiles the computer program mentioned in the last essay was used to work back from reported ranges.

Before looking for relationships among these data, we might take note of the extreme accelerations of small projectiles. A *Pilobolus* sporangium, at 100,000 m s^{-2}, approaches the acceleration of a rifle bullet, typically 500,000 m s^{-2}. A *Gibberella* spore, at this point the biological record holder, accelerates at a truly cosmic 8,500,000 m s^{-2}. Indeed, the sheer diversity of organisms in the table makes us suspect that extraordinary accelerations might not require extraordinary engines.

Logarithmic graphs do lovely (and all too often misleading) service in suppressing scatter and the uncertainties introduced by rough estimates, especially where data span many orders of magnitude. Figure 1 gives such a log-log plot for the data just tabulated; a linear regression of the logarithms gives a slope of – 0·86. That scaling exponent comes reasonably close to a value of – 1·0 especially when one considers the diversity of both organisms and engines. And including rhinos, polar bears, and other unlikely leapers among the big mammals would have offset the inclusion of some underperforming arthropods and pushed the exponent still closer to – 1·0. Our forefathers have been vindicated – asserting that all creatures can jump to the same height implies a scaling relationship for acceleration quite close to what we find.

But that diversity ought to raise a flag of suspicion. Why should an argument based on muscle work for systems that do their work with other engines? Muscle represents no typical biological engine – it ranks at or near the top in, for instance, power-to-mass ratio.

Also, the scatter ought to be examined. Sub-par performance should raise few eyebrows, since fitness need not turn on personal ballistics. Less easy to rationalize are systems that do better than expected. One kind of seed, that of the tropical tree, *Hura crepitans*, clearly outguns all other ballistic seeds for which I have found data – it certainly deserves additional investigation. Seeds in general do better than arthropods of about the same size. A *t*-test supports that impression, yielding a significant difference between the size-acceleration products for the two groups (excluding *Hura*, the extreme outlier). Where sizes overlap, both arthropods and mammals do better than frogs. Despite Mark Twain's famous short story, frogs don't jump all that well – they're just sedentary creatures from which single long-jumps can be elicited easily.

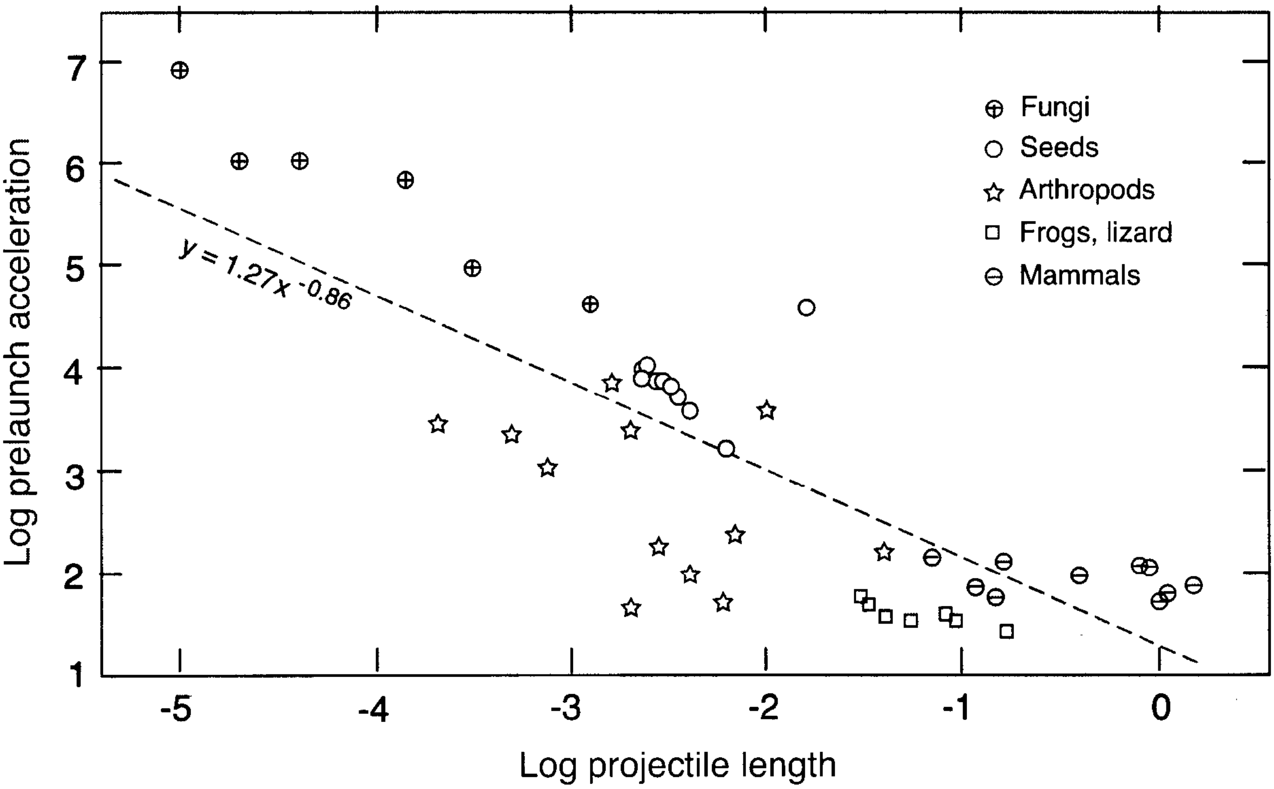

Figure 1. Projectile acceleration versus projectile size, with the linear regression line and equation for the data set. $r^2 = 0{\cdot}671$.

3. The limitations of real engines

But we cannot conclude that the near-inverse proportionality confirms our reasoning. Something else must be afoot – again, the original argument presumed isometric muscle-powered jumpers – and we get the odour of a more powerful basis for the scaling of projectile performance. The way the original rationale brushes aside the sometimes extreme effects of aerodynamic drag received attention in the preceding essay. In addition, it sweeps aside both a serious biological limitation and a major physical presumption. These will be our present concerns.

The biological limitation comes from way muscle performance scales – or fails to scale. The work a muscle can do, relative to its mass, depends little on its size or that of its animal. But consider jumpers impelled upwards by muscles that shorten as they jump – shortening, as one might say, in real time. An invariant launch speed demands that the muscles of the smaller animal do their work in a shorter time.

Skeletal muscle differs only a little either from muscle type to muscle type or with animal size in our broad-brush view. The resting length of the basic contractile unit, the sarcomere, is about 2·5 μm in vertebrates, and it varies by less than an order of magnitude elsewhere if we exclude a few odd extremes. Muscles consist of sarcomeres in series, so if all sarcomeres shorten at the same speed, then contraction speed should be directly proportional to muscle length. Or speed relative to length, called ‘intrinsic speed’ and given in units of reciprocal seconds, should not vary with body size. (‘Intrinsic speed’, not a speed in the strict length-per-time sense, equals minus the ‘strain rate’, as usually used in the engineering literature; the change of sign reflects a shift from shortening during normal muscle action to stretch during a tensile test.)

Muscle does not operate with equal effectiveness over a wide range of intrinsic speeds; an individual muscle does not operate with equal effectiveness over a wide range of actual speeds. A muscle pulls most forcefully (ignoring pulling during imposed extension) when not shortening at all, at zero speed. Force then drops off with speed until it hits zero at some maximum speed. Power, force times speed, peaks at about a third of that maximum speed. (McMahon 1984 gives a particularly good discussion of such matters.) In short, both force and power peak at speeds well below maximum.

Making a small animal power its jump by real-time muscle contraction forces its muscles to operate at high intrinsic speeds, speeds that either imply reduced effectiveness or cannot be reached at all. For example, consider two animals with launch speeds of 5 m s^{-1}. One is 1·2 m high and has to get up to launch speed in a third of its height, or 0·4 m. Working backwards from launch speed and distance gives an acceleration time of 0·16 s (and, incidentally, an acceleration of 31 m s^{-2}). It jumps, say, with muscles 0·3 m long that shorten by 20% of their length in the process, or 0·06 m. Thus its muscles shorten at a speed of 0·06 m/0·16 s or 0·375 m s^{-1}. Dividing by muscle length gives 1·25 s^{-1} as intrinsic speed, low enough to get a fine power output, perhaps a decent fraction of the 250 W kg^{-1} that approximates muscle’s practical maximum.

Contrast that with a similar animal a hundred times shorter, 12 mm. It must get up to speed in 1·6 ms (with an acceleration of 3100 m s^{-2}). Its 3-mm-muscles must shorten by 0·6 mm, thus at an identical speed of 0·375 m s^{-1}. Real-time muscle contraction of its shorter muscles takes a hundred-fold higher intrinsic speed, 125 s^{-1}, well beyond what vertebrate striated muscle can do. A mouse finger extender holds the upper record, 22 s^{-1}, but biological systems have difficulty getting reasonable (if suboptimal) outputs for power-demanding tasks above about 10 s^{-1}; where peak power matters, 5 s^{-1} is hard to exceed.

Therefore the old argument that all animals can jump to the same height cannot be correct if based on real-time muscle work – the physical presumption mentioned earlier. At best the rationale works above the body length at which necessary intrinsic speed becomes limiting. Jumping ability ought to drop off for animals less than 50 to 100 mm long. And judging by actual performances, even above that length size still seems to count. Cougar and kangaroo have muscles that yield more work per contraction relative to their masses and can jump higher than can jerboa and kangaroo rat. Drag is not the culprit (as noted in the last essay); they do indeed have higher launch speeds.

4. Amplifying power

Almost all the smaller jumpers evade this limitation on muscular performance by using power amplifiers to reach their necessarily higher accelerations. After all, conservation of work or energy does not imply conservation of power for non-sustained tasks. A system need only apply energy slowly and then release it rapidly – as done in archery.

A look at some large, muscle-powered weapons provides as good a direct comparison as I know between devices lacking and equipped with power amplifiers. Prior to the advent of cannon, Medieval Europe and Asia attacked fortifications with first one and then another version of a catapult, the two devices called, respectively, traction trebuchets and counterweight trebuchets (Hill 1973), as shown in figure 2. A traction trebuchet applied power in real time – the artillerymen pulled simultaneously downward on one arm, raising the arm and projectile-bearing sling on the other side of the fulcrum. A counterweight trebuchet stored energy gravitationally – artillerymen pulled downward on the arm with the sling and projectile, slowly raising a weight (of as much as

10,000 kg) on the other end. Releasing a catch on the lowered arm allowed the counterweight to plummet, raising that arm with its sling and projectile. Combining historical information with a few assumptions, the performances of representative trebuchets of the two kinds can be calculated (Vogel 2001).

A traction trebuchet could throw a 60 kg mass a distance of 90 m, implying a launch speed of 30 m s^{-1}. A human can pull downward for a distance of a meter with a force of 220 newtons, doing 220 joules of work per pull. Since the projectile needs $1/2 \times 60 \times 30^2$ or 27,000 joules per shot, at least 120 artillerymen had to pull – assuming massless arms and other unlikely idealizations. At Sind, now in Pakistan, in 708 CE, 500 people reportedly worked a single weapon – about 50 joules per operator per shot.

A counterweight trebuchet could throw a 225 kg mass a distance of 260 meters and thus with a launch speed of 50 m s^{-1} and an energy of 300,000 joules per shot. Without power amplification that would have demanded about 1400 artillerymen, again assuming perfect efficiency. As best we can tell, only about 50 were so employed – an effective output of not 50 but 6000 joules per operator per shot.

The anatomically simplest power amplifier uses antagonistic muscles to preload some elastic component. In that way a jump can be powered by the combined energy of direct muscular action and of elastic recoil of energy put in earlier. Most or all of the vertebrates – and certainly all the smaller ones – listed in table 1 augment direct, real-time muscular action with some preloading of elastic components. The main storage sites are the tendons in series with the jumping muscles themselves. These complex muscle-tendon systems have yet to be fully analysed, but they appear to involve initial crouching countermovements and some kind of catapult mechanism – at the least (Alexander 1988; Aerts 1998).

Calculations of the power outputs of jumping muscles often give values well above what isolated muscle can do – which, in the absence of more specialized devices, points to such preloading. A tree frog, *Osteopilus,* for instance, achieves a peak output in a jump about seven times higher that the maximum output of the muscles it uses (Peplowski and Marsh 1997); it may be extreme, but its power booster is not unique among frogs. With such amplification, frogs keep their intrinsic speeds fairly low, below about 5 s^{-1} (Marsh 1994). A jumping bushbaby (*Galago senegalensis*) does something similar; Aerts (1998) calculated that for direct action of leg extensors to power its jumps, those muscles would need to weigh twice that of the entire body. Frogs and bushbabies look like good jumpers, with big, specialized hind limbs. But little obvious structure underlies their power amplifiers – they seem to do their tricks the way an eager (and abusive) automobile driver races the engine before engaging the clutch or automatic transmission.

Power amplification appears almost universal among arthropods – only one group of jumpers clearly lack any specialized device. Despite the name, the salticid spiders do not jump especially well, at least by the criteria of launch speed and acceleration. Parry and Brown (1959) looked hard but found no amplifier. Spiders, though, are something of a law unto themselves, since they extend their legs with hydraulics rather than by direct muscular action.

Jumping with real-time muscle action or with simple preloading has its limits. Figure 3 fleshes out the picture presented by figure 1. A horizontal line seems to mark an upper limit for operating as do frogs, lizards, mammals, and salticid spiders. Nature, it appears, does not use real-time muscle action, even in mildly augmented form, as the main impetus for accelerations above about 150 m s^{-2}, whatever the size of creature or projectile. So we are

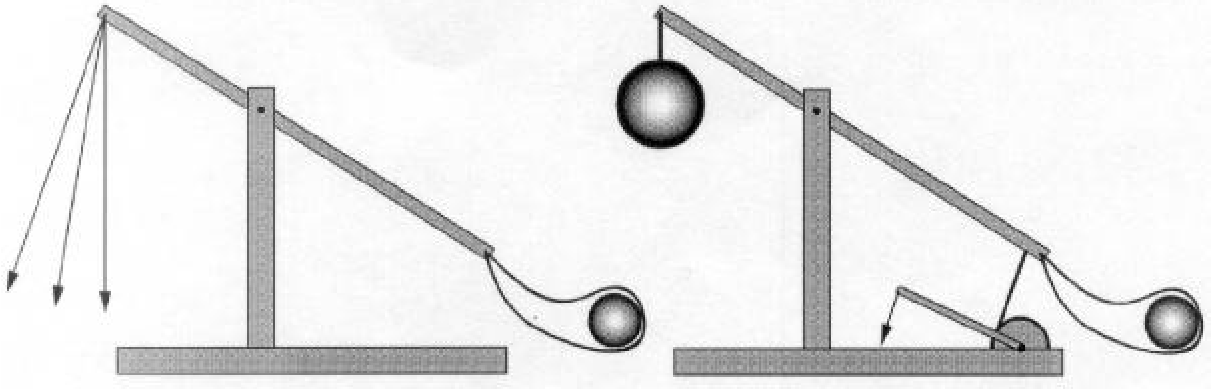

Figure 2. Diagrammatic versions of two large types of artillery powered by human muscle; both throw their projectiles from slings. The traction trebuchet, on the left, stored no energy except perhaps as kinetic energy of the moving arm; the counterweight trebuchet, on the right, made heavy use of gravitational storage.

left with the regression line in the figure with a slope (– 0·86) close to (indeed, statistically indistinguishable from) the value predicted (– 1·0) on the basis of an argument now revealed as specious, even for muscle-powered jumpers.

5. How to really amplify power

So far we have focused on muscle as engine. Again, both muscle's power relative to weight and speed relative to size put it near the high end among living engines. Direct or largely direct powering of ballistic launches may be possible for muscular systems, at least for large ones. By contrast, where other engines drive launches, power amplification must be an absolute necessity for any kind of ballistic travel.

What, then, are the options for serious amplification? Linking a slow input with a rapid output requires a way to store energy. Our human technology employs such things as flywheels and rechargeable chemo-electric bat-

Table 1. Projectile sizes and estimates of prelaunch accelerations for biological projectiles.

Projectile		Length	Acceleration	Source
Gibberella zeae spore	f	0·00001	8,500,000	Trail *et al* (2005)
Sordaria fimicola, spore	f	0·00002	1,100,000	Ingold and Hadland (1959)
Sordaria, 8-spore cluster	f	0·00004	1,100,000	Ingold and Hadland (1959)
Ascobolus immersus, spores	f	0·00015	630,000	Fischer *et al* (2004)
Moss mite (*Zetorchestes*)	a	0·0002	3,400	Krisper (1990)
Pilobolus sporangium	f	0·0003	100,000	Buller (1909)
Rat flea	a	0·0005	2,000	Bennet-Clark and Lucey (1967)
Box moss mite (*Indotritia*)	a	0·0008	1,200	Wauthy *et al* (1998)
Sphaerobolus glebal mass	f	0·0012	46,000	Buller (1933)
Geranium molle	s	0·0016	8,100	Stamp and Lucas (1983)
Flea beetle (*Psylliodes*)	a	0·002	2,660	Brackenbury and Wang (1995)
Springtail	a	0·002	47	Brackenbury and Hunt (1993)
Geranium carolinarium	s	0·002	10,300	Stamp and Lucas (1983)
Viola striata	s	0·0021	7,800	Stamp and Lucas (1983)
Ruellia brittoniana	s	0·0023	10,000	Witztum and Schulgasser (1995)
Vicia sativa	s	0·0027	7,500	Garrison *et al* (2000)
Skipper butterfly frass	a	0·0028	180	Caveney *et al* (1998)
Geranium maculatum	s	0·0029	7,600	Stamp and Lucas (1983)
Croton capitatus	s	0·0035	5,200	Garrison *et al* (2000)
Froghopper	a	0·004	4,000	Burrows (2003)
Flea beetle (*Altica*)	a	0·004	100	Brackenbury and Wang (1995)
Impatiens capensis	s	0·0051	1,650	Stamp and Lucas (1983)
Salticid spider	a	0·006	51	Parry and Brown (1959)
Desert locust, 1st instar	a	0·007	200	Katz and Gosline (1993)
Click beetle	a	0·010	3,800	Evans (1972)
Hura crepitans	s	0·016	41,000	Swain and Beer (1979)
Acris gryllus	h	0·027	64	Marsh and John-Alder (1994)
Pseudacris crucifer	h	0·029	58	Marsh and John-Alder (1994)
Desert locust adult	a	0·040	160	Katz and Gosline (1993)
Hyla squirella	h	0·044	29	Marsh and John-Alder (1994)
Hyla cinerea	h	0·056	26	Marsh and John-Alder (1994)
Jumping mouse	m	0·07	143	Nowak (1991)
Anolis carolinensis	h	0·07	45	Toro *et al* (2003)
Osteopilus septentrionalis	h	0·088	26	Marsh and John-Alder (1994)
Jerboa, kowari, kangaroo rat	m	0·12	75	Nowak (1991)
Red squirrel	m	0·15	60	Essner (2002)
Lesser galago	m	0·16	140	Bennet-Clark (1977)
Rana catesbiana	h	0·164	20	Marsh (1994)
Potoroo	m	0·4	100	Nowak (1991)
Springbok	m	0·8	125	Nowak (1991)
Impala	m	1·0	100	(various)
Cougar (mountain lion)	m	1·0	55	Nowak (1991)
Gray kangaroo	m	1·1	67	Nowak (1991)
Horse, eland	m	1·7	80	Nowak (1991)

Lengths in meters; accelerations in m s^{-2}. To convert the latter to multiples of gravitational acceleration ("*g*'s"), divide by 10. (f, fungus; s, seed; a, arthropod; h, frog or lizard; m, mammal.)

teries, schemes with only distant analogs in nature. Both human and natural technologies use gravitational storage – from counterweight trebuchets to pendulums in the former; stride-to-stride energy storage in legged walkers in the latter. One can imagine trees that toss fruit from wind-driven branches that sway as gravitational pendulums or seeds propelled by the drop of an elevated column of liquid, but I know no specific case of gravitational storage in biological ballistics. As far as I know, all prelaunch amplifiers depend on the same scheme, energy storage in deformed elastic materials.

Remind yourself of the simplicity and ubiquity of power amplification through brief elastic energy storage by flipping the nearest toggle switch, one that controls the room lights or some piece of household electronics. With most such switches, one pushes a lever with increasing force until it abruptly stops opposing your effort and switches to its alternative position. You have slowly loaded a spring, which then rapidly releases that energy to make a sudden and robust change of electrical contact. You may continue to push, but you do little additional work once the spring has shifted from absorbing to releasing energy.

A single-shot amplifier, as in most ballistic plants and fungi, can be self-destructive and thus even simpler than a spring-assisted switch. The fungus *Pilobolus*, for instance, bears its sporangium atop a liquid-filled hyphal tube, as in figure 4a. An osmotic engine raises the pressure in the tube until the sporangium suddenly detaches along a specific junctional line and takes flight (Buller 1909). That commonest of fungal schemes gets tweaked by ones such as *Sordaria* that manage to avoid self-destruction long enough to loose a series of up to eight spores in quick succession (Ingold and Hadland 1959). Another fungus, *Sphaerobolus*, uses a one-shot catapult in which an initially concave cup ('peridium') containing a millimeter-wide glebal mass of spores suddenly everts, becoming convex upward (Buller 1933, figure 4b). A similar bistable system has recently been described by Forterre *et al* (2005) in a higher plant, the Venus's flytrap. In both fungi and flytrap, the ultimate engine is osmotic, coupled hydraulically with the output device.

Many seed shooters use another single-shot system, one in which drying of an initially hydrated structure such as a seed pod gradually stresses some woody (cellulosic) material. The movement accompanying breakage then sends the seed (or a group of seeds) onward. In *Ruellia*, for instance (figure 4c) sudden lengthwise rupture of the seam between two external valves (each analogous to a half-shell of a bivalve mollusk) lets the valves bend

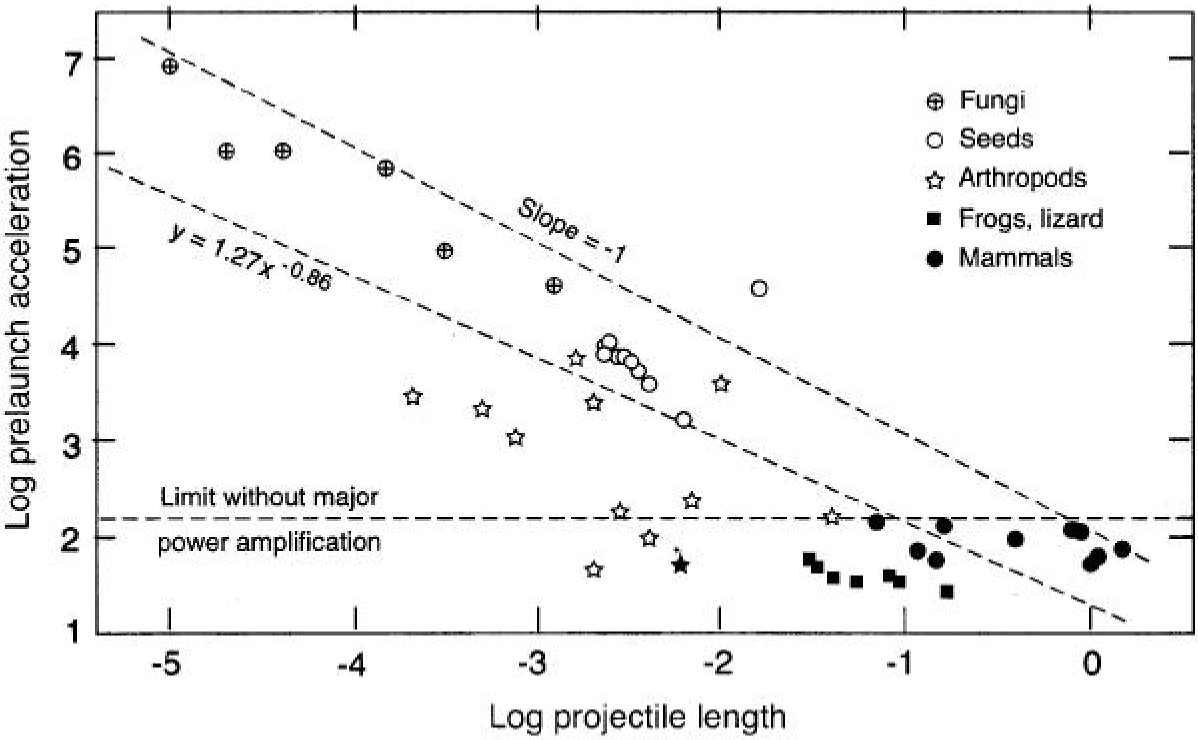

Figure 3. The same data and graphing conventions as in figure 1, but now showing which organisms make major use of elastic energy storage (solid symbols) and with an approximate performance limit line drawn on.

outward. That causes arms attached to the insides of valves to bend upward, whereupon each arm pushes a seed up and out (Witztum and Schulgasser 1995).

Among repetitive amplifiers, that of locusts and grasshoppers is especially straightforward. According to Bennet-Clark (1975), rather than directly powering a jump, the large extensor muscle of each hind tibia loads a pair of elastic elements. A catch near the junction of the (proximal) femur and (distal) tibia keeps the leg flexed – a jump must start with fully flexed tibias. Relaxation of the flexor muscle releases the catch, and the immediate power for the jump comes mainly from energy stored in the chitin of elastic cuticle. The peak power output of 0·75 watts from 70 mg of muscle represents almost 11,000 W kg^{-1}, around 40 times what muscle can do directly. Moreover, amplification permits locust jumping muscle to operate at an efficiently low intrinsic speed, peaking at less than 2 s^{-1}.

Fleas (Bennet-Clark and Lucey 1967) have a more intricate mechanism, about which I will give even less detail. A rabbit flea requires about 100 ms for the large trochanteral depressor muscles of its hind legs to deform a pair of elastic pads, here not chitin but the softer and spectacularly resilient protein resilin. A second pair of muscles trigger energy release by moving a strap sideways, undoing the catch. The jump itself lasts only 0·7 ms, quicker by nearly 150 times. Energy storage permits the muscles to operate at a very forceful intrinsic speed of 0·55 s^{-1} rather than an impossible 50 s^{-1} or more.

6. Storing energy elastically

Table 2 compares the key properties of several of the materials available for elastic energy storage, with spring steel included for reference. Of course ancestry constrains the choice of energy storage material. Thus only arthropods make resilin, and cellulose mainly occurs in plants. And the storage materials of most projectile-producers represent only mild modifications of those of non-projectile-producing forebears. For resilience, work regained relative to work put in, resilin, known best from insect wing hinges, beats any other biological material. It may have to be superb, not because a few percent gain in resilience matters much to fitness, but because the loss relative to perfect resilience (1·0) appears as heat, something not well tolerated in insect wing hinges, where it may be alternately stressed and released hundreds of times each second. Fleas just happen to be in an auspicious lineage.

Tendon is mainly collagen, our main elastic energy store (the protein elastin plays a lesser role); it also does

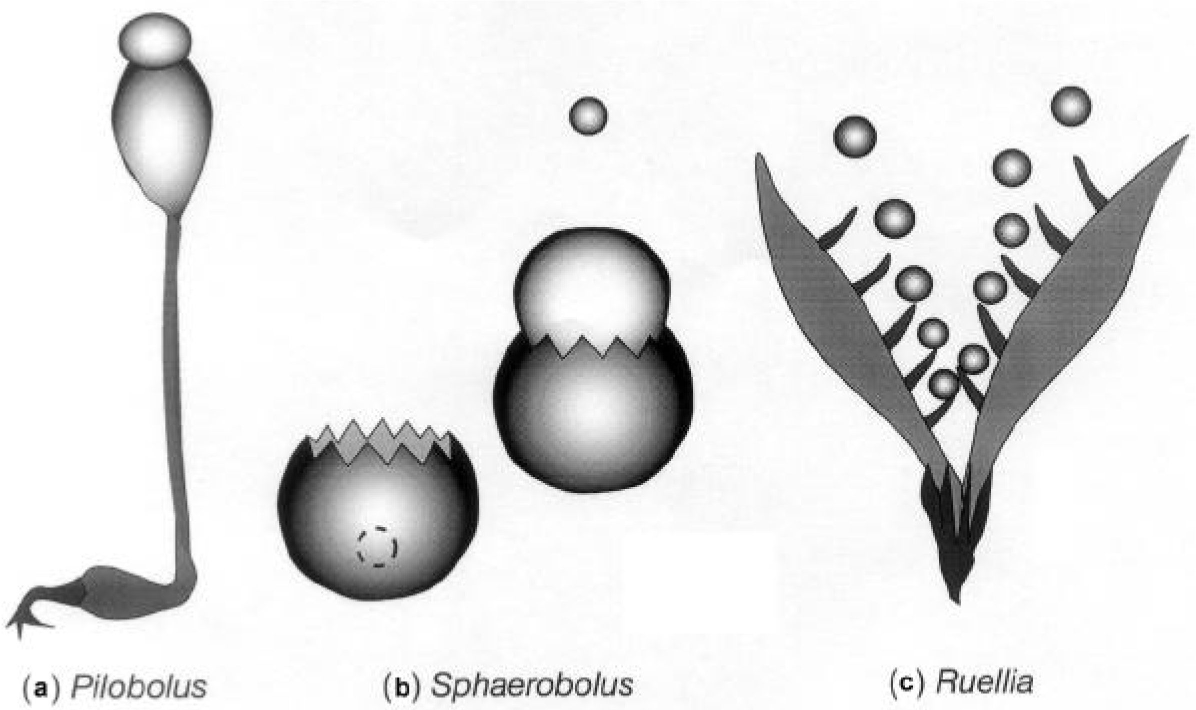

Figure 4. The diverse launching devices of three ballistic organisms. The fungus *Pilobolus*, is shown with the sporangium on top of the subsporangial swelling just before it shoots upward on a jet of cell sap. *Sphaerobolus* appears just before and just after a glebal mass of spores gets sent aloft by eversion of the floor of the cup. The seed *Ruellia* has been caught just before the end of launch, with each seed propelled upward by motion of the ejaculator beneath it.

well but has quite a different character and has to be used differently. Resilin, softer (elastic modulus 1·9 MN m^{-2}), can be usefully loaded in compression or shear. Collagen, stiffer (modulus 1,500 MN m^{-2}) works best in tensile applications, and even there it cannot be arranged like a conventional rubber – it operates at high forces and low extensions, crucial when linking muscles to bones.

Wood, despite our long use of wooden bows for archery, has poor resilience – we take advantage of its internal damping to get mellifluous resonators in many musical instruments. But woods vary widely (as instrument makers have long known), and fresh wood can be quite different from dry stuff. We know very little about the storage capabilities and resilience of woods (and other cellulosic materials) as used for energy accumulation and release in nature's ballistic devices. On the one hand, ballistic seeds do impressively well, suggesting high resilience. On the other hand, the structures involved represent a tiny fraction of the total mass of a plant and probably an even smaller fraction of its lifetime energy expenditure, so efficiency relative to either mass or work might not matter. We know still less about the properties of the materials that fungi use for energy storage.

Air makes a perfectly fine elastic material, taking loads in either compression (almost without limit) or tension (at least to 101,000 N m^{-2}, one atmosphere). I know of no case of its use in any ballistic system in nature, although I suspect that a moss, *Sphagnum*, might store energy for spore ejection by compressing air, as noted in the last essay. Water, by its ubiquity and the data in the table, looks attractive; but that proves illusory. Its extremely low compressibility (or high bulk modulus, the same thing) produces an awkward mismatch with biological solids. Squeeze water in a container of any such solid and the container stretches more than the water compresses – water requires operation at extremely high force with extremely low volume change. So, while water makes a superb medium for transmitting hydrostatic pressures, it turns out to be next to useless for storing energy.

The final column of the table 2 gives a severely idealized calculation of the minimum mass of elastic material relative to the weight of the projectile. In essence, it equates the initial kinetic energy of the projectile with the product of (i) the work (energy) of extension relative to mass of the elastic, (ii) the resilience of the elastic, and (iii) the mass of the elastic. The assumed launch speed of 5 m s^{-1} corresponds to a dragless vertical ascent of 1·25 m – a typical value for the present systems. On this perhaps biased basis, the biological materials look remarkably good.

7. What does limit acceleration and launch speed?

The old argument has been shredded. The work relative to mass of a contracting muscle deteriorates as animals get smaller rather than holding constant – a consequence of the requisite rise in intrinsic speed. Muscle need not and commonly does not power jumps in real time –elastic energy storage in tendons of collagen, in apodemes of chitin, and in pads of resilin provides power amplification. Finally, muscle powers none of those seeds and tiny fungal projectiles. Yet acceleration persists in scaling as the classic argument anticipates.

A look at the properties of elastic materials dispels any notion that their ability to store energy imposes a particular limit. Even the extreme case, launching a *Hura* seed with the energy of stretched or squeezed wood, would take an elastic mass only 5 or so times the mass of the projectile. That volume of elastic should be no problem, at least for shooters rather than jumpers. Distance (and thus speed) amplifying levers can compensate for inadequate speed of recoil of an elastic. And nature could probably enlarge muscular systems or run osmotic engines at higher pressures (although Alexander 2000 gives an argument against the first of these).

A possible alternative emerges from reexamination of the relationship between force and acceleration. If acceleration indeed scales inversely with length and mass di-

Table 2. The relevant properties of materials for brief elastic energy storage and release (Bennet-Clark 1975; Gosline *et al* 2002; Jensen and Weis-Fogh 1962; Vogel 2003). The numbers presume an uncomfortably large number of assumptions about such things as operating conditions and ignore large elements of biological variability.

Material	Energy/volume	Energy/mass	Resilience	Relative elastic mass
Arthropod cuticle	9·6 MJ m^{-3}	8,000 J kg^{-1}	~ 0·8	~ 0·2%
Tendon (collagen)	2·8	2,500	0·93	0·54
Wood	0·5	900	~ 0·5	~ 2·3
Resilin	1·5	1,250	0·96	1·04
Spring steel	1·0	150	0·99	8·42
Air	0·000500	417	1·00	0·75
Liquid water	0·18	180	1·00	6·94

rectly with the cube of length, then force should scale with the square of length. Or, put another way, force divided by the square of length should remain constant. Force over the square of length corresponds to stress. Perhaps our empirical finding that acceleration varies inversely with length tells us that stress in some manner limits these systems. A stress limit would represent no great biological novelty, having been recognized (or invoked) in remodelling of bone, resizing of blood vessels, and the growth of trees (see, for references, Vogel 2003).

The stress limit may go well beyond the maximum pull of a muscle. It might reflect a point of self-destruction, a limit that the propulsive equipment of a system might exceed only at risk, one might say, to life and limb. That could apply even to the largest jumpers, since experimental work on humans – anticipating rocket launches – shows that our bodies do not take kindly to accelerations much above those experienced by large mammalian jumpers. It also rationalizes the greater accelerations of seeds than of arthropods – seeds, simpler and sturdier, should be less easily damaged by high launch accelerations. I have to pick up small insects carefully lest I damage them; seeds I grind in mortar with pestle. (Of course seeds are not self-propelled, the basis of an alternative explanation.)

Figure 3 has one further line, a line with a slope of – 1 over its more than five orders of magnitude of size. It has been drawn so it roughly follows the extremes of acceleration (that eccentric seed of *Hura* is again an outlier). Maybe that line is the important one, a practical constraint imposed by the materials and structures of biological projectiles that must not be rendered dysfunctional by their ballistic episodes – these are whole animals and propagules, not bullets. That limit line, reflecting the scaling of force with the square of length, might be pointing to the size-independence of maximum stress tolerable by biological materials. It is consistent with (and may reduce to an example of) a more general scaling rule. Marden and Allen (2002) found just such scaling in the force output of a wide range of engines, ranging from molecular motors of myosin, kinesin, dynein, and RNA polymerase, through muscles to winches and rockets – their 'group 1 motors' – and attribute it to a common limit on just this capacity to withstand mechanical stress.

Several final notes: The present essay, following its predecessor, has focused on projectiles. Other biological systems achieve high accelerations, and these accelerations also vary inversely with size, despite their diversity of propulsive engines. So, for completeness, I ought to mention the ejectable nematocysts of the coelenterates, the retractable spasmoneme of vorticellid protozoa, and the protrusible tongues of many amphibians and reptiles.

The homogeneity of the seeds (one again omitting *Hura* and thus emphasizing its aberrant character) relative to the other groups comes as a surprise. These ballistic seeds span a notably narrow size range, with lengths ranging from just under 2 mm to just over 5 mm, and their accelerations vary only slightly more. Other explosively discharged seeds, such as those studied by Stamp and Lucas (1990) appear to fit into the same cluster. One suspects some as yet unidentified constraint.

And then we return to that assertion about all animals jumping to the same height. J B S Haldane attributed it to Galileo; I believe he erred. I can find no such assertion or anything closer to it than his comment on the scaling of bones. I confirm D'Arcy Thompson's attribution to Borelli, down to chapter and verse. Borelli was only translated into English long after Thompson wrote *On Growth and Form*; but, as an accomplished classical scholar, Thompson would have read Borelli in the original Latin.

Acknowledgements

I thank David Alexander, Peter Klopfer, Daniel Livingstone, Michael Reedy, and Frances Trail for help identifying sources, collecting material, and clarifying both my ideas and the presentation.

References

Aerts P 1998 Vertical jumping in *Galago senegalensis*: the quest for an obligate mechanical power amplifier; *Philos. Trans. R. Soc. London* **B353** 1607–1620

Alexander R McN 1988 *Elastic mechanisms in animal movement* (Cambridge: Cambridge University Press)

Alexander R McN 2000 Hovering and jumping: contrasting problems in scaling; in *Scaling in biology* (eds) J A Brown and G B West (Oxford: Oxford University Press) pp 37–50

Bennet-Clark H C 1975 The energetics of the jump of the locust *Schistocerca gregaria*; *J. Exp. Biol.* **63** 53–83

Bennet-Clark H C 1977 Scale effects in jumping animals; in *Scale effects in animal locomotion* (ed.) T J Pedley (London: Academic Press) pp 185–201

Bennet-Clark H C and Lucey E C A 1967 The jump of the flea: a study of the energetics and a model of the mechanism; *J. Exp. Biol.* **47** 59–76

Borelli G A 1680 *De motu animalium*; P Maquet, trans 1989 as *On the movement of animals* (Berlin: Springer-Verlag)

Brackenbury J and Hunt H 1993 Jumping in springtails: mechanism and dynamics; *J. Zool. London* **229** 217–236

Brackenbury J and Wang R 1995 Ballistics and visual targeting in flea-beetles (Alticinae); *J. Exp. Biol.* **198** 1931–1942

Buller A H R 1909 *Researches on fungi* vol. 1 (London: Longmans, Green)

Buller A H R 1933 *Researches on fungi* vol. 5 (London: Longmans, Green)

Burrows M 2003 Froghopper insects leap to new heights; *Nature* (*London*) **424** 509

Caveney S, McLean H and Surry D 1998 Faecal firing in a skipper caterpillar is pressure-driven; *J. Exp. Biol.* **201** 121–133

Essner R L 2002 Three-dimensional launch kinematics in leaping, parachuting and gliding squirrels*; J. Exp. Biol.* **205** 2469–2477

Evans M E G 1972 The jump of the click beetle (Coleoptera, Elateridae): a preliminary study; *J. Zool. London* **167** 319–336

Fischer M, Cox J, Davis D J, Wagner A, Taylor R, Huerta A J and Money N P 2004 New information on the mechanism of forcible ascospore discharge from *Ascobolus immersus*; *Fungal Genet. Biol.* **41** 698–707

Forterre Y, Skotheim J M, Dumais J and Mahadevan L 2005 How the Venus flytrap snaps; *Nature* (*London*) **433** 421–425

Garrison W J, Miller G L and Raspet R 2000 Ballistic seed projection in two herbaceous species; *Am. J. Bot.* **87** 1257–1264

Gosline J, Lillie M, Carrington E, Guerette P, Ortlepp C and Savage K 2002 Elastic proteins: biological roles and mechanical properties; *Philos Trans. R. Soc. London* **B357** 121–132

Haldane J B S 1926 On being the right size; *Harper's Monthly* **152** 424–427

Hill A V 1938 The heat of shortening and dynamic constants of muscle; *Proc. R. Soc. London* **B126** 136–195

Hill D R 1973 Trebuchets; *Viator* **4** 99–114

Ingold C T and Hadland S A 1959 The ballistics of *Sordaria*; *New Phytol.* **58** 46–57

Jensen M and Weis Fogh T 1962 Biology and physics of locust flight. V. Strength and elasticity of locust cuticle; *Philos. Trans. R. Soc. London* **B245** 137–169

Katz S L and Gosline J M 1993 Ontogenetic scaling of jump performance in the African desert locust (*Schistocerca gregaria*); *J. Exp. Biol.* **177** 81–111

Krisper G 1990 Das Sprungvermögen der mitbengattung *Zetorchestes* (Acarida, Oribatida); *Zool. Jb. Anat.* **120** 289–312

Marden J H and Allen L R 2002 Universal performance characteristics of motors; *Proc. Natl. Acad. Sci. USA* **99** 4161–4166

Marsh R L 1994 Jumping ability of anuran amphibians; *Adv. Vet. Sci.* **B38** 51–111

Marsh R L and John-Alder H B 1994 Jumping performance of hylid frogs measured with high-speed cine film; *J. Exp. Biol.* **188** 131–141

McMahon T A 1984 *Muscles, reflexes, and locomotion* (Princeton: Princeton University Press)

Nowak R M 1991 *Walker's mammals of the world* 5th edition (Baltimore: Johns Hopkins University Press)

Parry D A and Brown R H J 1959 The jumping mechanism of salticid spiders; *J. Exp. Biol.* **36** 654–664

Peplowski M M and Marsh R L 1997 Work and power output in the hindlimb muscles of Cuban treefrogs *Osteopilus septentrionalis* during jumps*; J. Exp. Biol.* **200** 2861–2870

Stamp N E and Lucas J R 1983 Ecological correlates of explosive seed dispersal; *Oecologia* (*Berlin*) **59** 272–278

Stamp N E and Lucas J R 1990 Spatial patterns and dispersal distances of explosively dispersing plants in Florida sandhill vegetation; *J. Ecol.* **78** 589–600

Swaine M D and Beer T 1977 Explosive seed dispersal in *Hura crepitans* L. (Euphorbiaceae); *New Phytol.* **78** 695–708

Thompson D'A W 1942 *On growth and form* (Cambridge: Cambridge University Press)

Toro E, Herrel A, Vanhooydonck B and Irschick D J 2003 A biomechanical analysis of intra- and interspecific scaling of jumping and morphology in Caribbean lizards; *J. Exp. Biol.* **206** 2641–2652

Trail F, Gaffoor I and Vogel S 2005 Ejection mechanics and trajectory of the ascospores of *Gibberella zeae* (anamorph *Fusarium graminearum*); *Fungal Genet. Biol.* (in press)

Vogel S 2001 *Prime mover*: *a natural history of muscle* (New York: W W Norton)

Vogel S 2003 *Comparative biomechanics: life's physical world* (Princeton: Princeton University Press)

Vogel S 2005 Living in a physical world II. The bio-ballistics of small projectiles; *J. Biosci.* **30** 167–175

Wauthy G, Leponce M, Banaï N, Silin G and Lions J C 1998 The backward jump of a box moss mite; *Proc. R. Soc. London* **B265** 2235–2242

Witztum A and Schulgasser K 1995 The mechanics of seed expulsion in Acanthaceae; *J. Theor. Biol.* **176** 531–542

*e*Publication: 18 May 2005

Series

Living in a physical world
IV. Moving heat around

STEVEN VOGEL
Department of Biology, Duke University, Durham, NC 27708-0338, USA
(Fax, 919-660-7293; Email, svogel@duke.edu)

1. Introduction – why move heat?

We care about temperature. All too often we feel either too hot or too cold. Our appliances come with thermostats, cooling fans, and thermal protection switches. The temperatures of organisms yield lovely data via thermocouples, thermal imaging equipment, and all manner of other thermometers. Temperature anomalies signal trouble, from personal fevers to global climate change. But the diverse and complex physical phenomena underlying temperature pose perilous pitfalls for explanations of such data. Furthermore, we are easily misled by our intuitive sense, that of a large, terrestrial animal that maintains a steady body temperature close to the maximum it encounters. We too easily forget that net photosynthetic rates for plants commonly peak at lower temperatures and that some of the most productive marine waters are quite cold.

In this and the next essay, I want to look at the complexities of temperature and heat, asking what physical phenomena matter most, what options are open to organisms, what devices organisms use, and what as yet undemonstrated devices might yet be uncovered.

In few terrestrial habitats do organisms lack some thermal challenge. Where I live, in southeastern North America, temperatures range from about – 19° to 37°C, on the old Fahrenheit scale a variation of no less than on of up to 1000 W m^{-2}, and air movement can range from imperceptible to overwhelming. Breathing, a convective process, comes with the evaporation's inevitable heat transfer. Our own heat production adds an additional complication – a resting human generates about 80 watts; were an adult human to retain that energy, body temperature would rise by about a degree per hour.

Why not just accept a body temperature determined by the local interplay of such phenomena? As is so often the case, we can, at best, make educated guesses, recognizing in the present case that some bacteria, for instance, tolerate truly infernal heat. Still, laissez faire might make either chemical or physical trouble, or, quite likely, both. Nearly all enzymatically-catalyzed reactions depend severely on temperature. Rates typically double or triple for every 10° rise in temperature, whether one looks at individual reactions or the overall metabolic rates of animals that do not regulate their temperatures. (To calculate proper temperature effects, the Arrhenius equation and Arrhenius constants should be used instead of this so-called Q_{10}) On top of that, most enzymes, as proteins, denature with ever increasing rapidity as temperatures rise above around 40°C. For instance, one protein that denatures a marginally tolerable 4·4% per day at 40°C, cooks (to use the appropriate vernacular) at 46% per day at 46°C.

As bad, perhaps, temperature-dependence varies from enzyme to enzyme, so sequences of reactions might demand something beyond simple mass-action effects to coordinate their operation. That may underlie the notably limited temperature range tolerated by many organisms. Most non-regulating inhabitants of niches that do not vary much in temperature cannot withstand body temperatures more than a few degrees above or below that normal range – even when well above freezing and well below the point of severe protein denaturation. The extreme in sensitivity must be creatures such as ourselves that regulate body temperature closely. Such constancy typically brings a loss of ability to survive – even briefly and dysfunctionally – without it.

Most physical variables change less with temperature. That same 10°C rise in temperature (using 20° to 30° for the examples) decreases air density by about 3·4% and the surface tension of water (against air) about 2%. It decreases the thermal capacity of water (on a mole basis) a mere 0·04% but increases the diffusion coefficients of ordinary gases in air by about 6%. One notes as a benchmark that something proportional to the absolute temperature will increase by 3·4% for a 10° rise – as does the reciprocal of air density. For instance, Weis-Fogh (1961) showed that the tensile stiffness (Young's modulus) of the protein rubber of insect wing hinges, resilin, stiffens by just that 3·4% per ten degree rise.

But a few physical quantities vary more widely. The viscosity of water decreases by over 20% from 20° to 30°C. An animal accustomed to pumping blood at 35°, say a reptile basking in the sun, must expend twice as much energy (or pump half as much blood) if it plunges into water at 5° – unless, as in so-called multiviscosity motor oils, its blood has a peculiarly low dependence of viscosity on temperature. And any increase in blood viscosity at low temperatures might well compound the problems of the temperature-dependent decreases in basal metabolic rate and maximum metabolic capability. (The maximum matters more if the animal is active while in the water – a drop in basal rate will decrease the need to move blood.)

Compounding the problem, the diffusion coefficients of solutes increase with temperature in parallel with the decrease in solution viscosity. So for a given solute at different temperatures, the product of viscosity and diffusion coefficient will remain nearly constant. That recognition came as one of Einstein's great achievements during that *annus mirabilis*, exactly a century ago, as he linked the viscosity (m) in Stokes' law for small-scale flows with the diffusion coefficient (D) in the Sutherland-Einstein relation or Stokes-Einstein equation (Pais 1982):

$$D = \frac{RT}{N}\,\frac{1}{6pmr}, \qquad (1)$$

where R is the gas-law constant, T the absolute temperature, N Avogadro's number, and r the radius of the solute molecules. (William Sutherland obtained the same result in the same year, hence Pais's suggestion of a hyphenated name.) So both biological transport processes, diffusion and convection, will be seriously impeded in liquid systems by a drop in temperature. At least if flow slows in proportion to viscosity, then Péclet numbers (ratios of convective to diffusive mass transfer: see Vogel 2004) will not change, and system geometries ought still be appropriate. Put less encouragingly, tinkering with system geometry cannot easily compensate for temperature change.

Nor do viscosity and diffusion coefficients mark the extremes. Once again looking at a rise from 20° to 30°C, the maximum concentration of water vapour in air (100% relative humidity) goes up from 17·3 to 30·4 g m^{-3} – a 75·6% increase. Put another way, water vapour makes up a mass fraction of 1·44% of saturated air at 20° and 2·61% at 30° – an increase of 81·6%. No wonder a lot of water condenses on a cool body in a hot, humid environment.

2. Heat-moving modes

How might a creature move heat from one place to another, whether shifting heat from one inside location to another, absorbing heat from its surroundings, or dumping heat onto those surroundings? A rather large array of options turn out to be available:

(i) *Radiation*: All objects above absolute zero radiate energy. A net radiative transfer of heat from warmer objects to colder ones occurs even if the objects are in a vacuum.
(ii) *Conduction*: Heat moves from warmer to colder parts of a material (or a contacting material) by direct transfer of the kinetic energy of its molecules.
(iii) *Convection*: Heat moves from warmer to colder places by direct transfer of the warmer material itself. Ordinarily its place is taken by either cooler material to close the cycle or yet more material from elsewhere.
(iv) *Phase change*: Vaporization takes energy, so it can absorb heat and leave a body cooler than otherwise. Fusion, likewise, takes energy, so melting a solid will cool either the rest of the solid or something else. Solid-to-gas change, sublimation, combines the two, absorbing even more energy.
(v) *Ablation*: The average temperature of an object of non-uniform temperature can be reduced by discarding some of its hotter-than-average portion, in effect exporting heat.
(vi) *Gas expansion cooling*: A contained gas exerts some pressure on the walls of its container; if it pushes those walls outward, thus doing work, either its temperature will drop or it will absorb heat.
(vii) *Cooling by unstressing an elastomer*: If an elastomer is stressed (stretching rubber, for instance), it warms. Elastic recoil as it is released cools the elastomer.
(viii) *Changing the composition of a solution*: Dissolving one substance in another – mixing two different liquids or dissolving a solute in a solvent – may either absorb or release heat.

Even without invoking ordinary chemical reactions or thermoelectric phenomena, we have at least eight modes of heat transfer, some of which can be divided further. All are reversible, and the last five can be used to move heat from something cool to something warm without doing violence to thermodynamics. Physics assuredly affords an abundance of possibilities that we should examine for biological relevance.

3. Radiative heat transfer

The temperature of an object determines the peak wavelength at which it either absorbs or emits radiation. How it behaves at (or near) that wavelength depends on its emissivity and absorptivity; since these do not differ, we use single measure, most often called the emissivity. (Were the two unequal, an isolated system might spontaneously move from temperature uniformity to non-uniformity, thermodynamically unlawful.) Not only peak wavelength but radiant intensity depends on temperature, the latter quite strongly. The first operative relationship, making the necessary distinction between emissivities (the e's) at incoming and outgoing wavelengths, is the Stefan-Boltzmann law:

$$q = sS(e_2T_2^4 - e_1T_1^4), \tag{2}$$

where q is the rate of energy transfer, T_1 and T_2 the Kelvin temperatures of the objects involved in the radiative exchange, S the effective exposed area, and s the Stefan-Boltzmann constant, $5{\cdot}67 \times 10^{-8}$ W m^{-2} K^{-4}.

The second is Wien's law (sometimes the "Wien displacement law"), asserting an inverse relationship between surface temperature, T, and peak emission wavelength, $l_{\max}$:

$$l_{\max} = \frac{0 \cdot 0029}{T}. \tag{3}$$

(The constant assumes temperatures in Kelvin and wavelengths in meters.) Thus peak emission of the sun, at about 5800 K, occurs at 500 nm, roughly in the middle of our visual spectrum. A organism at 30°C or 303 K will emit with a peak a little under 10 µm, far out in the infrared.

The solar peak at 5800 K, perceived by us as yellow, implies that both the photosynthetic machinery of plants and the visual systems of animals make good use of solar radiation. That may mislead slightly, an artifact of the way we ordinarily plot intensity against wavelength. The energy represented by radiation varies inversely with wavelength, something we mention parenthetically when cautioning against the hazards of the ultraviolet. So a better picture emerges from a graph with a scale on its abscissa inversely proportional to wavelength and thus independent of energy content. Wavelength inverts to frequency (f), with the speed of light as conversion factor ($f = c/l$), so frequency would work. In practice, something called "wave number", the unadjusted reciprocal, $1/l$, replaces frequency. Then equal areas under a line represent equal amounts of energy, wherever the areas might be located – a curve tolerates simple integration for energy, what matters when considering the heating effect of radiation.

Figure 1 gives such a spectrum for direct overhead solar illumination at sea level (from Gates 1965), along

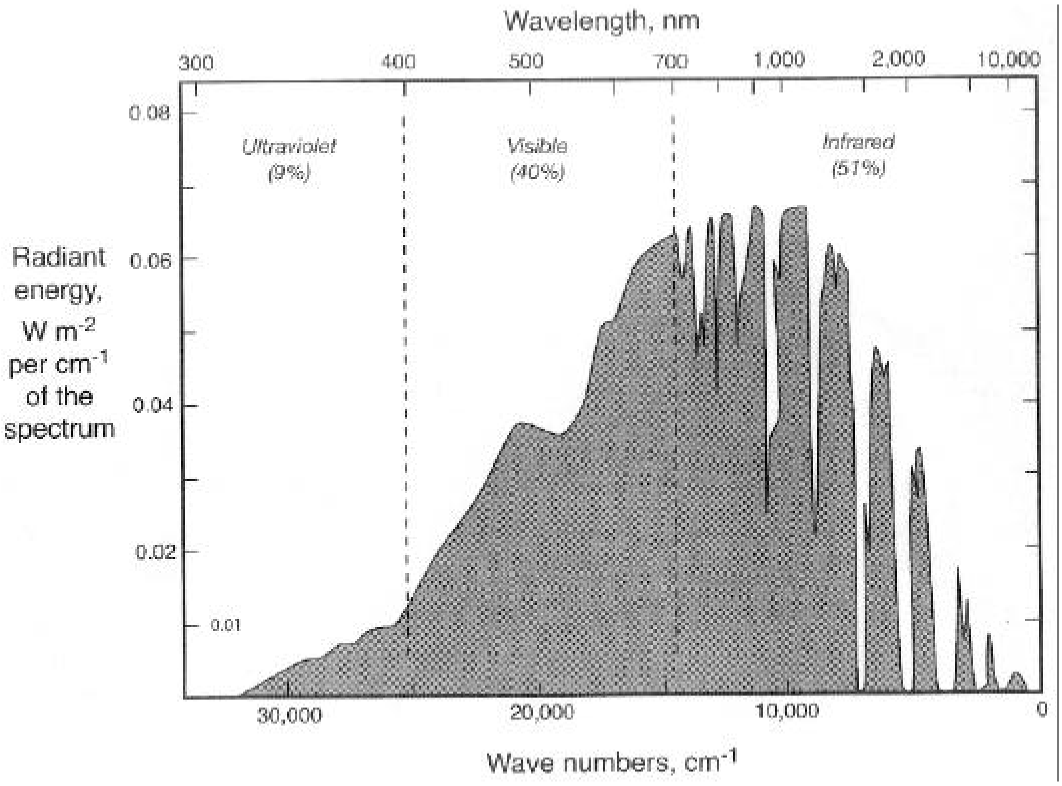

Figure 1. Spectral distribution of solar illumination of the Earth's surface on a plot in which energy is uniformly proportional to area under the curve.

with some fractional divisions (the latter from Monteith and Unsworth 1990.) Most of the energy we receive does not come in the visible at all. Fortunately, the ultraviolet, makes up only a small component; infrared radiant energy actually exceeds visible radiant energy. (The various bands absorbed by water give the curve its jagged appearance in the infrared.) That infrared radiation can make trouble for terrestrial organisms.

Consider a leaf exposed to full sunlight. It must absorb solar energy to split water and fix carbon. Yet the photons of solar radiation at wavelengths beyond 700 nm are insufficiently energetic for that purpose. If absorbed, though, they will convert to heat. We rarely worry that leaves might get intolerably hot, but the possibility should not be dismissed. The 1000 W m^{-2} of an overhead sun imposes no small thermal load – enough to heat a thin leaf by over 2° s^{-1}. By converting solar energy to a non-thermal form, photosynthesis might help, but its 5 W m^{-2} capture takes less than a per cent of the load. Leaves make a major dent in the problem by rejecting most of the infrared component of sunlight, reflecting or transmitting rather than absorbing about half the overall input. Photograph a tree with infrared-sensitive monochromatic film and a red filter to stop most of the visible light – the leaves will appear white (on a positive) against a starkly black sky.

That white 'colour' should be regarded as something special. Ordinary pigments, fabrics, animal skin and fur – all absorb infrared and thus look black. Among biological objects, are leaves unique in this respect? Unfortunately, radiative processes have drawn little attention from physiologists other than those concerned with terrestrial vascular plants. Bird eggs reflect most – sometimes over 90% – of the near-infrared. Bakken *et al* (1978) showed the independence of an egg's visible colour (commonly cryptic) from its infrared reflectivity, and the basis of the latter, the use of pigments other than the melanin typical of vertebrates. The shells and opercula of desert snails may also reflect most of the sun's direct infrared load (Yom-Tov 1971). And the spacing of the laminae in the cuticle of some iridescent red algae (from micrographs of Gerwick and Lang 1977) hints at infrared reflection; these organisms (*Iridaea* and others) can be exposed to both air and full sunlight at low tide. Still, one suspects investigative inattention rather than biological rarity. Some practical technology might come from knowing a bit more – adding a truly white roof could reduce the internal temperature of a sun-lit house in a hot place, and a truly white-crowned hat might provide shade with less concomitant heat.

The common lack of coincidence between visible colour and overall solar radiant energy absorption needs emphasis. Leaves and egg shells absorb little; fur of almost any colour absorbs a lot. That may deprive us of an easy visual assessment, but it permits organisms to decouple colour as seen by prey, predators, and conspecifics from effective radiant colour.

In addition to receiving solar radiation, organisms exchange infrared radiation with their more immediate surroundings, with intensities and wavelengths set by the Stefan-Boltzmann (eq. 2) and Wien (eq. 3) relationships. Ultimately, what matters is net transfer, something easy to forget when incoming greatly exceeds outgoing. One feels warmed on the side of the body that faces an surface above skin temperature, such as a stove, even when surrounded by air at a uniform temperature. Normally the temperatures and emissivities of organisms and their immediate surroundings are similar, so no great net heat transfer usually occurs. An exception is an open sky – a very large 'object' at a low effective temperature. According to Nobel (1999), with clean air, the effective temperature of a clear night sky may be as low as 220 K (– 53°C); with cloud cover that may rise to 280 K (+ 7°C). Thus something with a surface temperature of 30°C can radiate 3·6 times as much energy to a clear sky as it receives in return.

That asymmetry can be noticeable and significant. Ever since I became bald, I can feel whether a night sky is clear or cloudy without looking up, at least with no wind blowing. If I stand still for a few seconds beneath a clear sky, I get a particular tingle in my scalp. Of more consequence is radiation from foliage. On clear, windless nights, condensation often forms frost on low plants even when the atmosphere well above the ground remains above the freezing point – the foliage radiates sufficient energy to the sky to drop its temperature and, by conduction and convection, that of the air in its immediately vicinity, below freezing. The phenomenon can damage freeze-sensitive crops; prevention schemes include cover-

Table 1. Thermal conductivities of a variety of materials. Imagine heat transfer (W K^{-1}) through a rectangular slab of material oriented normal to heat flow. The linear dimension, m^{-1}, represents slab thickness (m^1) over slab area (m^2).

Material	Conductivity (W m^{-1} K^{-1})
Copper	385·0
Steel	46·1
Glass	1·05
Water	0·59
Skin	0·50
Muscle (meat)	0·46
Adipose tissue (human fat)	0·21
Wood (typical, dry)	0·20
Soil (inverse with air fraction)	0·25 to 2·0
Fur	0·024 to 0·063
Air	0·024

ings, sufficient wetting to overwhelm the temperature drop from the radiant emission, or (at least formerly) burning smoky fires, not to heat the crop but to obscure the sky.

One small tree may take action to avoid exposure to that cold night sky. *Albizzia julibrissin*, sometimes called the silk tree, is native to China but well established as an ornamental in the US southeast. Its doubly compound leaves with a few hundred leaflets give it a vaguely fern-like appearance. The leaves seem to have three distinct postures. In the shade, both leaves and leaflets extend horizontally; a light shining down on a leaf is almost fully intercepted. In the sun, the rachis of the leaf remains horizontal, but leaflets shift to near-vertical so the leaf casts only a minimal shadow (Campbell and Garber 1980 describe the motor responsible). At night, the entire leaf bends down to near-vertical, with the individual leaflets folded against the rachis – the leaf then looks like the tail of a horse. I suspect that the orientation in direct sunlight reduces exposure to a point source of radiation while the complete folding at night reduces exposure to a distributed radiation sink.

Postural control of solar irradiation has been documented for many terrestrial animals, mainly insects and lizards. Many of these either assume postures that minimize solar input, as in a leaf that takes up a vertical orientation during the heat of the day, or postures that maximize solar input – or both. Many insects absorb sunlight in preflight warm-ups that raise body temperatures well above ambient, taking advantage of their small size and consequently high surface-to-volume ratios. Wings often assist as shields against simultaneous convective cooling. (Heinrich 1996 gives an engaging account of the thermal devices of insects.) Lizards, larger, capitalize on their sit-and-wait predation mode to engage in more leisurely thermal basking.

Some mammals as well control solar radiation. A ground squirrel (*Xerus inauris*) that inhabits hot, dry areas of southern Africa, for instance, turns its back to the sun when conditions get especially challenging. That puts it in position to use its tail as a parasol to provide local shade. Bennett *et al* (1984), who describe the behaviour, calculate that the squirrel can thereby increase daytime foraging episodes from about 3 to 7 h.

Organisms may adjust emission as well as absorption. At the long wavelengths corresponding to their surface temperatures, desert plants have slightly higher emissivities than do plants from temperate regions, which are slightly higher than those from rain forest (Arp and Phinney 1980). All values, though, are high, most above 0·95. In general, at long wavelengths foliage, with emissivities of 0·96 to 0·98, emits more effectively than non-vegetated surfaces, typically about 0·91 (Kant and Badarinath 2002). What remains uncertain is whether the difference can confer a biologically significant additional heat loss.

Reradiation to the sky may underlie the peculiarly large and well-vascularized ears of many desert animals – jack rabbits (*Lepus* spp) in particular. As Schmidt-Nielsen (1964) points out, these animals are too small to cool by evaporating water, and most lack burrows as mid-day retreats. With air temperatures at or even above body temperature, their large ears look paradoxical. But by feeding in open shade, with hot ears exposed to a much colder sky (at an effective temperature of perhaps 13°C), an animal could off-load a large amount of heat.

4. Conductive heat transfer

The formal rules for conduction of heat parallel those for diffusion. Fourier's law (eq. 5) renames the variables in Fick's law (eq. 4), using energy transferred per unit time (q) instead of mass transfer rate (m/t), temperature difference ($T_1 - T_2$) in place of concentration difference ($C_1 - C_2$), and thermal conductivity (k) rather than diffusion coefficient (D):

$$\frac{m}{t} = DS\left(\frac{C_1 - C_2}{x}\right), \tag{4}$$

$$q = kS\left(\frac{T_1 - T_2}{x}\right). \tag{5}$$

Here S is the area over which transfer takes place and x the distance mass or heat has to move. In each process, a gradient – concentration or temperature – provide the impetus.

The only additional variable of concern is specific heat, usually given as c_p, which establishes a proportionality for a given material between energy input relative to mass and change in temperature. Water has a fairly high specific heat, 4·18 kJ kg^{-1} K^{-1} at ordinary temperatures; for air c_p is 1·01 kJ kg^{-1} K^{-1}, for soils c_p is typically (but not inevitably!) about 1·0 to 1·5 kJ kg^{-1} K^{-1}. Organisms, mostly water, rarely deviate much from its temperature-stabilizing high value.

For conduction through a slab of material, heat transfer varies inversely with thickness – as in eq. 5; for gain or loss from a solid body, rates (for most geometries) run inverse with the square of linear dimensions. And just as some diffusive step underlies every case of transfer of mass by bulk flow (as noted in Vogel 2004), conduction plays some role in all convective processes. (Advantage can sometimes be taken of that practical equivalence of diffusion and conduction. One can serve as proxy for the other, usually conductive heat transfer for diffusion, capitalizing on the greater ease of measuring temperature than chemical concentration – as, for instance, done by Hunter and Vogel 1986.)

In conduction lies the greatest divergence between heat transfer in nature and in human technology. Humans have access to metals, materials of high conductivity; non-human nature uses no metallic materials, either within organisms or in their surroundings. Metallic and non-metallic materials differ by orders of magnitude; table 1 gives a sampling of values. Between the low values of conductivity and its severe distance discount, conduction can play no great role in moving significant amounts of heat over appreciable distances in living systems. Again, consider a leaf in sunlight and nearly still air. The center of the leaf gets hotter than its margins because the latter make better thermal contact with the convective updraft induced by the hot leaf itself. Were the leaf made of metal, peak temperature would be lower – lateral conductance would move heat down the temperature gradient from center to edges. But a leaf is mainly air, water, and cellulose; and it cannot move enough heat to affect that temperature gradient, unlike the metallic heat sinks with which we protect heat-intolerant semiconductors (Vogel 1984).

Thus one should not (as have several studies) use radiantly-heated metallic models to study the thermal behaviour of leaves. Those models will have lower center and average temperatures; perhaps more importantly, as a result of their lateral heat transfer they will approach the condition referred to in books on heat transfer as “constant temperature” rather than “constant heat flux”. Unfortunately, those books reflect our metallic culture, so most of their formulas assume that unbiological near-constancy of temperature. Metal models are handy, but they must be heated in the middle rather than everywhere with a thickness of metal chosen to give the center-to-edge temperature gradient of real leaves.

In a sense pure conduction represents a gold standard for minimal heat transfer. Thus fur works by reducing convective air movement enough for overall transfer to approach the value for conduction in air. And heat exchangers (about which more in the next essay) drive the heat transfer due to blood circulation down toward the value for conduction in isolated tissue.

Nonetheless, a few organisms do employ conductive heat transfer as more than a short-distance link between a flowing fluid and an adjacent surface. Our elderly house cat rests on dry straw in the garden on cool days; on hot days he shifts to bare soil or pavement that never gets direct sunlight. The pattern is common among medium and large-size domestic animals with soft enough flesh and fur for effective contact with the substratum. More specific use of heat earthing has been documented in a desert rodent, the antelope ground squirrel (*Ammospermophilus leucurus*). For a diurnal desert animal it is especially small, which means that it heats up rapidly when foraging in the summer sun – 0·2 to 0·8°C min^{-1}. A squirrel deals with this heat load by tolerating brief bouts of hyperthermia (sometimes exceeding 43°C) and returning to its burrow as often as every 10 min. In the burrow, it loses heat rapidly by pressing itselfs against the walls, which are about 10°C cooler than its body (Chappell and Bartholomew 1981).

5. Convective heat transfer

Conduction poses few analytic problems, with reliable equations and only the peculiarities of biological geometries to complicate things. Radiative exchange may be less familiar, but, likewise, we can rely on straightforward rules. But whether looking at thermal phenomena within or around our creatures, we can rarely ignore convective transfer. And no such tidiness characterizes convection. While the textbooks for engineering courses (I particularly value Bejan 1993) provide reliable explanations, the equations they cite must be viewed warily. Most provide no more than rules of thumb, many presume conditions quite different from what organisms encounter, and even the first figure of their three-significant-figure constants may diverge from our reality. To list a few of the complicating aspects of convection:

(i) Internal versus external convection. We move lots of heat by pumping blood and other fluids through our various pipes and internal channels; flows of air and water around us transfer heat between ourselves and the environment. The basic phenomenon may be the same, but the practicalities depend strongly on whether the solid object surrounds the fluid or vice versa.

(ii) Flows may be laminar or turbulent, with major differences for heat transfer. In most laminar flows (such as in our capillaries) convection carries heat only with the overall flow – conduction drives transfer normal to the direction of flow. By contrast, the internal mixing of turbulent flow provides a major avenue for cross-flow transfer, and the thermal conductivity of the fluid loses most of its importance. For internal flows through circular pipes, the shift from laminar to turbulent occurs at a reasonably sharp value (2000 ± 1000) of a single variable, the so-called Reynolds number, *Re*:

$$Re = \frac{rlv}{m}, \tag{6}$$

where r and m are the fluid’s density and viscosity, l the diameter of the pipe or width of the channel, and v its average flow speed. External flows may have a similarly sharp transition, but the location of the transition depends a lot on texture and geometry – between *Re*’s of about 20 and 200,000, with l now taken as a variously defined characteristic length of the object in the flow.

(iii) Convection can be driven by density differences within the fluid – "free convection" – or it may be driven by some external current – "forced convection." Unlike the previous distinctions, regimes can be mixed. Another dimensionless number, the Grashof, *Gr*, provides an index of the intensity of free convection:

$$Gr = \frac{rgb(\Delta T)l^3}{m^2}. \qquad (7)$$

The only new variable is b, the volumetric thermal expansion coefficient; its value for liquid water is about $0{\cdot}3 \times 10^{-3}$ K^{-1}. All gases have about the same value of b. Since their volumes vary directly with the absolute temperature, $b = 1/T$; at 20°C, $b = 3{\cdot}4 \times 10^{-3}$ K^{-1}. Free convection mainly matters for external flows. It can be laminar or turbulent with a transition from former to latter at a Grashof number of about 10^9.

In substantial winds, for very large objects, for objects well above or below ambient temperature, forced convection will dominate the picture. But what of a small organism exposed, say, to sunlight in nearly still air? Another dimensionless index provides a rough-and-ready criterion, the ratio of the Grashof number to the square of the Reynolds number. In effect, this looks at the ratio of buoyant force to inertial force; viscous force, affecting both components, cancels out. Thus

$$\frac{Gr}{Re^2} = \frac{gb\Delta(T)l}{rv^2}. \qquad (8)$$

Some sources give the following rules of thumb. For ratios below about 0·1, forced convection predominates and free convection can be ignored. For ratios above about 16, free convection predominates and the effects of whatever wind might be present can be ignored. Higher thermal expansion coefficients, larger differences in temperature between organism and surroundings, and larger size raise the value and favour free convection; denser fluids and more rapid flows favour free convection, all intuitively reasonable.

By this criterion, mixed regimes cannot be ignored. Consider, yet again, a sun-lit broad leaf on a tree. A leaf 10 cm across will encounter a mixed regime at wind speeds between about 0·04 and 0·5 m s^{-1}. The lower figure is less than ambient wind ever gets for more than a few seconds in full sun. If nothing else, differential heating of ground and other foliage will generate that much convection. The higher figure, about our perceptual threshold for air movement, will nearly cool a leaf to air temperature – stronger winds make little further difference, and overheating ceases to be hazardous.

For that leaf, then, the only significant regime is a mixed one, the regime least amenable to anything other than direct measurements. Some years ago, a local engineering graduate student, Alexander Lim (1969), compared published formulas for mixed free and forced convection with measurements under conditions a leaf might encounter. He found even greater deviations than we expected, with discrepancies typically around 50% – in both directions. And he controlled variables that in nature would confound things even further. For instance, free convection carries air vertically, while forced convection need not be horizontal, since it includes not just ordinary wind but the upward free convection of adjacent leaves.

The main generalization one might make is that free convection will be insignificant for very small systems and a major consideration only for quite large ones. As Monteith and Unsworth (1990) point out, a cow might lose heat by free convection when the wind drops below about 1 m s^{-1}. Similarly, a camel, need not wait for a gentle breeze to dump heat convectively at night that it had acquired during the previous day. Judging from photographs of thermal updrafts around standing humans (using a technique which visualizes differences in air density), our large size permits some self-induced free convection. Still, even barely perceptible air movements help us avoid overheating when we work hard under hot and humid conditions. On yet larger scales free convection becomes yet more important; together with spatially irregular heating of the ground it produces the ascending thermal tori in which birds such as hawks and vultures soar.

6. Conduction versus convection

For biological systems, made of low conductivity materials, pure conduction with zero convection represents a kind of gold standard for minimal internal heat transfer. A warm human increases convective transfer by vasodilation of capillaries in the skin and the associated larger blood vessels – body temperature becomes less spatially variable. When cold, one reduces blood flow to the extremities, setting up internal temperature gradients closer to those of conducting systems. But we humans remain convection-dominated, reflecting both our high aerobic capacity and warm-climate ancestry.

How might one determine the relative importance of conduction and convection in an intact, living animal? Measuring blood flow will not give reliable results since heat exchangers (about which more in the next essay) can decouple heat flow from mass flow. A simple scaling argument suggests at least one possible approach – it adopts the rationale for circulatory systems of the Nobel laureate physiologist August Krogh (1941), merely substituting heat for oxygen.

If heat content depends on volume ($\propto l^3$) and heat loss depends on surface area ($\propto l^2$), then the rate of heat loss relative to volume will vary inversely with a typical linear dimension ($\propto l^{-1}$). That should happen where heat moves much more readily within the organism than to or from the organism. If, conversely, heat loss depends on the distance between core and periphery ($\propto l^1$), then the rate of heat loss relative to volume will vary inversely with the square of linear dimensions ($\propto l^{-2}$). That will happen when heat transfer to and from the organism presents less of a barrier than transfer within the organism. Muscle, fat, and other biological materials have low thermal conductivities (table 1 again), while circulating liquids make fine heat movers. So the first situation (loss $\propto l^{-1}$) will characterize convection-dominated cases, the second (loss $\propto l^{-2}$) conduction-dominated cases.

One needs only scaling data, at least at this crude level of judgment. Measurements of core body temperatures as equilibrated animals are heated or cooled will suffice, at least for ectothermic animals – temperature tracks heat loss per unit volume if heat capacity remains constant. Can such an easy model apply, or do confounding factors overwhelm it?

As a quick test, I created two sets of heat-transferring systems, one predominantly convective and the other exclusively conductive. Each set consisted of six ordinary laboratory beakers, of nominal capacities from 50 to 1000 ml, with each beaker filled to a depth equal to its internal diameter. A thermometer supported by a piece of corrugated paperboard extended down to the center of each beaker, as in figure 2. One set contained pure water while the other was filled with water plus 1% agar – the small amount of agar suffices to immobilize the water, preventing the free convection of self-stirring without significant effect on its specific heat. The twelve beakers were equilibrated overnight in an incubator at 49°C, moved at time zero to a room at 25°C, and their temperatures recorded every 5 min. Free convection stirred the water-filled beakers enough to make deliberate stirring unnecessary, and room air movement sufficed to minimize external resistance. Figure 3 shows the results, with log-log slopes satisfyingly close to the predicted values. An analogous exercise in which the beakers warmed after equilibration at 7°C gave much the same result – immobilizing the water gave greater reductions in the rate of temperature change in larger systems.

So this simplest of scaling rules can place systems on a spectrum from pure conduction to predominant convection. As an example, we might look at some old data for cooling lizards. For a variety of cooling varanids, Bartholomew and Tucker (1964) found a scaling exponent of – 1·156 (tripling their mass-based number), just a bit

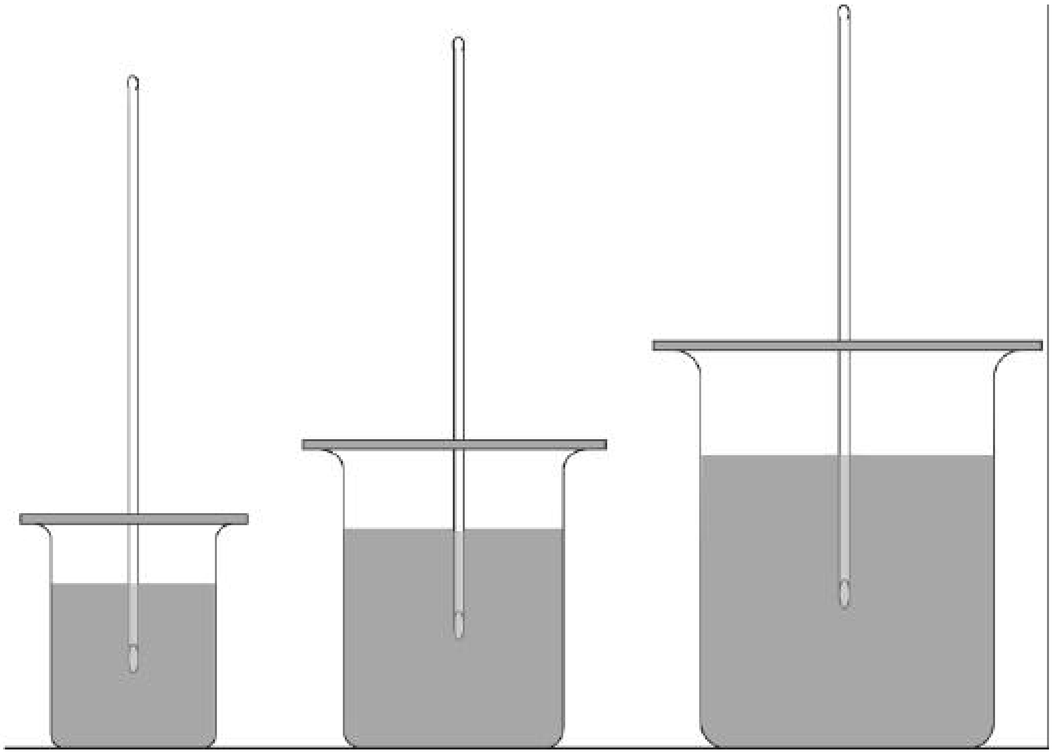

Figure 2. The arrangement of beakers and thermometers used to obtain the scaling data of figure 3.

greater than what we would expect for convection-dominated systems. By contrast, Bartholomew and Lasiewski (1965) reported an exponent of – 1·881 for Galapagos marine iguanas suddenly immersed in cold water, just short of what we anticipate for conduction-dominated systems. During dives, heart rates slow, but no more so than for the varanids. Somehow they must reroute their blood so it carries little heat peripherally. (Whether in air or water, the iguanas reheat much more rapidly.)

Cooling slowly makes adaptive sense for reptiles that bask on warm, sunny, shoreline rocks and then plunge into fairly cold water to feed. Charles Darwin gives a fine descriptions of iguana and its behaviour in *The Voyage of HMS Beagle* (1845) ("a hideous looking creature") as well as in his diaries (unoriginally, "imps of darkness"). Not that these iguanas do anything unprecedented. Immersed reptiles quite commonly heat faster than they cool, with the ratio increasing with body size, as noted by Turner (1987) and consistent with our scaling exponents.

One caveat. For systems surrounded by minimally moving air – insulated systems, the external resistance to heat transfer can approach the internal resistance. Thus reducing internal resistance by preventing convection within the system may not decrease the exponent for loss relative to size as much as expected. A somewhat unnatural comparison illustrates the effect. Turner (1988) gives heat loss exponents of – 1·8 for infertile eggs cooling in water, nearly the – 2·0 of our model, but only – 1·2 when cooling in still air.

7. Heat transfer by evaporation and condensation

Vaporization of a liquid (or sublimation of a solid) provides a particularly effective heat transfer mechanism, especially if the liquid has a high heat of vaporization, as does water. Indeed, the value most often found in non-biological sources, 2·26 MJ kg^{-1}, presumes boiling at 100°C and understates the case; at a more biologically reasonable 25°C, water's heat of vaporization is 2·44 MJ kg^{-1}, about 8% higher.

Several conditions, though, limit its use by organisms. The atmosphere into which water vaporizes must not be water-saturated, at least at the temperature of the evaporating surface (which, as our skin commonly is, may be above ambient temperature). Evaporation itself will reduce the temperature of the evaporating surface (again, as with our skin). And a copious supply of water must be available. A succulent plant, some lore to the contrary, cannot store enough water for significant evaporative cooling in a warm, dry habitat. Hard-working humans, cooling evaporatively as we do, must consume water at a great rate. At a metabolic rate of 400 W (a minimal esti-

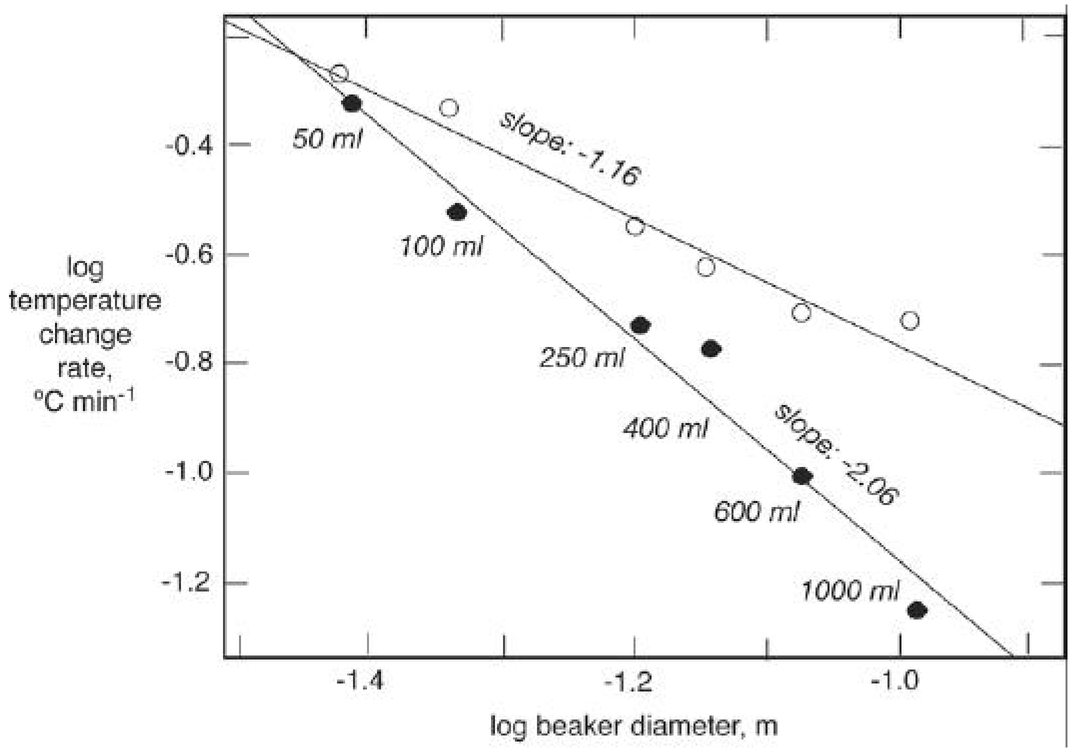

Figure 3. Log cooling rate versus log beaker diameter for water-filled (slope = – 1·16) and water-plus-agar filled beakers (slope = – 2·06). The slopes represent the exponent *a* in the expression (cooling rate ∝ diameter^a).

mate for a labourer working at 100 W output), dissipating metabolic heat by evaporation, our main mechanism, would take 0·6 l h^{-1}, or almost 5 liters for an 8 h working day.

Few small animals can rely on evaporative cooling as a principal mode during sustained activity – it demands too great a volume of water for the surface area exposed to a hot environment or for their metabolic rates (which scale nearly with surface area). Fortunately their higher surface-to-volume ratios improve the efficacy of convection. Not unexpectedly, they seem more often concerned with water conservation, with devices that reduce respiratory water loss and so forth.

Among animals that cool evaporatively, two routes play major roles; each has its points. Evaporation from skin (predominant in humans, cattle, large antelopes, and camels) takes advantage of the skin's large surface area. The concomitant vasodilation improves convective loss as well. On the debit side, cutaneous evaporation inevitably causes a loss of salt, which then becomes a particularly valuable commodity for herbivores active in hot climates. In addition, its requirement for exposed external surface conflicts with the presence of fur or plumage that might reduce heat loss under other circumstances.

Respiratory evaporation entails no salt loss, but it requires pumping air across internal surfaces, which costs energy and produces yet more heat. And the CO_2 loss in excess breathing drives up the pH of the blood. Animals such as dogs, goats, rabbits, and birds that use respiratory evaporation beyond that associated with normal gas exchange reduce both problems by panting – shallow breaths repeated at rates matching the natural elastic time constants of their musculoskeletal systems (Crawford 1962; Crawford and Kampe 1971).

Some mammals (rats and many marsupials) cool evaporatively by licking their fur and allowing the saliva to evaporate; the mode, though, is not used during sustained activity. Some large birds (vultures, storks, and others) squirt liquid excrement on their legs when their surroundings get hot (Hatch 1970), augmenting evaporative cooling. A few insects with ample access to water (nectar and sap feeders) derive clear benefit from evaporative cooling for dumping the heat produced by flight muscles despite their low surface-to-volume ratios – some cicadas, sphingid moths, and bees in particular (Hadley 1994; Heinrich 1996).

What about leaves? Again, many do get well above ambient temperature, pushing what look like lethal limits. Plants with broad leaves, the ones likely to run into thermal trouble, evaporate water ('transpire') at remarkable rates. Leaf temperatures calculated (from admittedly crude formulas) by Gates (1980) point up the thermal consequences of that evaporation. He assumes a wind of 0·1 m s^{-1} (as noted earlier, about as still as daytime air gets), solar illumination of 1000 W m^{-2} (again, an overhead unobstructed sun), an air temperature of 30°C, a relative humidity of 50%, and a leaf width of 5 cm. If reradiation were the only way the leaf dissipated that load, it would equilibrate (recall eq. 2) at a temperature of about 90°C. Allowing convection as well drops that to a still stressful 55°C. A typical level of evaporation cools the leaf to – hot but not impossibly so for a worst-case scenario. Evaporation cools leaves; it could not do otherwise. Typically broad leaves dissipate about as much energy evaporatively as they do convectively.

Less clear than its thermal consequences is the thermal role of this transpirative water loss. Plant physiologists (see, for instance, Nobel 1999) generally regard the loss as an inevitable byproduct of the acquisition of CO_2 – a leaf with openings (stomata) that admit inward diffusion of CO_2 will permit outward diffusion of water. CO_2 makes only about 0·03% of the atmosphere, and the diffusion coefficient of CO_2 is well below that of H_2O. So a lot of water must vaporize for even a modest input of the crucial carbon upon which plants depend. A representative value for water-use efficiency (Nobel 1999) is about 6 g CO_2 per kg H_2O. Functioning leaves have to lose water, whatever the thermal consequences. Indeed, transpiration sometimes depresses leaf temperatures 10°C or more below ambient. The situation resembles evaporative heat loss from our breathing, something of minor use (since we do not pant) for an excessively warm human but a distinct liability for one stressed by cold.

But that view cannot be wholeheartedly embraced. Water-use efficiencies vary widely. The extreme values come from measurements on those species (6 or 7% of all vascular plants) that only open their stomata at night, when temperatures are lower and relative humidities higher. They fix CO_2 as organic acids; decarboxylation the next day provides the input for photosynthesis. The trick can push water-use efficiency up an order of magnitude. So the adaptive significance of evaporative water loss from leaves remains uncertain. The question has drawn little attention – plant physiologists have worried less than have animal physiologists about primary – adaptive – versus secondary functions of multifunctional processes.

If evaporation cools, then condensation heats. Under at least one condition organisms may use condensation as a significant heat source. On cold, clear, calm nights, radiative cooling, as noted earlier, often drops leaf temperatures below both the local air temperature and the local dew point – the term "dew point" comes from the resulting condensation. It provides a major water source for some low desert plants. Sometimes water vapour condenses as frost; where that happens the heat of sublimation, greater by 13% at 0°C than the heat of vaporization, becomes the relevant factor. Condensation as dew or frost should offset some of that radiative cooling; again, the practical significance is uncertain. Frost *per se* causes little trouble – what damages plants is internal ice formation signalled by its appearance.

A wide variety of arthropods have been shown capable of condensing water from the atmosphere. In none does it seem to be such a simple physical process – the required temperature differences just do not occur, nor would they be likely in animals as small as ticks, fleas, and mites. Nor is a vapor-saturated atmosphere necessary – the minimum humidity can be as low as 50%. In none of these animals does condensation appear to confer any specific thermal benefit – obtaining liquid water is the pay-off (Hadley 1994).

A recent report implies a thermal role for still another form of phase change, one whose novelty may only reflect oversight. According to Dunkin *et al* (2005), a large fraction of dolphin blubber consists of fatty acids with melting points just below body temperature. The apparent thermal conductivity of the blubber of both young dolphins and pregnant females is well below that of human fat (as in table 1), and heat flux measurements suggest heat absorption by phase change as the mechanism.

8. Other modes – known and unknown

So far, we have only looked at half the heat transfer modes mentioned at the start – radiation, conduction, convection, and phase change. Some of the others can be either dismissed outright or their insignificance easily argued. Early spacecraft used *ablative cooling* when reentering the atmosphere. Animals, as noted, do void saliva and excrement, but the subsequent evaporation of the liquid from deliberately wetted skin or fur does far more to get rid of body heat than does ablation itself.

You can use *gas expansion cooling* to make excessively hot food or drink palatable by pursing lips and exhaling air that has been compressed by your thoracic muscles – air temperature can be dropped into the mid 20's according to a quick measurement on a cooperative colleague. But the muscle-powered compression-expansion sequence heats you more than it cools the food. Useful heat transfer by *stressing and unstressing elastomers* seems unlikely, even if the imperfect resilience of biological materials might be used (as in pre-flight warm-up in insects or in our shivering) as a small supplement to muscular heat generation. Similarly, transferring significant amounts of heat by *dissolving or extracting solutes* is unlikely, even though organisms commonly manipulate the composition of solutions.

What ought not be casually dismissed are novel combinations of the various heat transfer mechanisms. As an example of an unknown but biologically plausible scheme, consider a so-called heat pipe (figure 4), a device that combines phase change and convection. A liquid vaporizes at the warm end, absorbing heat. Vaporization produces a pressure difference that drives gas toward the cool end. There it condenses, releasing heat. Liquid then returns to the warm end by capillarity through some wicking material lining the pipe. A few uncommon bits of human technology use heat pipes since they can achieve effective conductivities orders of magnitude greater than that of copper bars of the same dimensions, but they have never become household items.

By contrast, heat pipes should be highly advantageous in nature inasmuch as organisms are made of materials of such low thermal conductivities. Having only water as a working fluid, though, imposes a serious limitation. Admittedly, water has a nicely high heat capacity. And the concentration of vapour at saturation is strongly temperature-dependent; recall the 81·6% increase in mass between 20° and 30°C that was mentioned earlier. But pressure-driven bulk flow from warm to cool end cannot drive vapour movement as it does in systems where nothing dilutes the substance that evaporates and condenses. Rough calculations suggest that diffusion, the obvious alternative, will not move enough water vapour over distances greater than about a millimeter. So such a system needs some local stirring of the gas phase – cross-flow thermal gradients, continuous flexing of the pipe, or something else.

Where in organisms might we find heat pipes? Air-filled passages with hydrophilic inner surfaces are not rare. I wonder about the insides (the spongy mesophyll) of small, succulent leaves. Several colleagues, Catherine Loudon and Thomas Daniel, suggest that insects might

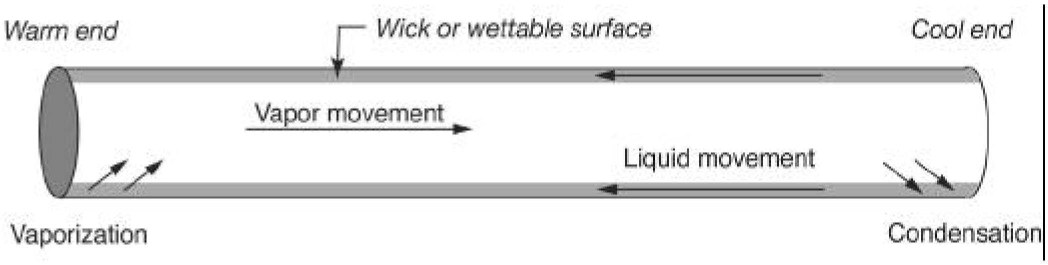

Figure 4. The operation of a heat pipe, with heat flow from left to right.

use the mechanism to move heat from flight muscles through their tracheal systems.

So the ease with which we measure temperature may all to easily obscure the complexity of thermal phenomena in general and the thermal behaviour of organisms in particular. Designing proper experiments challenges our ingenuity. For instance, putting an organism in a temperature-controlled chamber may not come close to mimicking the thermal character of a habitat of the same temperature. The walls will not behave like open sky, and the heat source will be unlikely to resemble the sun. The air movement needed to assure constant temperature will probably be unrealistically high – for instance for studying thermally stressed leaves. Or it may be unrealistically low – for, say, looking at the insulation fur provides for a mammal in the open. Beyond these difficulties lie the problems associated with the continuous variation in environmental temperatures, insolation levels, wind speeds, and so forth in nature.

Put in these terms, the obstacles appear daunting. I prefer to view the situation in a different light. Physical complexity instigates biological diversity, not just in phylogenetic terms, but as diversity of clever designs and devices awaiting elucidation. And identifying what nature does begins by recognizing the physical possibilities.

Acknowledgements

Thomas Daniel, Michael Dickison, Sönke Johnsen, Catherine Loudon, Knut Schmidt-Nielsen, Richard Searles, Vance Tucker, and William Wilson suggested biological cases and references. And both the clarity and accuracy of the text have benefitted from Scott Turner's suggestions.

References

Arp G K and Finney D E 1980 Ecological variations in the thermal IR emissivity of vegetation; *Environ. Exp. Bot.* **20** 135–148

Bakken G S, Vanderbilt V C, Buttemer W A and Dawson W R 1978 Avian eggs: thermoregulatory value of very high near-infrared reflectance; *Science* **200** 321–323

Bartholomew G A and Lasiewski R C 1965 Heating and cooling rates, heart rate, and simulated diving in the Galapagos marine iguana; *Comp. Biochem. Physiol.* **16** 575–582

Bartholomew G A and Tucker V A 1964 Size, body temperature, thermal conductance, oxygen consumption, and heart rate in Australian varanid lizards; *Physiol. Zool.* **37** 341–354

Bejan 1993 *Heat transfer* (New York: John Wiley)

Bennett A F, Huey R B, John-Alder H and Nagy K A 1984 The parasol tail and thermoregulatory behaviour of the cape ground squirrel *Xerus inauris*; *Physiol. Zool.* **57** 57–62

Campbell N A and Garber R C 1980 Vacuolar reorganization in the motor cells of *Albizzia* during leaf movements; *Planta* **148** 251–255

Chappell M A and Bartholomew G A 1981 Activity and thermoregulation of the antelope ground squirrel *Ammospermophilus leucurus* in winter and summer; *Physiol. Zool.* **54** 215–223

Crawford E C 1962 Mechanical aspects of panting in dogs; *J. Appl. Physiol.* **17** 249–251

Crawford E C and Kampe G 1971 Resonant panting in pigeons; *Comp. Biochem. Physiol.* **A40** 549–552

Darwin C 1845 *Journal of researches into the natural history and geology of the countries visited during the voyage of H M S Beagle under the command of Capt. Fitz Roy R N* (London: J Murray)

Dunkin R C, McLellan W A, Blum J E and Pabst D A 2005 The ontogenetic changes in the thermal properties of blubber from Atlantic bottlenose dolphin *Tursiops truncatus*; *J. Exp. Biol.* **208** 1469–1480

Gates D M 1965 Energy, plants, and ecology; *Ecology* **46** 1–13

Gates D M 1980 *Biophysical ecology* (New York: Springer-Verlag)

Gerwick W H and Lang N J 1977 Structural, chemical and ecological studies on iridescence in *Iridaea* (Rhodophyta); *J. Phycol.* **13** 121–127

Hadley N F 1994 *Water relations of terrestrial arthropods* (San Diego: Academic Press)

Hatch D E 1970 Energy conserving and heat dissipating mechanisms of the turkey vulture; *Auk* **87** 111–124

Heinrich B 1996 *The thermal warriors: strategies of insect survival* (Cambridge: Harvard University Press)

Hunter T and Vogel S 1986 Spinning embryos enhance diffusion through gelatinous egg masses; *J. Exp. Mar. Biol. Ecol.* **96** 303–308

Kant Y and Badarinath K V S 2002 Ground-based method for measuring thermal infrared effective emissivities: implications and perspectives on the measurement of land surface temperature from satellite data; *Int. J. Remote Sensing* **23** 2179–2191

Krogh A 1941 *The comparative physiology of respiratory mechanisms* (Philadelphia: University of Pennsylvania Press)

Lim A T 1969 *A study of convective heat transfer from plant leaf models*, MS Thesis, Duke University, Durham NC, USA

Monteith J L and Unsworth M 1990 *Principles of environmental physics* (Oxford: Butterworth-Heinemann)

Nobel P S 1999 *Physicochemical and environmental plant physiology*, 2nd edition (San Diego: Academic Press)

Pais A 1982 *Subtle is the lord–: the science and the life of Albert Einstein* (New York: Oxford University Press)

Schmidt-Nielsen K 1964 *Desert animals* (Oxford: Oxford University Press)

Turner J S 1987 The cardiovascular control of heat exchange: consequences of body size; *Am. Zool.* **27** 69–79

Turner J S 1988 Body size and thermal energetics: how should thermal conductance scale?; *J. Thermal Biol.* **13** 103–117

Vogel S 1984 The thermal conductivity of leaves; *Can. J. Bot.* **62** 741–744

Vogel S 2004 Living in a physical world I. Two ways to move material; *J. Biosci.* **29** 392–397

Weis-Fogh T 1961 Thermodynamic properties of resilin, a rubber-like protein; *J. Mol. Biol.* **3** 648–667

Yom-Tov Y 1971 Body temperature and light reflectance in two desert snails; *Proc. Malacol. Soc. London* **39** 319–326

Series

Living in a physical world
V. Maintaining temperature

STEVEN VOGEL
Department of Biology, Duke University, Durham, NC 27708-0338, USA
(Fax, 919-660-7293; Email, svogel@duke.edu)

1. Introduction

Little else in our immediate world varies as much as the thermal loads that we terrestrial organisms face. Too often we find ourselves too hot, too cold, too well illuminated by sunlight, too exposed to an open sky, or in too great contact with hot or cold solid or liquid substrata. Thermal loads vary in time scale as well as in magnitude. Air temperatures and radiative regimes change over every time scale relevant to their operation, from seconds to years, at the least; in addition both soil and water temperatures may be far from constant. Variation may be as regular as night following day or it may be predictable only in a general statistical sense. Terrestrial life – and sometimes even aquatic life – is rife with thermal challenges.

The last essay (Vogel 2005) argued that variable internal temperature could impose serious constraints on biological design. It looked first at the way temperature, both extremes and fluctuations, might affect the operation of organisms. It then turned to the various physical agencies that could move heat to, from, and within organism. Here I will take a complementary look at these same issues, exploring the ways in which organisms can mitigate those fluctuations, focusing for the most part on how creatures can avoid moving heat.

Ideally, holding internal temperature at a value different from that outside should cost no energy – in general, all cost reflects imperfect thermal isolation. We might venture a sweeping generalization, asserting that adaptations for maintaining appropriate temperatures in a world of extremes and fluctuations have a particular common character. All (or, to be on the safe side, almost all) work by minimizing the metabolic work expended on temperature control. While energy economy may not be the transcendent issue that many of us once presumed, its importance cannot be denied.

And we might assert another generalization, a bit less sweeping than the preceding. Conduction, whether through air, water, or tissue, most often establishes a base line; pure conduction represents a kind of gold standard. For transfer within an organism, the central challenge comes down to reducing the convective heat transfer accompanying flow in blood vessels and air passageways to a level at which conduction predominates. If that can be done, avoiding excessive temperature fluctuations with minimum energy expenditure can take advantage of the conveniently low thermal conductivities of life's two main media, air and water – or, much the same as the latter, flesh and bone.

Thus air and water set the standards. All gases have low thermal conductivities; air's value, $0{\cdot}024\ \mathrm{W\ m^{-1}\ K^{-1}}$, is ordinary for a gas or gas mixture – argon may be 32% and CO_2 36% lower, but hydrogen is 7 times higher. Liquid water, at $0{\cdot}59\ \mathrm{W\ m^{-1}\ K^{-1}}$, is quite as ordinary, here by comparison with other non-metallic liquids as well as solids – 40% lower than glass and 46% lower than limestone but about three times higher than pure fat, isolated whale or seal blubber (see Dunkin *et al* 2005), and common plastics such as the acrylics. Except, perhaps, for switching from watery muscle to minimally hydrated fat, reduction of thermal conductivity has little to offer.

(It should be noted that instead of thermal conductivity, animal physiologists often use thermal conductance, the combined rates of conductive and convective transfer per unit surface and per degree. With units of $\mathrm{W\ m^{-2}\ K^{-1}}$ rather than $\mathrm{W\ m^{-1}\ K^{-1}}$, it ignores the thickness of any insulating layer. That makes good sense when looking at experimental data from irregularly shaped and variably coated animals. By contrast, data for conductivity usually comes from *in vitro* measurements on pelts and tissue samples. Thus finding that thermal conductance varies inversely with thickness implies that thermal conductivity does not change.)

2. Circumventing convection

Air ordinarily moves unless prevented by some specific device, and it moves at speeds that matter. Speeds far less than our perceptual threshold of about 0·5 m s^{-1} still have thermal consequences. An oak leaf in the sun whose axial temperature at 0 to 0·01 m s^{-1} is 41°C will reach only 37° at 0·1 m s^{-1} (Vogel 1968). So "still air" in the meteorological sense may be presumed non-existent in the thermal world of organisms. If nothing else, any organism whose surface temperature differs from the surrounding air will experience self-induced free convection. Additionally, macroscopic organisms move fluids internally since some form of bulk transport is a practical prerequisite for getting much above cellular size. Such transport systems will move heat as well as material, and that heat transfer may have either positive or negative consequences.

One way to reduce internal convective heat transfer consists of simply reducing blood flow to the periphery and extremities by vasoconstriction. Small adjustments in relative vessel diameters can substantially reroute blood. We certainly do just that when inactive and exposed to cold, allowing our skin and appendages to stay at temperatures well below those of our brains and viscera. In the cold, it is normal for skin temperature to be 10° below that of the body's core. In one old experiment (DuBois 1939) nude males were asked to rest quietly in what was described as still air. Exposure to an ambient 22·5°C, perceived under the circumstances as quite chilly, dropped core temperature by about 0·5°C; it dropped average skin temperature by 7°C – hands somewhat less, feet by as much as 10°C. In more extreme cold exposure we begin to defend core temperature by increased metabolic activity, noticeable as the minimally-coordinated muscular contractions of shivering, rather than by further reduction in peripheral circulation. These responses appear fairly general among warm-blooded vertebrates, not just unfurry and unfatty ones such as ourselves.

In practice, vasoconstriction combines two physical agencies. It reduces convection by creating a peripheral region in which flow is minimal. And it lowers conduction because lengthening the distance between central and surface temperature in, say, an appendage reduces the steepness of the temperature gradient. Experimental studies rarely tease apart the mix; one presumes that it varies case to case and place to place.

Adding insulation works in much the same way as vasoconstriction, again through a pair of physical agencies. And it has two biological manifestations – internal insulation using peripheral layers of fat and external insulation of fur and feathers.

Fat, as noted earlier, has an agreeably low thermal conductivity, about three times lower than water or meat. In addition, few tissues approach the low metabolic activity of subcutaneous fat – the reason metabolic rates are often referred to "lean body mass" for comparisons among different animals. Thus addition of subcutaneous fat reduces peripheral circulation as well. And subcutaneous fat layers can be remarkably thick, getting up to about 50% of total body volume in aquatic mammals that swim in cold waters. With this blubber, a seal can have both a skin temperature about a degree above 0º and a core temperature in the mid 30's (Irving and Hart 1957). The significant insulating effect of subcutaneous fat in humans underlies the common observation that females, with thicker layers, tolerate full-body exposure to cold water better than do males, whether they are Korean pearl divers or (at least as I have observed) marine biologists.

Fur and feathers permit effective conductivities to approach the value of pure air by limiting both free and forced convection. In no case does their own conductivity, that of the protein keratin, take on particular importance. Again we lack good data on how much of their effectiveness represents restriction of flow (usually air, in this case) and how much comes from reduction in the thermal gradient over which conduction occurs. Another uncertainty concerns the effects of ambient wind. Designing a fur coat of greatest effectiveness for its cost and thickness should depend on the importance of the free convection of the warm animal itself relative to that of environmental air movement. A fur coat has a dynamic component as well. Piloerection permits some degree of adjustment of its thickness and thus its thermal effectiveness – although our own attempts, noticed as so-called goose-flesh or goose-bumps (reminiscent of plucked poultry) accomplish little.

In practical terms (sweeping such complications aside) a single number provides a simple measure of the effectiveness of the fur coat of a mammal, given the near-uniformity of mammalian core temperatures and our consistent preference for insulation over metabolic increase. One needs only the temperature below which insulation is insufficient to permit a mammal to maintain normal eutherian body temperature at basal metabolic rate, the temperature below which metabolites must be expended simply to stay warm. Naked human males (females, with more subcutaneous fat, do a bit better despite their smaller sizes) have to turn up the fire at about 27°C, which is not at all impressive – presumably we are still warm-country pursuit predators, better adapted for heat dissipation than for conservation. Sloths do still worse, with critical temperatures around 29°C. Even small mammals, with fur length limited by other considerations, can do better, with weasels at 17°C and ground squirrels at 8°C. Large mammals, especially arctic ones tolerate cold with remarkable economy – lower critical temperatures commonly run between 0°C and – 40°C. (Scholander *et al* 1950).

3. Offsetting convection with countercurrent exchangers

A convective link between hot and cold locations need not transfer heat. The agency can be turned against itself – if it can carry heat one way, it should be able to carry it in the other quite as well. In the context of a warm animal in a cold place, the trick consists of transferring heat from blood flowing peripherally, not to the environment, but to blood flowing axially. The engineering literature refers to the device for doing that as a counterflow exchanger, physiologists prefer the word 'countercurrent' (often spelled 'counter-current'). The key element is a region, typically near the base of an appendage, in which arteries and veins lie in sufficiently intimate juxtaposition for that heat transfer. If blood were to travel in the same direction in both arteries and veins, the best that could be achieved would be an output that averaged hot and cold inputs. But a counterflow arrangement, as in figure 1a, runs into no fundamental limit on transfer; practical limits are set by the intimacy of the vessels, flow rates, the conductivity of blood and vessel walls, and the outer insulation of the exchanger. Exchange is not limited to heat – diffusion, again, follows the same rules as conduction – and countercurrent exchangers conserve such substances as dissolved oxygen and water.

Figure 1b shows a device with which students in a course I once taught explored the operation of such exchangers. In practice they were asked to compare two, a countercurrent one in which flows ran in opposite directions (as in the figure) and one in which reversing a pair of connections made flows run concurrently. We quantified their deficiencies as the difference in temperature between input (T_{in}) and output (T_{out}) divided by the overall temperature difference between hot and cold ends (ΔT); subtraction from unity expressed data as exchange efficiency, e_e:

$$e_e = 1 - \frac{T_{out} - T_{in}}{\Delta T}. \tag{1}$$

For both, plots of efficiency against flow speed showed that, as expected, faster flow reduced the effectiveness of exchange while slower flow gave better performance. But as flow speed decreased, the concurrent exchanger never quite reached 50% efficiency while, even this crude countercurrent device often exceeded 90%.

The recognition of countercurrent exchangers in organisms has a curious and instructive history. That the large arteries and veins of our appendages commonly lie close together had been recognized for over three centuries before the suggestion, by the father of physiology, Claude Bernard, in 1876, that the combination might function as a heat exchanger. And early anatomists noted the wide distribution among other animals of local arrays of vessels branched to form, in cross section, networks of intermingled arteries and veins. They called such a structure a "rete mirabile" (plural 'retia'), literally a "wonderful net", or a "red gland" for the colour imparted by all that blood. Among others, Francesco Redi (1626–1697), still remembered for his experimental evidence against the spontaneous generation of maggots from meat, recognized retia. J S Haldane (father of the more flamboyant J B S Haldane) in his classic book, *Respiration* (1922), had the right idea about the fish swimbladder rete (an exchanger of dissolved gas, not heat). He drew an analogy with "a regenerating furnace, where the heat carried away in the waste gases is utilized to heat the incoming air".

Somehow the common function of these retia escaped notice. Why? Traditional anatomists did not think in either functional or non-biological terms. Physiologists, with only rare exceptions until recently, focused on particular functions and particular animals, mainly humans, who happen to lack blatant examples of such exchangers. Take your pick of explanations. But once someone drew sufficient attention to the basic function of a rete, practically every known instance was quickly reexamined and assigned a functional role. Variants appeared, as did exchangers of less definitive anatomical character and less efficient operation. For instance, two veins (venae comitantes) surround the brachial artery of our upper arms, forming the exchanger noted by Bernard. The trio develops a lengthwise thermal gradient, though, of only about 0·3°C cm^{-1}, and we conserve more heat by shunting blood away from superficial vessels (Bazett *et al* 1948) than by its action.

Retia, then, have long been known; how they worked as countercurrent exchangers that could conserve either heat or diffusible molecules was first brought to general attention (one wonders whether the word 'discovered' applies) by an especially creative physiologist, Per Scholander (1905–1980) in the 1950's. He credited Haldane, who credited Redi and others. His 1957 article in *Scientific American* seems to have provided that catalyst for the transition from obscurity to fashion. The first formally described function was not heat exchange but transferring dissolved gas in the vessels supplying the swimbladder of deep sea fish (harkening back to Haldane); the device allowed them to secrete and maintain gas in the bladder, gas that pressures of up to several hundred atmospheres should return to the blood, gills, and then ocean (Scholander and Van Dam 1954). The flukes and tail fins of small whales provided the first definitive

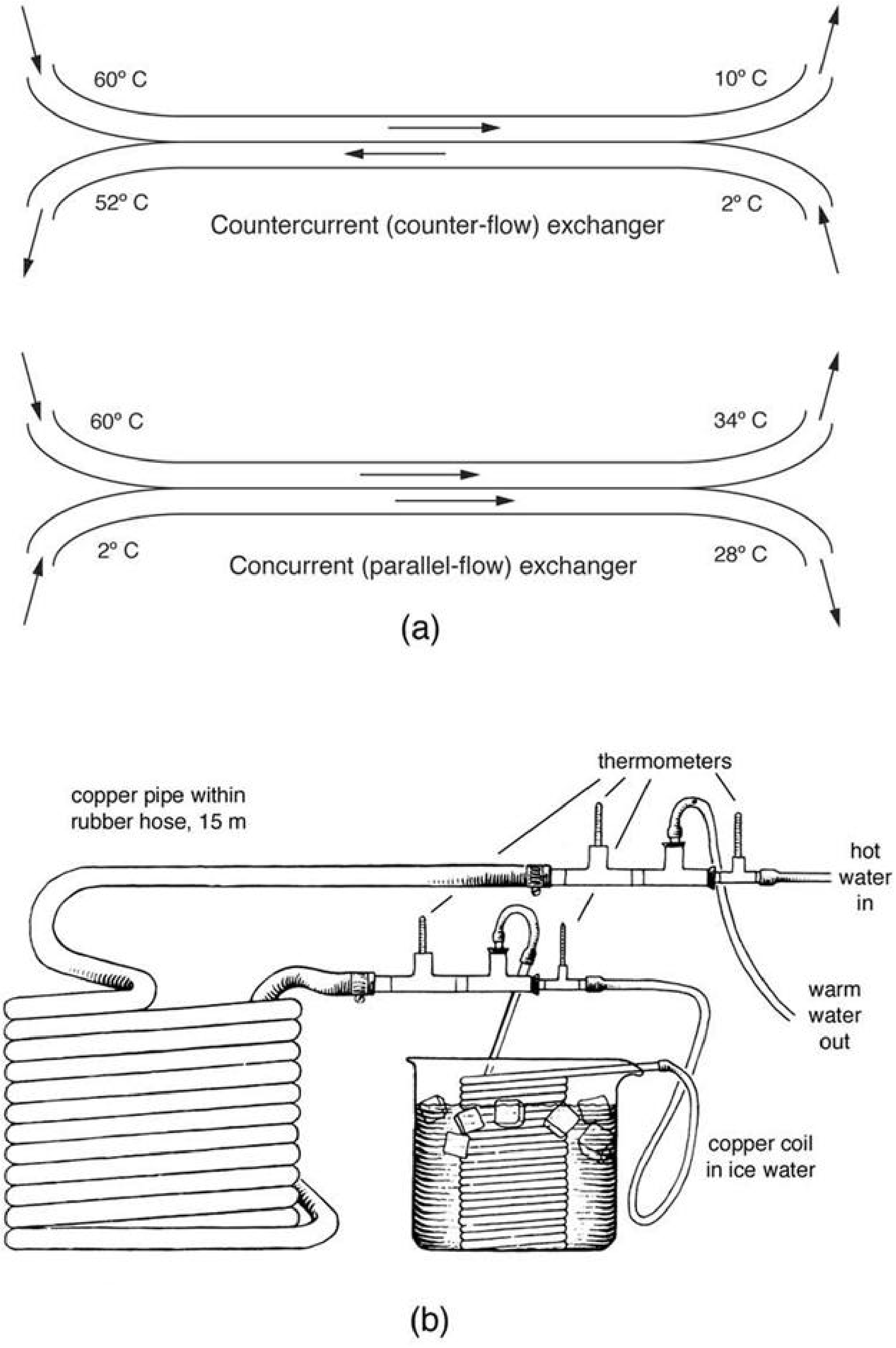

Figure 1. (**a**) Two heat exchangers, one in which fluid in the two channels flows in opposite direction and another in which it moves in the same direction. The temperatures represent typical results obtained by students using the device below. (**b**) A device that can be used as either a countercurrent (as here) or a concurrent exchanger. It consists of axial and annular channels and is made of ordinary flexible copper household plumbing, about 1·5 cm in diameter, rubber automotive heater hose, about 3 cm in diameter, copper plumbing fittings, and laboratory stoppers, thermometers, and tubing.

examples of heat exchangers (Scholander and Schevill 1955). Blubber, noted earlier, provides superb insulation, but thickly coated appendages would be ineffective as propulsors. Exchangers allow these animals to supply effectively cold-blooded fins with blood from an otherwise warm-blooded body and to do so without a futile investment of metabolic energy in heating the global ocean.

Highly effective countercurrent heat exchangers have now been described in the bases of the appendages of sloths, anteaters and some lemurs (Scholander and Krog 1957), the legs of wading birds (Scholander 1955; Kilgore and Schmidt-Nielsen 1975), the tails of muskrats (Irving and Krog 1955; Fish 1979), beavers, and manatees (Rommel and Caplan 2003), the legs of leatherback turtles (Greer *et al* 1973), the testicular blood supply of marsupials, sheep (Barnett *et al* 1958), bulls (Glad Sorensen *et al* 1991), and dolphins (Rommel *et al* 1992). They isolate the warm, dark, lateral muscles of large, fast-swimming tuna and mako sharks from the colder water passing along the body and across the gills (Carey and Teal 1966, 1969; Dewar *et al* 1994). Gazelles, sheep, and some other ungulates keep their brains from getting as hot as the rest of their heat-stressed bodies with a carotid rete, in which ascending arterial blood is cooled by venous blood coming from evaporatively cooled nasal passages (Baker and Hayward 1968). Honeybees and some other Hymenoptera isolate their abdomens from their hotter thoraces in flight with exchangers in their narrow, wasp-waist petioles (Heinrich 1996). Most of these exchangers can be bypassed by opening shunting vessels, so an animal can use an appendage as a heat dissipation device when (usually during locomotion) needed.

All the preceding countercurrent exchangers operate as steady-state devices. Unsteady versions that briefly store heat occur in both mammals and birds as well, again a scheme whose wide use was evident only after recognition of the first. Here the nasal passages of a North American desert rodent, the kangaroo rat, provided the initial case. Jackson and Schmidt-Nielsen (1964) showed that during exhalation heat moved from the air stream to the walls of the passages, so air left an animal near – in a dry atmosphere slightly below – ambient rather than body temperature. During inhalation, heat moved from passage walls to air, warming it and cooling the walls. In desert rodents its primary function appears to be water conservation, with over 50% of respiratory water loss (their principal mode of leakage) avoided by this condensation during exhalation and revaporization during inhalation. But they economize on heat as well, in amounts significant relative to overall metabolic rates, recapturing over 60% of the energy used to heat and humidify inhaled air (Schmidt-Nielsen 1972). Camels use their enormously surface-endowed nasal turbinates in the same manner; for them, concomitant thermal economizing may be detrimental rather than advantageous (Schmidt-Nielsen 1981).

Both children and adult humans exhale air close to core temperature. I wonder, though, about neonatal humans. My son, when about a week old (a smaller-than-average baby who is now a larger-than-average adult), seemed to be exhaling air that was quite a lot cooler than what came out of my own nose.

4. Buffering fluctuations through short term storage

We may be less immediately aware of the problems of temperature variation than of inopportune temperature *per se*. Our large size buffers us from changes in ambient temperature and radiant regime, and our mobility usually enables us to quickly reach more salubrious locations. Our perceptual world remains distant from that of a marine snail caught on a large rock in summer sunlight at low tide or of a sun-lit leaf on a tree when the normally ubiquitous air movements briefly abate. But, as Denny and Gaines (2000) remind us, the distribution of organisms more likely reflects local extremes, particularly temporal ones, than it does regional averages.

What constitutes a temporal extreme, though, depends on size. As large creatures we can ignore most events that last only seconds and need not take seriously most minute-scale phenomena. I can move a finger through a candle flame without discomfort, much less injury; and I recall watching students on a Canadian campus going without coats from one building to an adjacent one despite a temperature of about – 30°C. At the same time, few, even well bundled, waited in the open for buses. So we encounter yet another problem of scale, that ever-lurking consideration in each of these essays.

While minute-to-minute fluctuations in heat load may not matter to large animals, variation on scales of hours clearly do. A particularly interesting case is that of a camel in a hot desert, faced with problems of both too much heat and too little water, something investigated by Schmidt-Nielsen *et al* (1957) and later put into a general context (Schmidt-Nielsen 1964). Comparison of normal and shorn animals showed that fur reduces both heat gain and evaporative water loss. Beyond using fur, camels (dromedaries in North Africa) take peculiar advantage of the predictability of the main temporal fluctuation that affects them. When their access to water is limited, they permit their core temperatures to rise from about 34°C to 40°C during the day, secure in the knowledge that night will follow with cooler air and (usually) an open sky. They thereby reduce evaporative water loss (less sweating, mainly) almost three-fold and halve overall water loss.

One might expect that only large creatures can play this particular game – a few large mammals, perhaps some of the more massive cacti (Nobel 1999 calculated time constants for the latter of several hours). Remarkably, at least one group of small desert succulents heats slowly enough to do so as well. These so-called stone plants (*Lithops* spp.) live largely buried in the soil of the Namib desert and the Karoo scrubland of South Africa; they protrude only about 2 mm above the surface but extend downward about 30 mm, as in figure 2. A translucent window on the top of each of the paired leaf-analogs admits light into the interior, with the photosynthetic tissue (the chlorenchyma) lining the bottom somewhat as our retinas line the inner rear surface of our eyeballs. Turner and Picker (1993) found that daily temperature cycled between extremes of 12°C and 46°C, as very nearly did plant surface, plant interior, and the surrounding soil 1 cm below the surface – all rose rapidly though the morning, peaked in the afternoon, and slowly dropped through the night. That may be more variation than experienced by a camel, but direct solar exposure without coupling to the surrounding soil would make matters worse – plants surrounded by styrofoam insulation became considerably hotter than those in full contact with the soil.

Thus by combining its thermal mass with that of the surrounding soil, *Lithops* buffers its daily temperature changes and, most importantly, reduces peak daytime temperatures. In addition, it takes advantage of the steep vertical thermal gradient in the soil, coupling not to the hotter surface but to the cooler soil a short distance beneath and by locating its most metabolically active tissues well down from that hot surface.

To emphasize the connection between the size of the system and the relevant temporal scale of fluctuations we might return to broad leaves during periods of what we think of as still air. Convection, whose magnitude depends strongly on air speed, provides a major avenue of heat transfer. The speed of “still air” fluctuates rapidly and continuously, the result of passing turbulent structures and local convection. And leaves, with lots of surface and little volume, are effectively small and thus have very rapid thermal responses.

Some years ago I tried to get a sense of a leaf’s thermal situation on a still, sunny summer afternoon with a model leaf mounted near the top of the forest canopy. The model, of cellulose acetate with black ink dots, had both the shape and thickness of a sun leaf of white oak (*Quercus alba*) – testing in the laboratory assured me that its absorptivity and time constant came close to those of real leaves. A tiny bead thermistor glued to its lower surface monitored mid-blade temperature, while a heated thermistor tracked adjacent air movement. Figure 3 shows a typical pair of tracks. Air temperature remained almost constant, and model temperature invariably exceeded it. The temperature of the leaf model was anything but constant; when the wind dropped, it rose, with only a short lag. One rarely, if ever, thinks of leaf temperature as such a wildly fluctuating variable; once alerted, one wonders about the metabolic implications of its rapid and continuous change.

Nobel (1999) calculated a time constant below 20 s for a broad leaf, quite consistent with the data from my model in figure 3. As one can see from that figure, even a modest increase in such a time constant would yield significant thermal buffering, so rapidly does air speed change. Thus improved protection against temperature extremes would require vastly less mass than a camel or

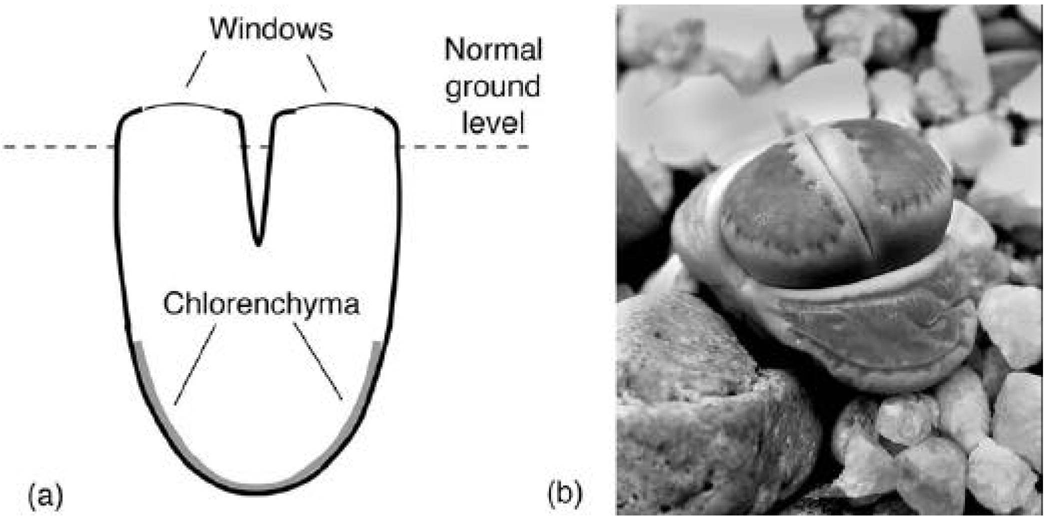

Figure 2. (**a**) Diagrammatic cross section of a mature *Lithops*. (**b**) *Lithops*, as grown in a greenhouse and less deeply buried than it would be in nature. The above-ground portion is about 2 cm across.

stone-plant-plus-soil. And thus increased thickness might well constitute a specific adaptation to assure lower peak temperatures during brief episodes of especially low wind – as opposed to an incidental consequence of some other functional demand. Plants with small, thick leaves have long been termed 'xerophytes' for their prevalence in dry habitats; the leaf structure is then 'xeromorphic'. Perhaps the plants might instead be called 'thermophytes', the lack of local water for evaporative cooling simply contributing to the thermal challenge they face.

A functional explanation that focuses primarily on heat and only secondarily on water might explain the peculiar prevalence of plants with xeromorphic leaves in some well-watered places such as the swampy bogs of eastern North America. Traditional explanations invoke some kind of physiological dryness or deficiencies of nitrogen or calcium. But the results of a comparative morphological study by Philpott (1956) are consistent with a thermal rationale. She matched leaves of 19 species from forest-surrounded bogs in Carolina (called 'pocosins' in the region) with those of 14 related plants from the Appalachian mountains directly inland. Whether looking at specific genera or at averages, the bog plants had smaller and thicker leaves. Small size would give better convective coupling to the surrounding air and therefore less deviation from ambient temperature; thicker leaves would heat more slowly during lulls. Thus the low wind and high humidity that makes these bogs notoriously unpleasant for people may be just the factors that challenge the local plants.

Somewhat more direct evidence that leaves may decrease size and increase thickness to lower peak temperatures through short term heat storage comes from work of Kincaid (1976). He collected holly (*Ilex*) leaves of a variety of species that experience different thermal extremes and exposed them to a wide variety of regimes in a very low speed wind tunnel in my laboratory. Among other manipulations, he subjected radiantly-heated leaves to pulses of moving air, alternating 10 s of still air ($< 0{\cdot}01$ m s^{-1}) with 10 s of winds of 0·1 and 0·5 m s^{-1}, conditions of light and air movement that he showed were in a range they might normally encounter on a hot, windless day. Larger and thinner leaves heated significantly faster and further during lulls than did smaller and thicker ones. The variation in behaviour among the different species in the wind tunnel correlated satisfyingly with estimates of the importance of short term heat storage from field data.

5. The possibility of counterconvection

In examining how the physical world affects the adaptations and aspirations of organisms, this series of essays

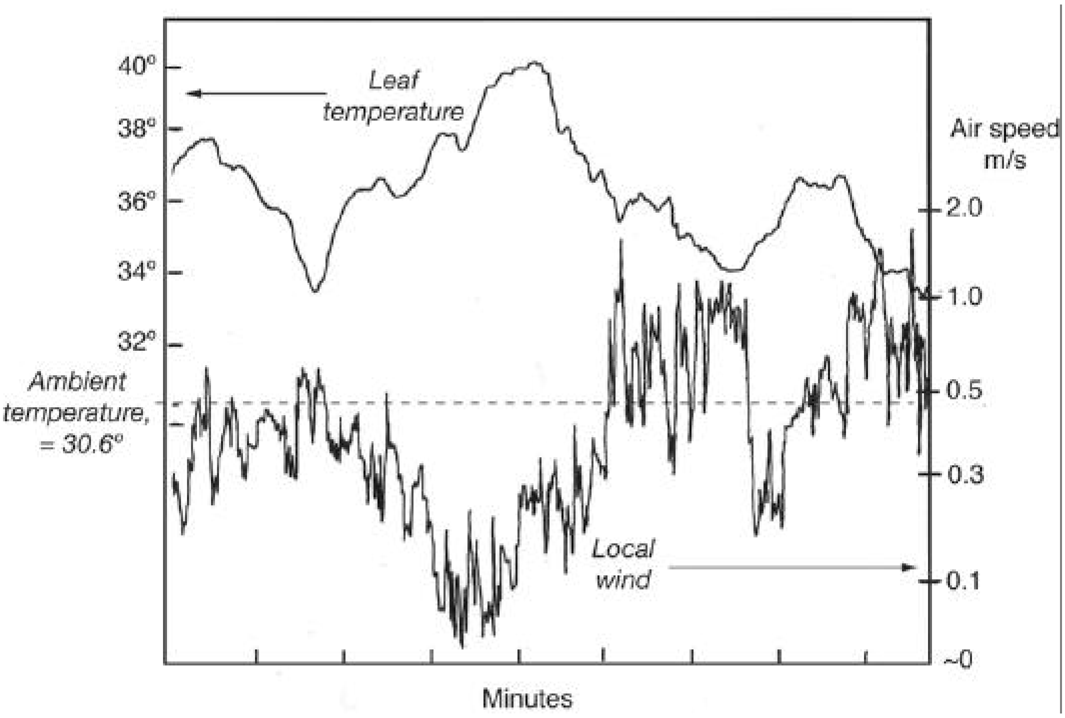

Figure 3. Representative data for air temperature, mid-blade leaf temperature, and wind speed for a model sun leaf of white oak, *Quercus alba*, on a typically windless summer afternoon in the Carolina piedmont. Note the non-linear scale for wind speed.

attempts general perspectives rather than conventional reviews. I want to include in my domain physical devices as yet unknown in living systems – one can, as an alternative to our normal search for functional explanations of specific features of organisms, look for organisms that use some hypothetical but plausible device. Per Scholander's recognition of the commonness, diversity, and general function of biological countercurrent exchangers – as well as much else he did – certainly shows the utility of the approach. In a sense, he played Hamlet to us Horatios; as Shakespeare put it, "There are more things in heaven and earth, Horatio, than are dreamt of in your philosophy".

Consider a related scheme not yet known in a biological system. Countercurrent devices combine convective with conductive or diffusive transfer – fluid moves axially through pipes while heat or molecules conduct or diffuse laterally through the fluid and across the walls separating the pipes. These two modes of moving heat or molecules can be combined in another way, one that could permit heat transfer to be driven below what we have been treating as a baseline, pure conduction – achieving, in effect, perfect insulation. Credit for asking about its possible roles should go to an engineer, the late Lloyd Trefethen. He described the scheme and asked me whether it found use; I could offer no specific instance. Perhaps some reviewer or reader will recognize a case of what has been called 'counterconvection', It operates in the following way – focusing on heat transfer, but bearing in mind that diffusive material transfer and a concentration gradient could replace heat conduction and a temperature gradient.

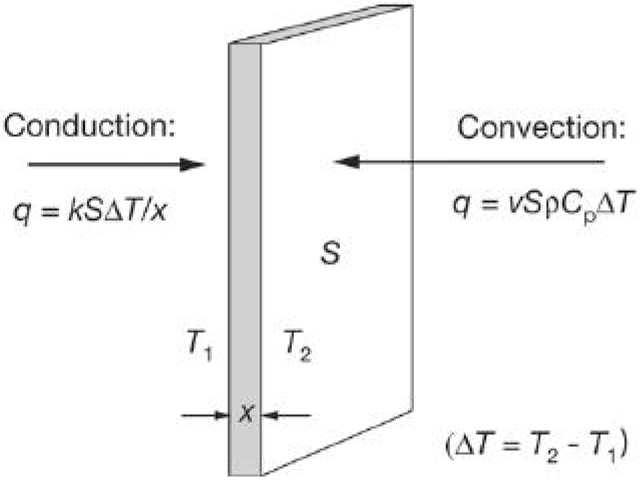

Figure 4. Heat conduction, left to right, and convection, right to left, in a counterconvective arrangement. *S* and *x* are slab area and thickness respectively, *k* is thermal conductivity, *v* is cross-slab flow speed, *r* is fluid density, and C_p is the heat capacity of the fluid.

Imagine a porous, conductive barrier between two compartments that differ in temperature, as in figure 4. Heat ought to be conducted from warmer to cooler side. That conduction, though, is exactly offset by fluid forced through the barrier, so that all the heat that would otherwise be conducted down the thermal gradient gets transferred to fluid flowing up that gradient. And fluid flowing up the thermal gradient, now preheated, no longer cools the warmer compartment as it enters. In effect, heat moves down a thermal gradient while fluid moves down a pressure gradient, with conduction in one direction balanced by convection in the other. Balance will be achieved when

$$\frac{k}{x} = vrC_p, \tag{2}$$

where *k* is thermal conductivity, *x* is the thickness of the barrier, *v* is flow speed, *r* is fluid density, and C_p is heat capacity (or specific heat at constant pressure).

The principal difficulty, to provoke proper skepticism at the start, is that the mechanism does not (at least as I see it) lend itself to operation as a closed cycle. Fluid will accumulate in one compartment, so draining it in any ordinary way will offset anything gained. Actively pumping fluid will leave the system still worse. This suggests examining systems where fluid ordinarily enters or leaves and can be secondarily pressed into counterconvective service or systems that operate only part time, perhaps during periods of particular environmental stress.

Does the possibility pass quantitative muster? Consider two cases in which hypothetical organisms find themselves in dangerously hot circumstances:

(i) A spherical animal with 1 m^2 of outer surface (0·56 m in diameter) and an insulating layer of fat 0·01 m thick is exposed to an outside temperature 10°C above body temperature; high humidity or a liquid external medium prevents evaporative heat transfer. If fat's conductivity is 0·21 W m^{-1} K^{-1}, Fourier's law for conduction predicts heat entry at 210 W. Expelling it in the form of water, with a heat capacity of 4·2 kJ kg^{-1} K^{-1} would take only 5 ml s^{-1}. Still, that amounts to 18 l h^{-1}, which would use up the entire volume of the animal in just a little over 5 h, making the scheme an unattractive long-term fix. That 210 watt heat entry would normally cause the animal to heat (initially at least) at about 1·9 K h^{-1}, which ought to be tolerable for short periods. So counterconvection would not work well for long periods and would be unnecessary for short periods. Still, the scheme cannot be dismissed as impossible for all scales of size, time, and temperature.

(ii) Another spherical animal of the same size and faced with the same temperature difference has no insulating fat; instead it has a fur coat of the same (0·01 m) thickness. Heat conducts inward through the fur while perspired liquid water is wicked outward and then obligingly disappears with no additional thermal consequences. Conductivity is now 0·025 W m^{-1} K^{-1}, and heat will enter at 25 W. That requires an outward flow of water of only 0·6 ml s^{-1} or 2·15 l h^{-1}. The animal thus contains about 44 h supply, enough, one might guess, to deal with a hot afternoon in the secure knowledge that night will follow in a few hours. But one further calculation puts this sanguine scenario in a less favourable light. Dealing with an input of 25 W by evaporative cooling, given water's heat of vaporization of 2·44 MJ kg^{-1}, would take only 0·037 l h^{-1}, about 60 times less. Thus the scheme, while possible, makes sense only where evaporative cooling cannot be relied upon.

What should we conclude? While we should not dismiss the possibility of counterconvection, the requirements for it to be worthwhile turn out to be daunting. Still, conduction through a material of low conductivity and flow through a porous barrier, the requirements for it to happen, are biologically ordinary. One can produce enough bulk flow through such a barrier with only a modest pressure gradient, and organisms often either absorb or excrete liquid water for other purposes at appropriate rates.

Acknowledgements

I thank Dwight Kincaid and Knut Schmidt-Nielsen for steering me to important sources of information and the Duke University greenhouse for access to *Lithops*. I remain indebted to the late Jane Philpott and Lloyd Trefethen for stimulating my interest in aspects of the present topic, and I treasure the memory of my brief encounters with Pete Scholander.

References

Baker M A and Hayward J N 1968 The influence of the nasal mucosa and the carotid rete upon hypothalamic temperature in sheep; *J. Physiol., London* **198** 561–579

Barnett C H, Harrison R J and Tomlinson J D W 1958 Variations in the venous system of mammals; *Biol. Rev.* **33** 442–487

Bazett H C, Love L, Newton M, Eisenberg L, Day R and Forster R II 1948 Temperature changes in blood flowing in arteries and veins in man; *J. Appl. Physiol.* **1** 3–19

Bernard, C 1876 *Leçons sur la chaleur animale, sur les effets de la chaleur et sur la fièvre* (Paris: J-B Baillière)

Carey F G and Teal J M 1966 Heat conservation in tuna fish muscle; *Proc. Natl. Acad. Sci. USA* **56** 1464–1469

Carey F G and Teal J M 1969 Mako and porbeagle: warm-bodied sharks; *Comp. Biochem. Physiol.* **28** 199–204

Denny M and Gaines S 2000 *Chance in biology: Using probability to explore nature* (Princeton, NJ: Princeton University Press)

Dewar H, Graham J B and Brill R W 1994 Studies of tropical tuna swimming performance in a large water tunnel. II. Thermoregulation; *J. Exp. Biol.* **192** 33–44

DuBois E F 1939 Heat loss from the human body; *Bull. N.Y. Acad. Med.* **13** 143–173

Dunkin R C, McLellen W A, Blum J E and Pabst D A 2005 The ontogenetic changes in the thermal properties of blubber from the Atlantic bottlenose dolphin *Tursiops truncatus*; *J. Exp. Biol.* **208** 1469–1480

Fish F E 1979 Thermoregulation in the muskrat (*Ondatra zibethicus*): the use of regional heterothermia; *Comp. Biochem. Physiol.* **A64** 391–397

Glad Sorensen H, Lambrechtsen J and Einer-Jensen N 1991 Efficiency of the counter current transfer of heat and ^{133}xenon between the pampiniform plexis and testicular artery of the bull; *Int. J. Androl.* **14** 232–240

Greer A E, Lazell J D and Wright R M 1973 Anatomical evidence for a countercurrent heat exchanger in the leatherback turtle (*Dermochelys coriacea*); *Nature (London)* **244** 181

Haldane J S 1922 *Respiration* (New Haven, CT: Yale University Press)

Heinrich B 1996 *The thermal warriors: Strategies of insect survival* (Cambridge, MA: Harvard University Press)

Irving L and Hart J S 1957 The metabolism and insulation of seals as bare-skinned mammals in cold water; *Can. J. Zool.* **35** 497–511

Irving L and Krog J 1955 Temperature of the skin in the Arctic as a regulator of heat; *J. Appl. Physiol.* **7** 355–364

Jackson D C and Schmidt-Nielsen K 1964 Countercurrent heat exchange in the respiratory passages; *Proc. Natl. Acad. Sci. USA* **51** 1192–2297

Kilgore D L Jr and Schmidt-Nielsen K 1975 Heat loss from ducks' feet immersed in cold water; *Condor* **77** 475–478

Kincaid D T 1976 *Theoretical and experimental investigations of Ilex pollen and leaves in relation to microhabitat in the southeastern United States*, Ph.D. Thesis, Wake Forest University, Winston-Salem, NC, USA

Nobel P 1999 *Physicochemical and environmental plant physiology*, 2nd edition (San Diego, CA: Academic Press)

Philpott J 1956 Blade tissue organization of foliage leaves of some Carolina shrub-bog species as compared with their Appalachian mountain affinities; *Bot. Gaz.* **118** 88–105

Rommel S A and Caplan H 2003 Vascular adaptations for heat conservation in the tail of Florida manatees (*Trichechas manatus latirostris*); *J. Anat.* **202** 343–353

Rommel S A, Pabst D A, McLellan W A and Potter C W 1992 Anatomical evidence for a countercurrent heat exchanger associated with dolphin testes; *Anat. Rec.* **232** 150–156

Schmidt-Nielsen K 1964 *Desert animals: Physiological problems of heat and water* (Oxford, UK: Oxford University Press)

Schmidt-Nielsen K 1972 *How animals work* (Cambridge, UK: Cambridge University Press)

Schmidt-Nielsen K 1981 Countercurrent systems in animals; *Sci. Am.* **244** 118–129

Schmidt-Nielsen K, Schmidt-Nielsen B, Jarnum S A and Houpt T R 1957 Body temperature of the camel and its relation to water economy; *Am. J. Physiol.* **188** 103–112

Scholander P F 1955 Evolution of climatic adaptations in homeotherms; *Evolution* **9** 15–26

Scholander P F 1957 The wonderful net; *Sci. Am.* **196** 96–107

Scholander P F and Krog J 1957 Countercurrent heat exchange and vascular bundles in sloths; *J. Appl. Physiol.* **10** 405–411

Scholander P F and Schevill W E 1955 Counter-current vascular heat exchange in the fins of whales; *J. Appl. Physiol.* **8** 279–282

Scholander P F and Van Dam C L 1954 Secretion of gases against high pressures in the swimbladder of deep sea fishes; *Biol. Bull.* **107** 247–259

Scholander P F, Walters V, Hock R and Irving L 1950 Heat regulation in some arctic and tropical mammals and birds; *Biol. Bull.* **99** 237–258

Turner J S and Picker M D 1993 Thermal ecology of an embedded dwarf succulent from southern Africa (*Lithops* spp: Mesembryanthemaceae); *J. Arid Environ.* **24** 361–385

Vogel S 1968 "Sun leaves" and "shade leaves": differences in convective heat dissipation; *Ecology* **49** 1203–1204

Vogel S 2005 Living in a physical world IV. Moving Heat Around; *J. Biosci.* **30** 449–460

*e*Publication: 22 November 2005

Series

Living in a physical world
VI. Gravity and life in the air

STEVEN VOGEL

Department of Biology, Duke University, Durham, NC 27708-0338, USA

(Fax, 919-660-7293; Email, svogel@duke.edu)

1. Introduction

In our perceptual world, no physical agency imposes itself with greater immediacy than does gravity. We depend on it to walk or run; it injures us if we trip. It makes each of us about half a centimeter shorter at the end of each day than when we first arise. Our flesh sags as we age; more slowly, the glass of a window thickens at the bottom and thins at the top. We dream of escaping its constant crush, although our recent experiences in orbiting spacecraft reveal an addiction with a difficult withdrawal. Physicists may regard the gravitational attraction between two objects as the universe's definitional weak force, but to us large, terrestrial creatures it feels anything but weak.

Since the consequences of gravity depend on one's size, scaling will loom at least as large in this and then the next as in any of the preceding essays. Even more important than the ways gravity's effects, scale will be another message–the surprisingly wide range of biological situations in which it plays some role. One knows that no massless world exists; I would argue that a weightless world is almost as hard to imagine.

That contrast, mass versus weight, needs a few words. Newtonian mechanics lumped two distinct kinds of mass, inertial mass and gravitational mass. Establishing the basis of their apparent equality awaited the 20th century. An inertial mass resists acceleration, as expressed in Newton's first and second laws; quantitatively, mass equals force divided by acceleration–the familiar $F = ma$. Weight follows from the other kind of mass. A gravitational mass attracts any other mass, exerting a force equal to the product of the two masses, divided by the square of the distance between them, times a universal gravitational constant, in proper SI units, 6.67×10^{-11} N m^2 kg^{-2}. (One should avoid using "gravitational constant" for the acceleration of gravity at the surface of the earth, commonly designated g.) In our world that other mass is that of the earth itself, 5.976×10^{24} kg, and the basic distance is that from the earth's surface to its center of mass, 6,370 km. These data give $g = 9.8$ m s^{-2}; with $F = mg$ we can then convert mass (kilograms) to terrestrial weight (newtons).

Our lack of intuitive feeling for the difference between mass and weight just reflects inexperience with situations in which the constant of proportionality differs from the terrestrial 9.8 m s^{-2} or, because of our buoyancy, an effective value near zero when submerged. Think back to video images of astronauts ambulating on the surface of the moon. Unsurprisingly, they adopted a rather bouncy gait when going straight ahead; one could easily fail to notice the greater forward tilt needed to get going. Turning, though, looked glaringly unfamiliar with a greater lean needed to change direction. Greater tilts to start, stop, and turn provided sufficient force so these weight-deprived individuals could accelerate their unchanged masses, to apply what weight they retained with sufficient effect. Fortunately, the human neuromuscular machine turns out to cope remarkably well with this completely novel six-fold reduction in weight.

2. The other forces that matter

Besides that of gravity, life contends with a diversity of forces–for instance that inertial force from the unwillingness of mass to accelerate; the force of surface tension from the cohesion of liquids in gases or in other, immiscible, liquids; and viscous force from the resistance of both liquids and gases to shearing motion. In most situations, though, we only need worry about one or two forces, and the first item in an analysis commonly consists of identifying the forces at work and their relative importance. An extreme example should emphasize the point.

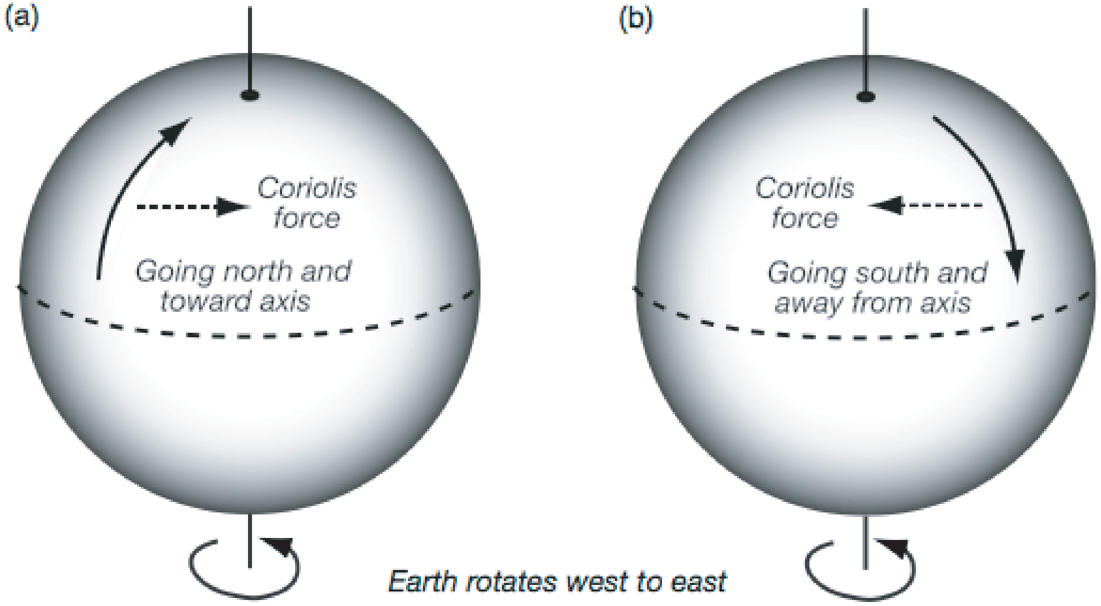

Figure 1. The origin of the Coriolis force for something moving **(a)** northward and **(b)** southward, in the northern hemisphere. While the force may be eastward and then westward relative to the earth, it remains to the right with respect to the mover.

Some time ago, I was asked to evaluate a claim that the asymmetries of the mammalian body, in particular of our own, could be traced to effects of the Coriolis force from the earth's rotation on the evolution of terrestrial animals. That force (really a pseudoforce, like so-called centrifugal force) results from the spherical rather than cylindrical shape of the rotating earth. Thus (as in figure 1) an object in the northern hemisphere moving north must move inward toward the earth's axis of rotation as well. Angular momentum being conserved, it should rotate faster–the effect will be felt as an eastward force, a force to the right of its path. When moving south, the object will move outward and thus try to rotate more slowly–an effect now felt as a westward force, but still to the right of its path. Clearly a slightly sturdier right leg ought to confer an advantage, making an animal, one might say, a leg up. The same argument was applied most ingeniously to our many anatomical asymmetries.

I found against the plaintiff, so to speak, making my case by comparing the magnitude of the Coriolis force with that of the gravitational force. The former is twice the product of the object's mass (m), the speed at which it moves north or south (v), the earth's angular velocity (Ω), and the sine of the latitude (ϑ). Mass times gravitational acceleration (mg) gives the gravitational force. In their ratio, mass cancels, and we get

$$\frac{2v \cdot \Omega \cdot \sin\vartheta}{g}. \tag{1}$$

A most-favourable-possible-scenario might consider an animal living at 45º latitude and spending its life going north or south at 1 m s^{-1}. Under these conditions the ratio is 1:100,000. That seemed to me to offer evolution precious little advantage with which to work; for more evolutionarily reasonable lower speeds and latitudes, the ratio would be still less auspicious. In short, little about our persons can be attributed the Coriolis force–however dramatic its effects on, for instance, weather patterns.

[As noted by a reader of the manuscript, the equation should not be applied in this simple form to the large bodies of air responsible for our weather. It ignores buoyancy, tacitly assuming that the density of the mass at issue far exceeds that of the atmosphere. Persson (1998) provides an engaging introduction to Gaspard Gustave de Coriolis (1792–1843) and his force.]

While the terrestrial Coriolis force may be summarily dismissed relative to gravity, many other forces cannot. Hydrostatic and aerostatic forces squeeze or expand organisms. Tensile, compressive, and shearing forces variously distort their shapes. The viscosity and dynamic pressures of flows impose both drag and lift. The inertia of fluids can exert major transient forces, as when the surface of a body of water is slapped–by a hand or, more significantly, when a basilisk lizard runs across a stream (Glasheen and McMahon 1996). Transpiring trees as well as water striders depend on surface tension. And so on. What most often determines the practical consequences of gravity is its magnitude relative to the other forces at work.

The engineering community, the fluid mechanists in particular, have long used a variety of dimensionless ratios of one force to another to evaluate their relative importance. Gravitational force contributes to many of them, either as numerator or denominator depending on the prejudice of the particular field in which the ratio first found use–which force carried the load and which constituted a nuisance. For instance, the Bond number, below, has mainly been used for gravity-driven flows with interfaces in porous media, so gravity makes the system go, while surface tension acts as a brake. Thus using gravitational force as numerator and surface tension as denominator makes high rather than low values desirable. If, instead, the ratio had been contrived by a biologist concerned with animals supported by surface tension atop a pond, gravitational force would have dropped to the denominator.

Among the dimensionless ratios that include gravitational force (from Weast *et al* 1987; new editions of the *Handbook of Chemistry and Physics* no longer give dimensionless ratios)….

(i) Bond number–as mentioned, gravitational force to surface tension force:

$$Bo = \frac{(\rho_o - \rho_m) l^2 g}{\gamma}. \tag{2}$$

ρ_o and ρ_m are the densities of object and liquid medium respectively, l is a characteristic length of the system, the choice depending on the particular phenomenon at hand, and γ is surface tension.

(ii) Froude number–inertial force to gravitational force:

$$Fr = \frac{v^2}{gl}. \tag{3}$$

The choice of l, again, depends on the system.

(iii) Bagnold number–drag to gravitational force:

$$Ba = \frac{3 C_d \rho_m v^2}{4 d \rho_o g}. \tag{4}$$

C_d is the object's drag coefficient and d its diameter. It resembles the Froude number because the underlying formula for drag ($1/2\ C_d \rho_m S v^2$, with S for projecting area normal to flow, as in eq. 12, below) tacitly presumes it an inertial force and ignores buoyancy by assuming an object much denser than the medium.

(iv) Grashof number–buoyant force to viscous force:

$$Gr = \frac{\rho^2 g \beta (\Delta T) l^3}{\mu^2}. \tag{5}$$

β is the coefficient of thermal expansion of the fluid, ΔT the temperature difference, and μ the viscosity. The Grashof number appeared previously in essay 4 (Vogel 2005b).

(v) Galileo number–gravitational force to viscous force:

$$Ga = \frac{l^3 g \rho^2}{\mu^2}. \tag{6}$$

Since gravity underlies buoyancy, similarity between this one and its predecessor should be no surprise. Both the Grashof and Galileo numbers, as well as a few others, include as a factor the Reynolds number,

$$Re = \frac{\rho l v}{\mu}, \tag{7}$$

the ratio of inertial to viscous force–thus density and viscosity appear as second powers in both.

In all these dimensionless numbers (as well as others), the larger the system, the more important gravity becomes relative to other forces. Whether g appears as numerator or denominator, some size factor appears with it. I can think of no exception to that rule, although I hesitate to assert its universality.

3. Going up and down

Besides keeping our atmosphere from drifting away, gravity makes its outer portions squeeze down on the inner portions; thus a pressure increase accompanies an approach to the earth's surface. (Only a tiny part of that increase comes from the increase in gravitational force as the earth's center is approached.) As in any ordinary gas mixture, atmospheric density follows pressure–some consequences of altitude change result from density change, others from pressure change. In particular, pressure affects the solubility of gases in liquids. A carrier of respiratory gas such as haemoglobin, suitable for reversible binding with oxygen at one altitude, will not work as well at a very different one, and mammals adapted by ancestry (as opposed to individual experience) to high altitudes have haemoglobin variants with greater affinities for oxygen (Hall *et al* 1936).

The volume of a helium- or hydrogen-filled balloon will increase as it rises; if its buoyancy varies with volume and its drag with surface area, its ascent speed will gradually increase. Organisms, though, do not use buoyant bags to ascend in air. Still, the volume increase does matter, requiring that internal air containers either be surrounded by stretchy walls or be vented to the outside. We vent our middle ears into our respiratory passages through a pair of Eustachian tubes, and ascents and descents in aircraft or elevators with plugged tubes cause pain and temporary auditory impairment. Birds, facing the problem in more severe form, vent all their air-filled bones.

The external effects of that volume increase with altitude (or with anything else that lowers pressure) may be more important. If a patch of ground heats more than the surrounding area, the locally warmer air above it may rise. It initially forms a column, then a round bubble, and finally a torus. That rising torus, typically over a highway or plowed field, can provide an elevator for pollen, spores, seeds, and small organisms.

One might expect that any ascent will be brief, since such objects will always be descending relative to the local air and must soon fall out of the ascending torus. That need not be the case – the torus forms because air at the periphery of the bubble is slowed by the surrounding air. Thus air near the periphery descends relative to the overall structure, and air near the inner portion of the ring must rise. So something near the inside margin of the toroidal ring can fall steadily without falling out. Many birds appear to do just that, soaring in circles whose radii are smaller than the radii of the cores of the tori. That need to glide in fairly tight circles has been invoked to explain the typically short and broad wings of terrestrial – by contrast with marine – soarers. Other creatures, such as tiny spiders and insects may exploit the same opportunity by paying out long threads that partly enwrap vortices, although we lack specific documentation.

Locally warmed air rises; locally cooled air should fall. Such cold, downslope currents, mainly at night, often occur in hilly or mountainous terrain. Extreme versions go by names such as "air avalanches," 'mistrals' and 'williwaws' and may reach 40 m s^{-1}; Geiger (1965) describes ones of remarkably regular short-term periodicity.

Large scale mixing may not always suffice to keep the atmosphere uniform, and the resulting local changes in atmospheric composition can have serious physical and biological consequences. Local enrichment with a light gas such as methane will produce upward bubbles, columns, and toroids just as does local heating, if on a smaller scale and less portentously. Local enrichment with a heavy gas will, in the absence of significant wind, lead to a stable, enriched layer at ground level. A ground-level layer of carbon dioxide, not normally regarded as a serious toxin, led to the Lake Nyos disaster of 1986 in West Africa (Kling *et al* 1987). After the outgassing that began at some supersaturated spot deep within the lake, CO_2 reached sufficient concentration to cause the immediate death of the 1700 or so people of the surrounding villages.

4. Falling in air

When acting directly on the mass of an organism, gravity has consequences both more common and serious than anything resulting from changes in atmospheric pressure and density. Unrestrained, a body accelerates downward at 9.8 m s^{-2}, so

$$v = \sqrt{19.6d}. \qquad (8)$$

The impact speed after a fall of a meter will be 4.4 m s^{-1}, tolerable for most organisms under most circumstances. A 10 m fall will give an impact speed of 14 m s^{-1} and a 100m fall of 44 m s^{-1}, both ordinarily hazardous.

In reality, a body falling in air accelerates ever more gently, asymptotically approaching a state where its downward gravitational force equals its upward drag, where the Bagnold number (above) approaches 1.0. The apparent calculational simplicity, though, proves deceptive. As noted when considering trajectories in essay 2 (Vogel 2005a), C_d, the drag coefficient, varies peculiarly. For very small things falling in air – fog droplets, pollen grains, and so forth – it varies inversely with speed (more specifically, with the Reynolds number and as described by Stokes' law), while for large, fast, dense things it remains very nearly constant. In between, from ejected spore clusters, falling seeds, and small, flying insects to medium-sized flying birds – no biologically trivial realm – we find several shape and Reynolds-number dependent transitions, some of them abrupt (Vogel 1994).

Figure 2 gives terminal speeds for spheres of the density of water falling in air of sea-level density, calculated using the formulas of essay 2. (Streamlined bodies, at least ones of large or moderate size, will reach higher terminal speeds, while irregularly shaped or tumbling bodies will descend more slowly.) Bigger inevitably means faster, but for small spheres terminal speed is especially size-dependent, increasing with the square of diameter according to Stokes' law, while for large ones terminal speed increases with only the square root of diameter.

Still, one should not assume that drag must always be taken into account. In figure 3 the same equations have been used to view the approach to terminal speed for spheres of a range of sizes (again at one atmosphere and of water's density). For large bodies, long drops precede achievement of near-terminal speeds, while small ones get there so shortly after release that one can assume they accelerate instantaneously. For instance, to get within 5% of terminal speed, a 100 mm sphere needs about 100,000 mm of fall; a 10 mm sphere needs about 2,000 mm of fall; a 1 mm sphere needs about 100 mm of fall; a 0.1 mm sphere needs about 0.1 mm of fall; a 10 μm sphere needs a mere 0.01 μm of fall. A 10^4-fold decrease in diameter yields a 10^{10}-fold decrease in the dropping distance to get to 95% of terminal speed.

Alternatively, one can view (as in figure 3) the scaling in terms of the diameters of the spheres, using falling distance over diameter as a nicely dimensionless length. To get to our benchmark 95% of terminal speed, the 100 mm sphere takes 1000 diameters, the 10 mm sphere takes 200 diameters, the 1 mm sphere takes 100 diameters, the 0.1 mm

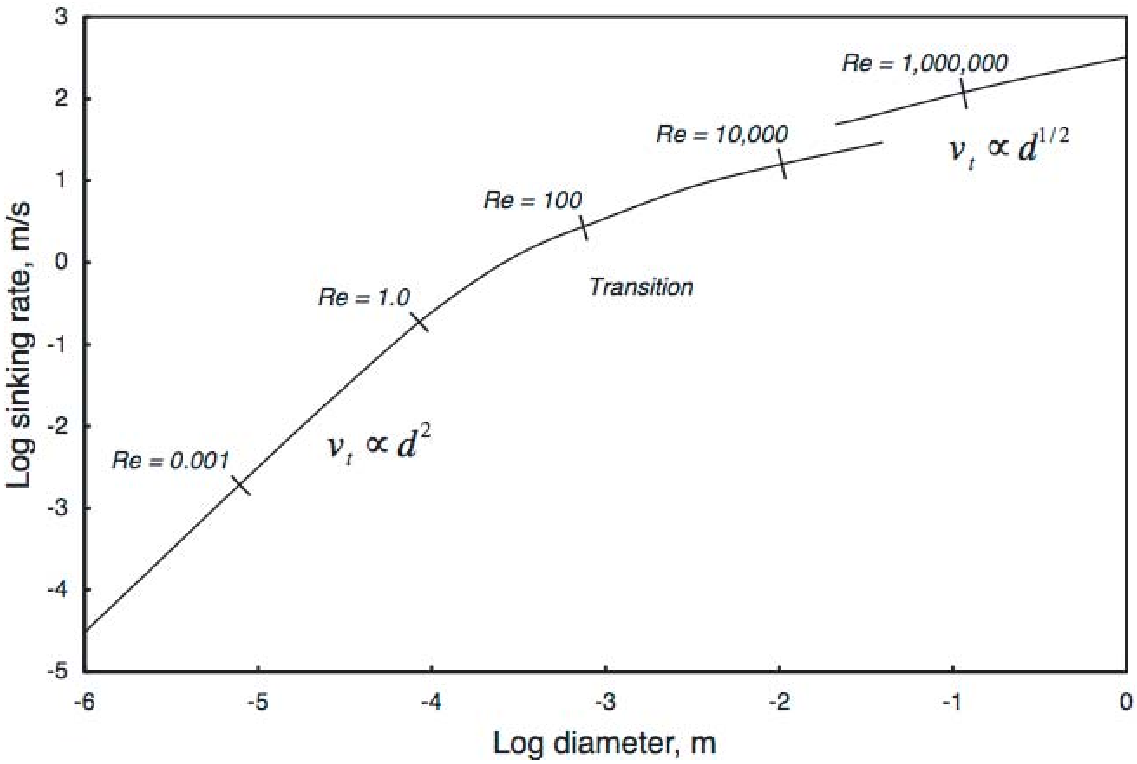

Figure 2. Terminal sinking rates of spheres without induced internal motion and of the density of water as a discontinuous function of diameter and Reynolds number (eq. 7, using the density and viscosity of air).

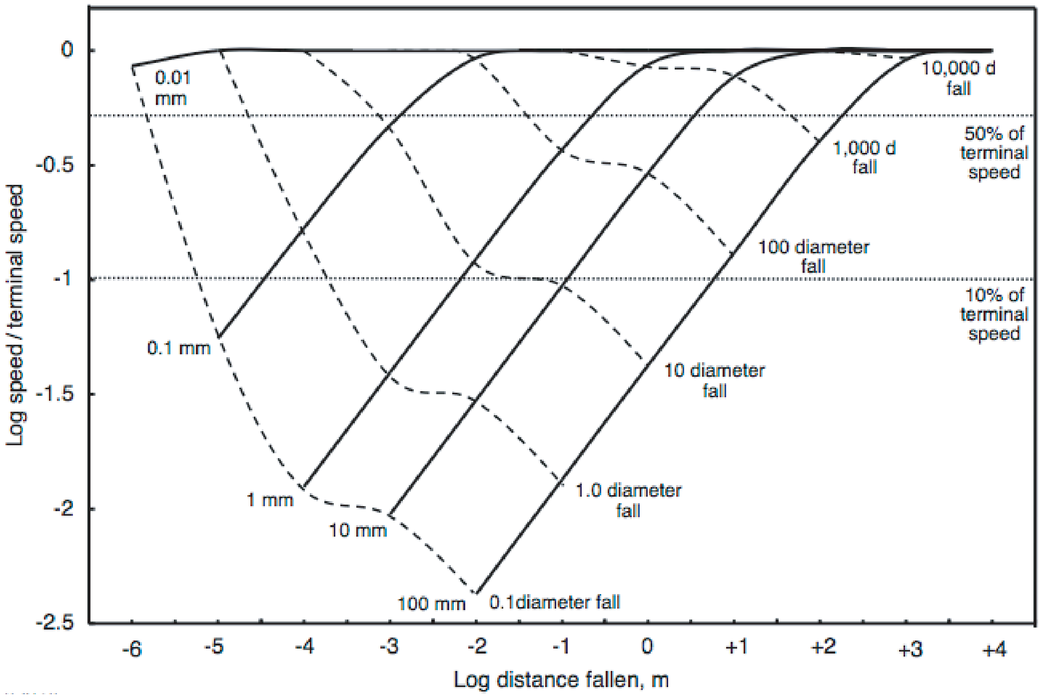

Figure 3. How the approach to terminal velocity varies with the size of spheres of density 1000 kg m^{-3} falling in air.

sphere takes 1 diameter, while the 10 μm sphere takes a mere 0.001 diameter. (The peculiarly irregular variation of drag with speed for falling spheres and most other ordinary shapes makes the sequence somewhat erratic.)

In short, when falling in air, drag and terminal speed carry biological significance mainly for small objects– unless, of course, one considers selectively questionable behaviour such as sky-diving by humans. Still, even small ones may run out of range of Stokes' law, reliable only for Reynolds numbers decently below 1.0, so one may need to do calculations such as those that generated figures 2 and 3. A sphere 1 mm in diameter will exceed $Re = 1.0$ after a fall of less than a millimeter, long before reaching terminal speed (at which $Re = 230$), indeed well before drag has begun to alter its motion much at all. Even a 0.1 mm sphere, about at our visual threshold, will exceed $Re = 1.0$ at terminal speed.

What happens below diameters of about 10 μm? Stokes' law gives unambiguous results with no uncertain coefficients. Unfortunately an exceedingly basic assumption, one only infrequently made explicit, begins to break down and restrict its reliability. For the most part, fluid mechanics treats fluids as continua–non-particulate, infinitely divisible without loss of character. While 10 μm remains far above molecular dimensions, molecular phenomena nonetheless start to intrude. The terminal speeds of particles begin to deviate from those anticipated by Stokes' law as their diameters approach the mean free paths of the molecules of their surrounding fluid. For air, mean free paths are of the order of 0.1 μm, only two orders of magnitude smaller than the 10 μm spheres considered here and closer still to, for instance, a 4-μm spore of the fungus *Lycoperdon* (Ingold 1971).

In effect, the Brownian motion due to random collisions with moving gas molecules rises to the same scale as that caused by gravity, so motions become irregular, eventually having only a statistically-downward bias. The relative magnitude of the effect increases rapidly with decreasing particle size both because gravitationally-driven descent speed decreases and because the effective Brownian displacement speed increases. (The later, as noted in the first of these essays is a peculiarly duration-dependent speed, here the square root of the quotient of twice the diffusion coefficient for a particle of a particular size divided by a reference time.) For a 10 μm particle (still of water's density sinking through air) and a reference time of 1 s, Brownian displacement speed is less than a thousandth of gravitational speed; for a 1 μm particle (perhaps an airborne bacterium), Brownian displacement speed rises to a fifth of gravitational speed; for a 0.1 μm particle, Brownian speed approaches a hundred times gravitational speed (Monteith and Unsworth 1990; Denny 1993).

Further clouds on the horizon need more mention than they usually get, suggesting caution in adopting textbook equations. The equations that generated figure 3 assume quasi-steady motion in an unbounded fluid; that is, they take no account of any special phenomena associated with acceleration. At least two unsteady phenomena can take on importance, one largely independent of the fluid's viscosity, the other its direct consequence.

First, when a body accelerates in one direction, fluid must accelerate in the other. The latter requires force no less than the former; it goes as the "acceleration reaction force" or simply the "acceleration reaction". For a sphere, one calculates the extra force by presuming that the body has an additional mass equal to half that of the volume of fluid it displaces–that half is the "added mass coefficient" of a sphere (Daniel 1984, Denny 1988). So accelerations are less than calculated for a quasi-steady case–as if drag were increased, but with the effect scaling with volume rather than diameter or cross section. Decelerations are also reduced, with the acceleration reaction now opposing drag. For a sphere of biologically relevant density in air, the acceleration reaction will usually be negligible next to drag. It should matter, though, for a buoyant balloon just after release. In water, the acceleration reaction can be a major factor. In at least one circumstance it dominates—for the initial ascent of a bubble of gas in a liquid. Here the mass, even half the mass, of the displaced fluid far exceeds that mass of the accelerating body, so neglecting the acceleration reaction gives an acceleration overestimate of several orders of magnitude (Birkhoff 1960).

Second, setting up a steady-state flow pattern around a body takes time, so a history term may be significant during acceleration. Again, accelerations are reduced, here because velocity gradients and thus shear forces are more severe than otherwise expected. Again, the effect, often called the "Bassett term," (Michaelides 1997; Koehl *et al* 2003) will only rarely be important for ordinary bodies accelerating in air.

[The Bassett term is analogous to the long-known Wagner effect (Wagner 1925; Dudley 2000), a delay in the development of aerodynamic lift as an airfoil begins to move. Moving those initial 7 or 8 wing widths cannot greatly tax the run-up to take-off of an airplane; but it demands special devices for animals that, lacking rotational propellers, must flap their wings, starting each wing twice during each stroke.]

Nor do the usual equations worry about wall effects or interparticle interactions, which occur whether a body is accelerating or moving steadily. They result mainly from viscosity and the resulting velocity gradients. A body falling near a wall falls more slowly, sometimes much more slowly; the lower the Reynolds number, the more severe the effect and the more distant can be a confounding wall. And one starting from a surface will have a lower initial acceleration. Conversely, a body falling in the wake of another will experience lower drag and tend to catch up; a cloud of tiny bodies can thus coalesce as the bodies fall. These effects are more likely to be more important in air than are the acceleration reaction and the Bassett term, at the same time they are easier to identify and avoid. Not that biologists do so consistently–a substantial literature for sinking rates in nature comes from measurements of the descent speeds of clouds of individuals in worrisomely narrow tubes.

5. Another way to descend slowly

The higher a body's drag, the slower gravity will make it descend, at least, when drag has acted for a sufficient time. Alternatively, descents can also be slowed with some kind of lift-generating airfoil–a device that produces a force component at right angles to the oncoming airflow as well as the inescapable drag, parallel to flow. Not only does this mode of descent-slowing have quite a different aerodynamic basis, but it imposes an antithetical requirement. The effectiveness of a lift-producing airfoil depends on its lift-to-drag ratio. That implies minimization rather than maximization of drag.

Higher Reynolds numbers allow airfoils to achieve greater lift-to-drag ratios, so using lift to slow descents becomes increasingly attractive as systems enlarge. Thus airborne seeds (or fruits or seed-leaves–more generally, 'diaspores' or 'propagules') that slow descents by increasing drag mostly have masses below 50 to 100 mg; very few of what Augspurger (1986) terms 'floaters' exceed that benchmark. Ones heavier than this, such as the samaras of maples, ashes, and tulip poplars, mostly employ lift-producing airfoils; conversely samaras come no lighter than about 10 mg (Azuma 1992). By contrast, animals such as flying squirrels that can control their aerial postures blur the boundary, with no hard and fast distinction between drag-based parachuting and lift-based gliding–more about these shortly.

What the lift-to-drag ratio sets is the angle with which a gliding airfoil descends, whether a glider moves in one direction or, as does a samara, takes a helical path as it autogyrates downward. Specifically, the lift-to-drag ratio, *L*/*D*, (for the entire craft if made of more than a single airfoil) equals the cotangent of the angle relative to the horizontal, ε, of a steady-speed descent:

$$\frac{L}{D} = \cot\varepsilon \text{ and } \frac{D}{L} = \tan\varepsilon. \quad (9)$$

But although it sets the path, the ratio does not fully determine the speed at which the craft approaches the earth. For a steady glide, descent speed depends as much on the amount of upward force needed, which must equal the weight of the craft. Since lift (like drag) is proportional (putting aside some secondary matters) to the square of speed, that square of speed varies directly with weight. Doubling the weight of a glider increases its steady-state speed (both overall and descent speeds) by 1.414. Thus in still air (and assuming a unidirectional glide, not an antogyrating vertical descent), the heavier glider will go about as far when released from a given height, but it will get there faster.

That independence of glide angle and weight may underlie the large size of some fossil fliers, whether insects (Paleodictyoptera) or reptiles (pterosaurs). One hastens to add at least one caution, though. One might imagine that, assuming biologically-ordinary tissue densities, an increase in the weight of a glider will be offset by the increase in wing area and thus lift, which varies direct with it. But the scaling of wing area and body mass, about which more in the next section, undercuts that offset. For isometric craft of constant density, lift will vary with wing area, *S*, and thus body area and length squared, while weight will vary with length cubed, so

$$W/S \propto l. \quad (10)$$

That variable, *W*/*S*, goes by the name "wing loading." As a consequence of its scaling with length, bigger must mean faster, at least if size increases isometrically. That demands some combination of shorter glides, more wind-dependent take-offs, and harder landings. Perhaps those large fossils tell us that back when few fast terrestrial predators lurked, isometry could be put aside–the increased fragility of light construction and disproportionately large wings may have been less disadvantageous.

(Wing loading may enjoy a weight of tradition, but it ignores at least one potentially confounding factor. Long, narrow wings do better than short, wide ones, something now well understood, but evident only empirically in the early days of flight. To avoid giving equal weight to length and width in wing area, an alternative variable, "span loading", the ratio of weight to the square of the wing span, sometime finds use. Choice of variables matters little for wings of ordinary proportions or for comparisons among wings of similar shapes.)

So both glide angle and wing loading (the relevant form of weight) enter the picture. The first looks at first glance as if size-independent, while the second is inimical to large craft. In addition, there is a third variable, one inimical to small craft. Lift, or properly the lift coefficient (C_l) of a high-quality airfoil, depends only slightly on Reynolds number, at least for *Re*'s above those of fruit-flies, around 100; it usually has a value or about 1.0 or a bit less at a maximal lift-to-drag ratio. By contrast, drag, expressed as the analogous drag coefficient (C_d), drops with increases in the Reynolds number. Formally defining those coefficients, we have

$$L = C_l \rho_m v^2 S/2 \text{ and} \quad (11)$$

$$D = C_d \rho_m v^2 S/2, \quad (12)$$

where ρ_m is the density of the medium and *S* is the projecting area, normal to flow, of the airfoil. In effect, the best achievable lift-to-drag ratio will increase (if complexly) with size, putting small gliders at a disadvantage when it comes to glide angle.

How much so? Looking at how a specific airfoil's or aircraft's performance varies with Reynolds number will most likely mislead us, since effective airfoil design itself varies with *Re*. So we might compare a heterogeneous collection, ones that have proven effective under their individual operating conditions. Figure 4 gathers data for lift-to-drag ratios for airfoils as a function of Reynolds number; these, bear in mind, are data from diverse sources with embedded estimates to render them commensurate. A competitive sailplane may have a ratio of 50; a top-flight bird, an albatross, of 20; the hindwing of a desert locust about 8; flies and bees around 2; test airfoils at $Re = 10$ and $Re = 1$ of 0.43 and 0.18 respectively. Clearly small fliers cannot achieve glide angles as low as those of large fliers, and the increasing glide angles at Reynolds numbers below about 500 make gliding itself impractical. Some large insects glide, at least occasionally; small ones do not.

In one sense, though, the inferior glide angles of insects (and, although less extreme, of birds) may mislead us. In that earlier assertion that as wing loading went up with body size so must flying speed lies a compensating advantage of small, if not very small, size. In the real world, gliding in a temporally and spatially uniform atmosphere represents both a worst and an uncommon case. We know quite a few ways gliders can take advantage of atmospheric structure, what we have taken to call 'soaring' as opposed to simple gliding. Most schemes for soaring depend as much or more on time aloft than on the horizontality of simple gliding. Time aloft, of course, varies inversely with sinking speed, so time aloft is no worse for a flier that descends twice as steeply if it flies half as fast. With still-air time aloft as the criterion, gliding/soaring retains utility down into the large insect range–it may even improve. A limit line drawn through the upper left set of points in figure 4 has a slope of 0.217. Converting from lift-to-drag ratio to sinking speed tells us the latter will vary with $Re^{0.116}$ for isometric gliders of equal lift coefficients (eq. 11). So the smaller glider will approach the earth somewhat less rapidly in still air or, of more relevance, slower ascending air will suffice to keep it aloft. Tucker and Parrott (1970) make just this point,

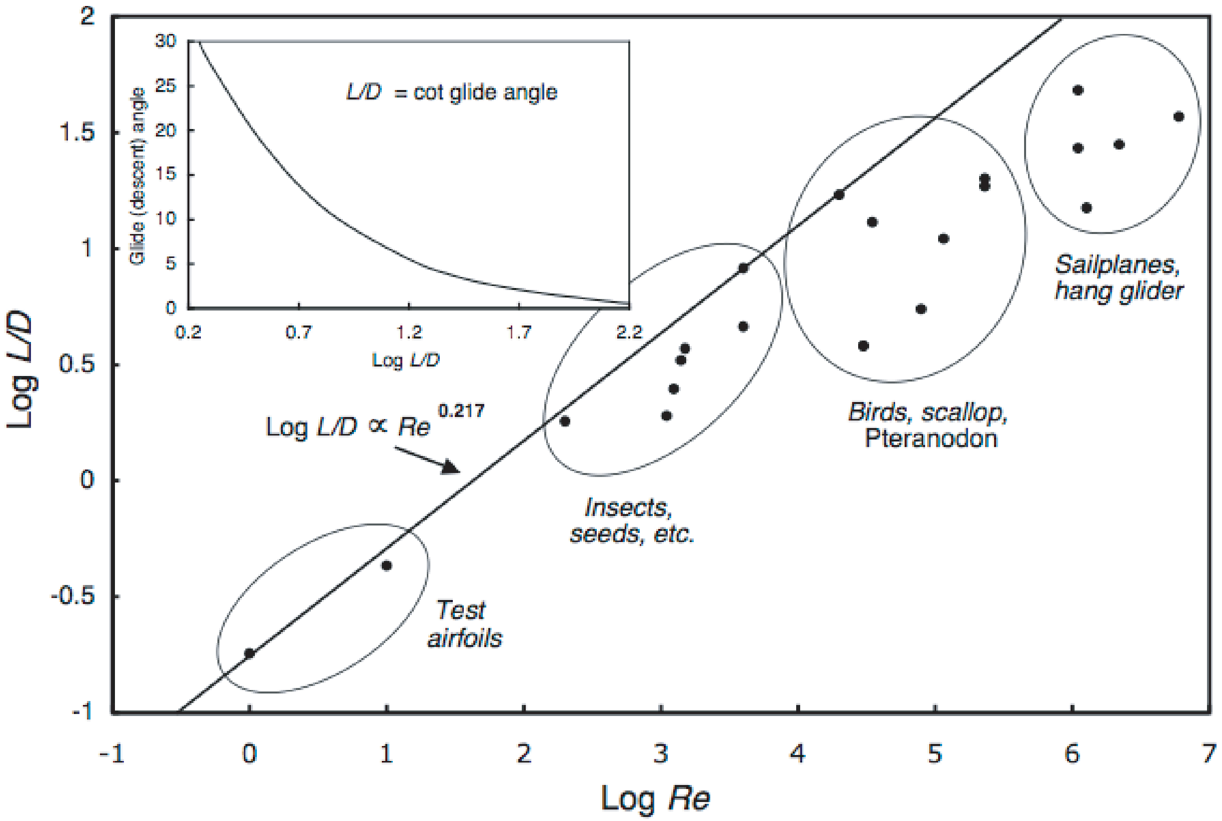

Figure 4. The scaling of the lift-to-drag ratio with Reynolds number, with an empirical limit line. The inset provides a conversion of the ratio to the steady-state glide angle with respect to the horizontal. Data for human-carrying craft from commercial websites; *Pteranodon* from Bramwell (1971); scallop from Hayami (1981); birds from Withers (1981); insects from sources in Vogel (1994, p 249); seeds from Azuma (1992); test airfoils from Thom and Swart (1940).

noting that soaring birds such as some vultures can achieve lower minimum sinking speeds than high-performance sailplanes.

So why do not tiny, even microscopic, gliders fill the skies? Flapping fliers blur the issue, with what appears to be a gradual diminution with decreasing size of the extent to which they employ intermittent gliding; locusts, butterflies and dragonflies glide at least a little; bees and flies do not. Purely passive gliders provide a clearer dichotomy. Among plants–those wind-dispersed, autogyrating samaras, mainly–as noted earlier, gliders get no smaller than about 10 mg and fly at Reynolds numbers no lower than about 500. Among animals, purely passive gliders drop out below about 1 g (McGuire and Dudley 2005). That initially puzzling 2-order-of-magnitude difference may just be a matter of adaptational opportunities or lack of alternatives such as active flight. Still, the absence of a fauna analogous to the samaras seems odd. Arboreal insects have the option of flight, but where are the gliding arboreal spiders?

The data collected in figure 5 may provide a bit of insight. Perhaps that slight improvement with decreasing size eventually competes with an alternative that offers still better scaling. The left, linear portion of figure 2 implies that sinking speed for objects retarded by drag will vary with $Re^{0.67}$, more drastic scaling than the $Re^{0.116}$ for lift-based retardation, an implication well confirmed by the real-world data of figure 5. Smaller becomes not a bit but a lot better than for craft that slow their descents by maximizing drag. Blurring the contrast just a little, the drag-based floaters, while generally lighter than the lifters, operate in a realm with an intermediate scaling exponent–with sinking speed proportional to $Re^{0.33}$ as on the right side of figure 2. And without obvious exception, they keep sinking rates reasonable through drastic surface proliferation, equipping themselves with all manner of hairs, fluff and appendages.

Too quick a look at such data for both glide angle and sinking speed may mislead us in another way. It accords poorly with the repeated evolution of remarkably bad gliders in several groups of terrestrial vertebrates–frogs that glide using oversize webbed feet, lizards that glide with lateral trunk extensions, squirrels and phalangers that glide with thin skin that stretches from fore to hind legs, even snakes that glide with a bit of body flattening and a

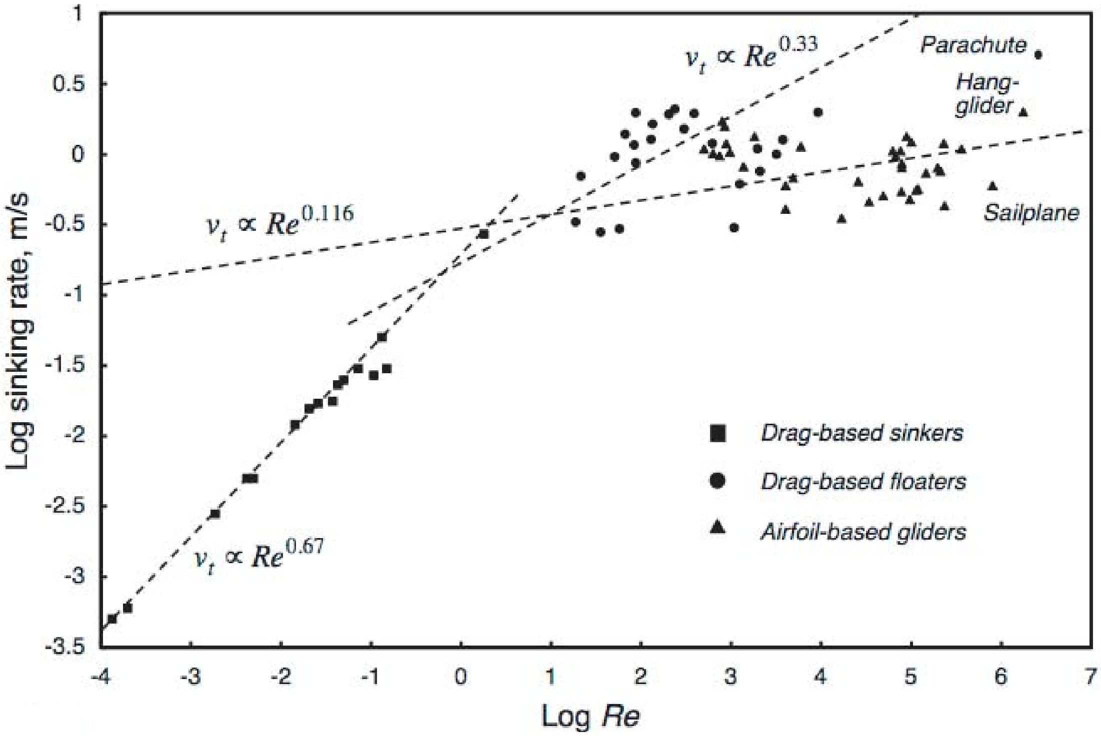

Figure 5. Sinking rates of passively sinking or gliding systems spanning an especially wide size range and anything but isometric. Note that while the exponents of the scaling lines can be justified, their positions are arbitrary. The exponent of 0.116 comes from the empirical limit line of figure 4; the exponents of 0.33 and 0.67 are theoretical and come from figure 2. Sources: Azuma (1992), Bramwell (1971), Gibo and Pallett (1979), Ingold (1971), Jensen (1956), McGahan (1973), Niklas (1984), Okubo and Levins (1989), Parrott (1970), Pennycuick (1960, 1971, 1982), Rabinowitz and Rapp (1981), Tennekes (1996), Trail *et al* (2005), Tucker and Heine (1990), Tucker and Parrott (1970), Verkaar *et al* (1983), Ward-Smith (1984), Werner and Platt (1976), and Yarwood and Hazen (1942).

cross-flow body orientation (see, for instance, Dudley and DeVries 1990; Norberg 1990). None of these achieves an especially high lift-to-drag ratio for its size; values run from a little over 2 to a little under 5 (Socha 2002). The apparent paradox may stem from the way both glide angle and descent speed tacitly assume steady-state activities. These may be the exception rather than the norm in these animals–why I omitted them from figure 5. A large part of a trajectory typically consists of an initial outward and then downward leap, with only a minor aerodynamic component; the path then becomes ever less vertical. Major and deliberate drag increases may precede landings, raising lift at the expense of speed and the lift-to-drag ratio, in a sense reinvesting the momentum of the initial leap when airspeed is no longer an asset.

Recent work on flying snakes, genus *Chrysopelea*, (Socha *et al* 2005; see also www.flyingsnake.org) and on lizards, genus *Draco* (McGuire and Dudley 2005) provide object lessons. Clearly the old and often quoted distinction between parachuting and gliding, whether the trajectory descends more or less steeply than 45º, is worse than arbitrary; its implied scenario diverges misleadingly from reality.

One point of figure 5, the superiority making lift rather than drag for staying aloft–at least for Reynolds numbers high enough for decent lift-to-drag ratios–can be argued in another way. Consider a hypothetical drag-based descender that loses altitude at the same rate (0.41 m/s) as the winged seed-leaf of the Javanese cucumber *Alsomitra* (*Zanonia*), which operates at $Re = 4000$ (Azuma and Okuno 1987; Alexander 2002). If the descender weighed no more than that seed-leaf (210 mg) and took the form of a flat horizontal disk (thus normal to the upward relative flow: $C_d = 1.2$), it would need an area of 3.4 times that of *Alsomitra's* 0.005 m^2. And the latter operates at the unimpressive lift-to-drag ratio of 3.7, apparently accepting a lesser value than its best 4.6 to gain the intrinsic stability critical for a totally passive glider. For a similar reason, windmills with blades rotating in a plane normal to flow became common and displaced ones turning horizontally, like cup anemometers, about a thousand years ago. Ships with propellers displaced most drag-based side-wheel and stern-wheel boats, starting a century and a half ago. Both transitions preceded the aircraft-stimulated development of propellers that could achieve respectable L/D ratios.

The dichotomy between drag-based and lift-based descent-slowing carries a further message. That size-dependent shift from drag as good to drag as evil may constitute an odd adaptive barrier–a device well-attuned to one mode will ordinarily be especially bad in the other. Active flight has evolved from gliding flight whenever it has appeared, but gliding flight seems never to have evolved from drag-based descent retardation. One serious suggestion that flying insects took that route (Wigglesworth 1963) has never gained substantial support. The nearest thing to an evolutionary switch I can think of occurs in a few Lepidoptera such as the gypsy moth (*Lymantria dispar*), a notorious pest in North America. Instead of basing dispersal on actively flying adults, the first instar caterpillar does the job by paying out long silk strands as if a newly-hatched spider.

Not that one cannot imagine plausible designs that might permit fairly easy shifts from, say, drag maximization to gliding in passive craft. A round horizontal disk with a mass on a rigid stalk beneath its center will descend with lots of drag. Moving the stalk and mass closer to an edge could convert the device to something like a hang-glider, with better still-air dispersal distance as a selective reward. I would not place a bet, even at good odds, against the reality of such a scenario–some seed-leaves look like good candidates.

6. Flying–why big craft should fly swiftly

In simple gliding, gravity provides the motive force, and energy to sustain the process comes from the steady loss of altitude; in soaring, the energy ultimately comes from atmospheric structures. In the sustained, active flight of airplanes, birds, pterosaurs, insects, and bats, the lift of paired wings again plays the key role. But sustaining altitude without that gravitational or atmospheric free ride demands some engine, typically either a propeller directing air rearward with a fixed wing deflecting the craft's airstream downward or else flapping wings that create both rearward and downward airstream momentum. Averaged over all but the briefest of time spans, the upward aerodynamic resultant must precisely equal the downward gravitational force, the weight of the craft, just as for steady gliding. So the same basic scaling rule appears applicable. As in eq. (10), weight divided by wing area, or wing loading, ought to vary directly with body length for an isomorphic set of fliers or, assuming constant density as well, with body mass to the 1/3 power. And similarly, bigger should mean faster; from eq. (11) we see that

$$v \propto (W/S)^{1/2} \propto l^{1/2} \propto m_b^{1/6}, \tag{13}$$

a specific prescription for how much faster larger aircraft must fly.

In a lovely book, Tennekes (1996) makes this a major point, drawing a single line on a graph that appears to indicate compliance (without even a shift in the constant of proportionality, 0.38 in SI units) from fruit flies to the largest passenger aircraft, a Boeing 747. Wing loading, W/S, goes up as the cube root of mass, m, and eq. (13) predicts cruising speed quite well. Other sources such as McMahon and Bonner (1983), Azuma (1992) and Dudley (2000) cite the same rule. Airplanes fit almost perfectly, at least if one excludes gliders and human-powered craft, which keep wing loading and therefore cruising speed deliberately low. Birds fit the same regression line, wing loading again going up with mass$^{1/3}$, with both the same proportionality constant and scaling exponent.

Insects, though, scatter a lot more, with the scaling line recognizable only by lumping some very lightly wing-loaded butterflies and moths with heavily loaded beetles, bees, and flies and following downward the pre-established trend. Except for dragonflies, the insects we regard as smooth, fast fliers weigh several times more relative to their wing areas than the scaling relationship predicts–as do hummingbirds. Furthermore, eq. (13) predicts flying speeds considerably in excess (roughly double) what the all-too-few reliable measurements (and other considerations) show.

Why this fly-in-the-ointment? I think nothing especially obscure underlies the deviation. In a sense, the problem combines etymology and entomology (my apologies to the hummingbirds). The smaller the flapping flyer, the more the function of what we call a wing approaches that of a propeller and the less it resembles that of a paradigmatic airplane wing. In effect, a flapper uses its wings more often than does a fixed-wing craft. Indeed we see the greatest deviations from the rule where wingbeat frequencies are highest, more specifically, where the speed of the wings in their upstroke-downstroke oscillation most exceeds that of the insect's or hummingbird's forward flight. In effect, a flapper uses its wings more often than does a fixed-wing craft, and the speed most relevant becomes the tip speed of each beating wing rather than the forward speed of the craft. (alternatively, the area most relevant becomes the area swept by the wings in a stroke rather than the area of the wings themselves–"disk loading" thus replaces wing loading.)

So we need another parameter, the ratio of the forward movement of the craft to up-and-down wing movement. The propeller designers provide one, the so-called advance ratio, *J*, although for applicability to animal flight it has to be altered slightly–a wing swings, down plus up, through less than an angle of 360º, and its additional parameter, amplitude, can itself vary. As usually given for flying animals (Ellington 1984),

$$J = \frac{v_f}{2\phi n R}, \tag{14}$$

where v_f is flight speed, ϕ amplitude (or "stroke angle"), n wingbeat frequency, and R wing length.

Amplitude varies too little to matter here. Wing length, of course, goes down with size, which would push up the advance ratio. But small insects suffer more from the pernicious effects of viscosity and must make do with lower *L*/*D* ratios, as we saw earlier. So they have to beat their wings at high frequencies and fly slowly–their wings go up and down a lot for only a little forward progress–which more than offsets their small size. *J* for a bumblebee peaks at about 0.66, for a black fly 0.50, for a fruit fly 0.33. By comparison, ducks and pigeons fly at about 1.0 (Vogel 1994). Halving the advance ratio roughly doubles the effective wing area, about what we see when comparing birds, which follow the scaling rule, with these fairly fast insects, which have greater wing loading and fly faster than it predicts. For the particulars of how small insects achieve frequencies that may reach 1000 s^{-1}–fabulously lightweight wings, special neuromuscular devices, and so forth–one should look at Dudley (2000) and other specialized accounts.

From the viewpoint of scaling, the relatively high flight speeds of some tiny insects–around 1 m s^{-1} for a fruit fly (500 body lengths per second; a duck does less than 100)–might be surprising, whatever their obvious utility in an atmosphere that is rarely still. After all, their higher surface-to-volume ratios mean a relatively greater cost to deal with drag and a lesser cost to offset gravity. In fact, while true, the force needed to oppose drag remains modest–for a falcon less than 1% of weight or lift (Tucker 2000); for a teal (a duck) about 2% (Pennycuick *et al* 1996); for a desert locust, about 4% (Weis-Fogh 1956); for a bumblebee, 8% (Dudley and Ellington 1990); for a fruit fly about 10% (Vogel 1966). Those percentages, incidentally, suggest that drag reduction through streamlining can only marginally affect the cost and practicality of flight. Gravity remains the chief opponent.

7. The value of gravitational acceleration

Both whole organisms and their parts inevitably exceed the density of air–nature makes no blimps or ascending balloons. So any biological system that keeps a bit of atmosphere between itself and the earth's surface must contend with gravity. The particulars prove complex physically, complex biologically, complexly size-dependent, and at least occasionally counterintuitive.

One final example may persuade the reader of that last. We expect that a greater gravitational acceleration, as would characterize a larger planet, would make passively aerial organisms descend faster. That same increase in *g*, though, ought to increase atmospheric density–in one scenario (which we will assume) increasing directly with *g* (see, for instance, Taylor 2005). In the Stokes' law world of the small and slow, terminal descent speeds will indeed increase, because they depend on the difference between the densities of object and medium, and even doubling the latter will still leave it insignificant. The change in atmospheric density will not affect the all-important viscosity.

At Reynolds numbers above one, drag becomes increasingly dependent on atmospheric density (eq. 12) and decreasingly dependent on viscosity. Ignoring some complications, drag will vary directly with density. If both drag and weight vary directly with gravitational acceleration, then drag-based terminal descent speeds will not – which strikes one as odd. By contrast, if lift also varies with density (eq. 11), then lift-to-drag ratios and glide angles (eq. 9) will be independent of air density. So the increased weight of a glider will make it descend faster – as in the drag-based Stokes' law range but not the drag-based higher *Re* range. Gravity always drives the aerial system earthward, but that does not imply inevitable importance for the particular value of *g*.

Acknowledgements

Much of this essay was written while I was Visiting Scholar at the Darling Marine Center of the University of Maine; for arranging that visit I am grateful to Sara Lindsay and to Kevin Eckelbarger, its director. For particular insights I am indebted to Pete Jumars, Larry Mayer, Laura Miller, and Jim Price.

References

Alexander D E 2002 *Nature's flyers: birds, insects, and the biomechanics of flight* (Baltimore: Johns Hopkins University Press)

Augspurger C K 1986 Morphology and dispersal potential of wind-dispersed diaspores of neotropical trees; *Am. J. Bot.* **73** 353–363

Azuma A 1992 *The biokinetics of flying and swimming* (Tokyo: Springer-Verlag)

Azuma A and Okuno Y 1987 Flight of a samara, *Alsomitra macrocarpa*; *J. Theor. Biol*. **129** 263–274

Birkhoff G 1960 *Hydrodynamics: a study in logic, fact, and similitude* 2nd edition (Princeton: Princeton University Press)

Bramwell C D 1971 Aerodynamics of *Pteranodon*; *Biol. J. Linn. Soc*. **3** 313–328

Daniel T L 1984 Unsteady aspects of aquatic locomotion; *Am. Zool.* **24** 121–134

Denny M W 1988 *Biology and the mechanics of the wave-swept environment* (Princeton: Princeton University Press)

Denny M W 1993 *Air and water*: *the biology and physics of life's media* (Princeton: Princeton University press)

Dudley R 2000 *The biomechanics of insect flight: form, function, evolution* (Princeton: Princeton University Press)

Dudley R and DeVries P 1990 Tropical rain forest structure and the geographical distribution of gliding vertebrates; *Biotropica* **22** 432–434

Dudley R and Ellington C P 1990 Mechanics of forward flight in bumblebees. II. Quasi-steady lift and power calculations; *J. Exp. Biol.* **148** 53–88

Ellington C P 1984 The aerodynamics of hovering flight; *Philos. Trans. R. Soc. London* **B305** 1–181

Geiger R 1965 *The climate near the ground,* Revised edition (Cambridge: Harvard University Press)

Gibo D L and Pallett M J 1979 Soaring flight of monarch butterflies; *Can. J. Zool*. **57** 1393–1401

Glasheen J W and McMahon T A 1996 size-dependence of water-running ability in basilisk lizards (*Basiliscus basiliscus*); *J. Exp. Biol*. **199** 2611–2618

Hayami I 1991 Living and fossil scallop shells as airfoils: an experimental study; *Paleobiology* **17** 271–276

Hall F G, Dill D B and Barron E S G 1936 Comparative physiology in high altitudes; *J. Cell. Comp. Physiol.* **8** 301–313

Ingold C T 1971 *Fungal spores: their liberation and dispersal* (Oxford: Clarendon Press)

Jensen M 1956 Biology and physics of locust flight. III. The aerodynamics of locust flight; *Philos. Trans. R. Soc. London* **B239** 511–552

Kling G W, Clark M A, Compton H R, Devine J D, Evans W C, Humphrey A M, Koenigsberg E J, Lockwood J P, Tuttle M L and Wagner G L 1987 The 1986 Lake Nyos gas disaster in Cameroon, West Africa; *Science* **236** 169–175

Koehl M A R, Jumars P A and Karp-Boss L 2003 Algal biophysics; in *Out of the past* (ed.) A D Norton (Belfast: British Phycological Association) pp 115–130

McGahan J 1973 Gliding flight of the Andean condor in nature; *J. Exp. Biol. 58* 225–237

McGuire J A and Dudley R 2005 Comparative gliding performance of flying lizards; *Am. Nat*. **166** 93–106

McMahon T A and Bonner J T 1983 *On size and life* (New York: Scientific American Library)

Michaelides E E 1997 Review–the transient equation of motion for particles, bubbles, and droplets; *J. Fluids Engin*. **119** 233–247

Monteith J L and Unsworth M 1990 *Principles of environmental physics,* 2nd edition (Oxford: Butterworth-Heinemann)

Niklas K J 1984 The motion of windborne pollen grains around conifer ovulate cones: implications on wind pollination; *Am. J. Bot.* **71** 356–374

Norberg U M 1990 *Vertebrate flight: mechanics, physiology, morphology, ecology and evolution*. (Berlin: Springer-Verlag)

Okubo A and Levins S A 1989 A theoretical framework for data analysis of wind dispersal of seeds and pollen; *Ecology* **70** 329–338

Parrott G C 1970 Aerodynamics of gliding flight of a black vulture *Coragyps atratus*; *J. Exp. Biol*. **53** 363–374

Pennycuick C J 1960 Gliding flight of the fulmar petrel; *J. Exp. Biol*. **37** 330–338

Pennycuick C J 1971 Gliding flight of the white-backed vulture, *Gyps africanus*; *J. Exp. Biol*. **55** 13–38

Pennycuick C J 1982 The flight of petrels and albatrosses (Procellariiformes), observed in South Georgia and its vicinity; *Philos. Trans. R. Soc. London* **B300** 75–106

Pennycuick C J, Klaassen M, Kvist A and Lindström A 1996 Wingbeat frequency and the body drag anomaly: wind-tunnel observations on a thrush nightingale (*Luscinia luscinia*) and a teal (*Anas crecca*); *J. Exp. Biol*. **199** 2757–2765

Persson A 1998 How do we understand the Coriolis force?; *Bull. Am. Meteorol. Soc.* **79** 1373–1385

Rabinowitz D and Rapp J K 1981 Dispersal abilities of seven sparse and common grasses from a Missouri prairie; *Am. J. Bot.* **68** 616–624

Socha J J 2002 Gliding flight in the paradise tree snake; *Nature (London)* **418** 603–604

Socha J J, O'Dempsey T and LaBarbera M 2005. A three-dimensional kinematic analysis of gliding in a flying snake, *Chrysopelea paradisi*; *J. Exp. Biol.* **208** 1817–1833

Taylor F W 2005 *Elementary climate physics* (Oxford: Oxford University Press)

Tennekes H 1996 *The simple science of flight: from insects to jumbo jets* (Cambridge, MA: MIT Press)

Thom A and Swart P 1940 The forces on an aerofoil at very low speeds; *J. R. Aero. Soc.* **44** 761–770

Trail F, Gaffoor I and Vogel S 2005 Ejection mechanics and trajectory of the ascospores of *Gibberella zeae*; *Fungal Genet. Biol.* **42** 528–533

Tucker V A 2000 Gliding flight: drag a torque of a hawk and a falcon with straight and turned heads, and a lower value for the parasite drag coefficient; *J. Exp. Biol. 203* 3733–3744

Tucker V A and Heine C 1990 Aerodynamics of gliding flight in a Harris' hawk, *Parabuteo unicinctus*; *J. Exp. Biol.* **149** 469–489

Tucker V A and Parrott, G C 1970 Aerodynamics of gliding flight in a falcon and other birds; *J. Exp. Biol.* **52** 345–367

Verkaar H J, Schenkeveld A J and van de Klashorst M P 1983 The ecology of short-lived forbs in chalk grasslands: dispersal of seeds; *New Phytol.* **95** 335–344

Vogel S 1966 Flight in *Drosophila*. I. Flight performance of tethered flies; *J. Exp. Biol.* **44** 567–578

Vogel S 1994 *Life in moving fluids* 2nd edition (Princeton: Princeton University Press)

Vogel S 2005a Living in a physical world II. The bio-ballistics of small projectiles; *J. Biosci.* **30** 167–175

Vogel S 2005b Living in a physical world IV. Moving heat around; *J. Biosci.* **30** 449–460

Wagner H 1925 Über die Entstehung des dynamischen Auftriebes von Tragflügeln; *Z. Angew. Math. Mech.* **5** 17–35

Ward-Smith A J 1984 *Biophysical aerodynamics and the natural environment* (Chichester: Wiley-Interscience)

Weast R C (ed.) 1987 *CRC handbook of chemistry and physics,* 68th edition (Boca Raton: CRC Press)

Weis-Fogh T 1956 Biology and physics of locust flight. II. Flight performance of the desert locust; *Philos. Trans. R. Soc. London* **B239** 459–510

Werner P A and Platt W J 1976 Ecological relationships of co-occurring goldenrods (*Solidago*: Compositae); *Am. Nat.* **110** 959–971

Wigglesworth V B 1963 Origin of wings in insects; *Nature (London)* **197** 97–98

Withers P C 1981 An aerodynamic analysis of bird wings as fixed aerofoils; *J. Exp. Biol.* **90** 143–162

Yarwood C E and Hazen W E 1942 Vertical orientation of powdery mildew conidia during fall; *Science* **96** 316–317

*e*Publication: 9 February 2006

Living in a physical world
VII.
Gravity and life on the ground

Steven Vogel
Department of Biology, Duke University, Durham, NC 27708-0338, USA
(Fax, 919-660-7293; Email, svogel@duke.edu)

1. Introduction

Unless some energy-demanding process counteracts its effect, gravity inevitably makes aerial life descend. For terrestrial life, gravity acts less obviously, less immediately, and less consistently. Sometimes it matters; sometimes other agencies eclipse its effects. Sometimes it acts as impediment or nuisance; sometimes it plays a crucial positive role. In short, gravity has more diverse consequences and has elicited a wider range of biological devices for organisms that live on the ground.

For one thing, much more depends on the distinction between gravity, thus weight, and inertia, thus mass. Steadily lift an object, and you work against gravity; pull downward, and you enlist gravity's assistance. Sliding an object steadily sideways may entail no irreduceable resistance, but the frictional force you do feel still comes from gravity, from the press of the object against the substratum. But accelerate an object, and you work against its mass. Big Neanderthal thrusting spears put gravity to use, working best with heavy bodies that leaned forward over well-planted feet to get sufficient purchase on the ground. Lighter, thrown spears depended more on inertial mass – a running body, in effect no purchase at all, could aid a launch. Similarly, a lighter person has to lean further outward when opening a substantial door. The lesser weight needs to be more effectively applied to produce the sideways force that will accelerate the mass of the door. Muscularity is a secondary matter.

For another, organisms consist of both solids and liquids. In practice the two phases of matter face gravity in slightly different guises that reflect the difference between compressive stress and hydrostatic pressure. Both variables have dimensions of force per unit area, but in a specific direction for stress while omnidirectional for pressure. Stack solid bricks ever higher (with pads between to ensure uniform force transfer), and eventually the lowest will crush. That crushing point is reached when the compressive strength of brick, or force-resistance relative to cross section, of about 20 MPa (or MN m^{-2}), equals the weight of the column relative to cross section. If made of bricks whose density is 2000 kg m^{-3}, the column will be about 1000 m high. Taper changes the picture – a column tapering upward can extend farther; one expanding upward will not reach as far. With similar reasoning, Weisskopf (1975) estimated the maximum height of a mountain as 10 km, about 10% higher than our present highest; in his analysis, plastic flow rather than crushing set the limit, so taper mattered little.

Extend a pipe of liquid water upward in the air, and the pipe eventually bursts at ground level. The column of water extending upward stresses (in the sense above) the material of the pipe, but it does so in proportion only to the height of the column – cross section and contained volume have no direct relevance. The pressure difference, Δp, across the walls of the pipe will be the product of the liquid's density, ρ, gravitational acceleration, g, and the column's height, h, in the familiar equation for both manometry and conversions of pressure units:

$$\Delta p = \rho g h. \tag{1}$$

Transforming that pressure to tensile stress (σ_t) in the wall of the pipe depends, obviously, on the thickness of the wall of the pipe (Δr, assumed well below the radius r) and, less obviously, on its size, here the radius:

$$\sigma_t = \frac{\Delta p\, r}{\Delta r} = \frac{\rho g h r}{\Delta r}. \tag{2}$$

This last equation is prescient with biological implications. For a given pressure and a wall material of a given tensile strength, a narrower pipe (lower r) will manage with a thinner wall (Δr). For example, your capillaries withstand pressures about 1/3 of that in your aorta despite having walls 2000X thinner. They manage that apparently paradoxical feat (convenient for material exchange) because their diameters are about 4000X less than that of the aorta. As one can see from eq. (2), they feel about 6X less tensile stress in their walls rather than the many times more that one might guess (Zweifach 1974; Caro *et al* 1978). Or, anticipating just a bit, since neither cardiac blood pressure (Δp) nor maximum muscle stress (σ_t) changes with body size, the thickness of the ventricular wall (Δr) will remain a constant fraction of heart radius itself (r) (Seymour and Blaylock 2000).

Here I will examine three situations in which gravity plays a role, asking what sets blood pressures for animals of different sizes and with what consequences; what determines the gait transition speeds for legged animals; and what sets the heights of trees and forests.

2. Circulation and hydrostatics

The scaling of the circulatory components of vertebrates, especially mammals, has come in for renewed attention in recent years. Heart mass and total blood volume increase in direct proportionality to body mass ($\propto m_b^{+1}$). Capillary length goes up slightly with mass ($\propto m_b^{+1/5}$), while capillary density ($\propto m_b^{-1/6}$) and maximum heart rate ($\propto m_b^{-1/5}$) go down. Maximum oxygen consumption and cardiac output go up but not as fast as body mass itself (both $\propto m_b^{+7/8}$). But not all variables vary with mass; in particular, blood viscosity, capillary and red blood cell diameters, aortic flow speed, and average arterial blood pressure remain nearly the same. (Exponents from Baudinette 1978, Calder 1984 and Dawson 2005.)

In looking for gravity's consequences, we ought to take a closer look at that size-independence of blood pressure. That constancy, first noted over half a century ago, has become ever better supported. For mammals, the average of systolic peaks and diastolic minima (often taken as a third of systolic plus two-thirds of diastolic to get closer to a true time-averaged mean), is about 12,900 Pa (97 mm Hg). So we humans are typical, with our systolic pressure of about 16,000 Pa (120 mm Hg) and diastolic pressure of 10,500 Pa (80 mm Hg). For birds average pressure runs somewhat higher, 17,700 Pa (133 mm Hg) (Grubb 1983).

From our present viewpoint, constancy of blood pressure seems paradoxical. Terrestrial animals amount to ambulatory manometers, obeying eq. (1), with a blood density of about 1,050 kg m^{-3} for ρ, and thus with a pressure gradient of 10,300 Pa m^{-1} from head to toe. Without auxiliary pumps, blood pressure at head height has to drop as body height increases. Thus a normal human has a diastolic blood pressure of about 5,300 Pa (40 mm Hg) in the head and 20,000 Pa (150 mm Hg) in the feet (Schmidt-Nielsen 1997). While gravity cannot be turned off, the relatively high pressure gradient needed to keep blood flowing through the resistive vessels ordinarily exceeds that gravitational gradient. Health care people learn to cuff the arm at heart height when taking blood pressures, although (by my informal survey) almost none of them know just why or what error an improper height introduces. With that 5,300 Pa (a little lower if hypotensive) we manage to keep blood flowing steadily and our brains decently supplied with oxygen – I have seen no claim that mental agility decreases with body height. Roughly 4,000 to 6,000 Pa (diastolic) appears sufficient to keep a mammalian brain in business.

An animal with its head a meter above its heart should be in serious trouble at standard mammalian cardiac output pressure – during diastole, blood will cease flowing at all. Half a meter should be about the limit, with gravity dropping diastolic pressure by 5,100 Pa (almost 40 mm Hg).

In fact, animals that hold their heads high do not have normal mammalian blood pressure. Most sources of scaling exponents include some parenthetical remark such as "excluding the giraffe" (Calder 1984) after noting the standard and its nearly size-independent scaling. That exclusion represents not some special case but a necessary threshold for gravitational compensation. We might view the situation with the aid of a dimensionless ratio, a "gravitational hazard index" (*GHI*). Such an index puts the height of animal in pressure units, that is, as if it were a blood-filled manometer obeying eq. (1); it divides this "manometric height" (ρgh) by average arterial (heart-high) blood pressure:

$$GHI = \frac{\rho g h}{\Delta \bar{p}}. \tag{3}$$

Figure 1 considers average blood pressures relative to this *GHI*. (One should recall just how labile a variable is one's own pressure and recognize the limitations of data from animals of even less certain disposition.) Two limits on blood pressure can be discerned. A lower horizontal line must represent the minimum average arterial pressure needed to overcome the resistance of the systemic system of conduits; it has a value of about 10,000 Pa (75 mm Hg). A vertical line to the right of the data points represents the limit set by the need to supply a brain at some minimum pressure after gravity exacts its tax on cardiac output. It appears to have a value (dimensionless) of about *GHI* = 1.7. Bear in mind the use of overall height instead of heart-to-head height and zero rather than some necessary minimum cranial pressure.

Small mammals can ignore gravity, while large (at least tall) ones most definitely must care. In effect, mammals tolerate the gradual diminution of cranial blood pressure with increasing height – up to a point. That point corresponds to animals only slightly taller than ourselves. (And thus slightly hypotensive or unusually tall humans manage quite well.) While we have too few reliable data for mammals taller than ourselves, we have no reason to suppose that their blood pressure does anything other than tracking the sum of two components, that set by the resistance of the system and that set by the need to raise blood against gravity.

The perceptive reader may think of a simple evasion of this problem of getting blood up to the head – in a word,

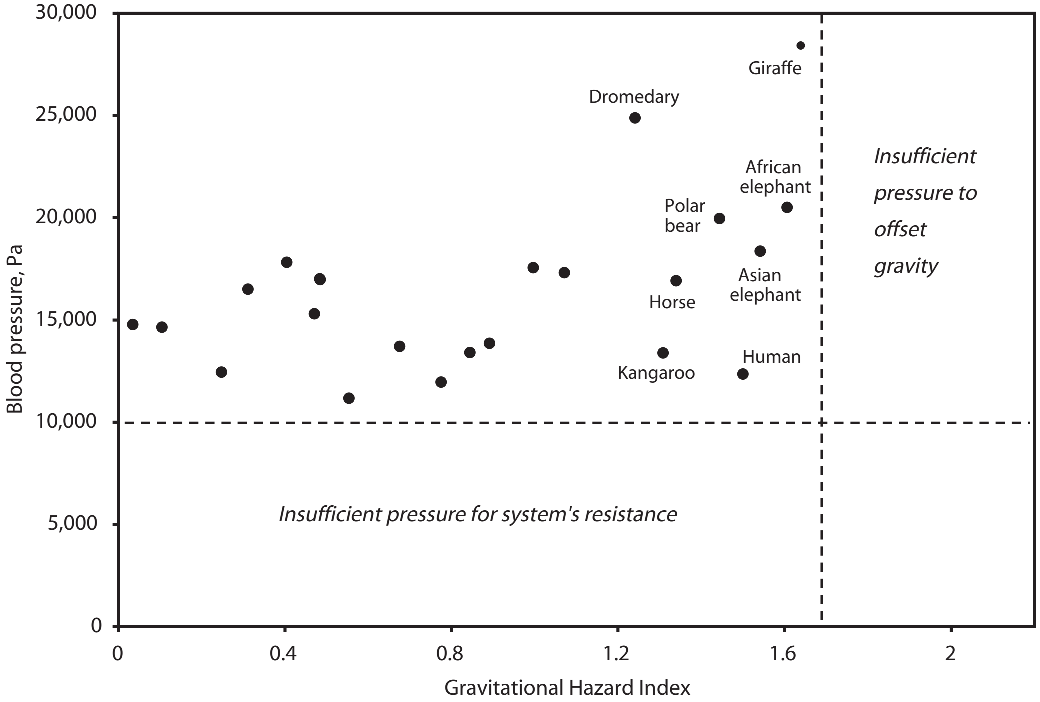

Figure 1. Blood pressures of mammals, with names of the larger ones, plotted against gravitational hazard index (*GHI*). Most of the pressure data come from Seymour and Blaylock (2000), with a few additions and confirmations from *www.ivis.org;* heights have been estimated from the photographs and shoulder heights in Nowak (1991).

siphoning. Since vessels are full, descending blood could draw blood upward, reinvesting the energy of decent to raise blood. The issue of siphoning has provoked no small amount of controversy; at present the weight of evidence opposes it (Pedley *et al* 1996). Tall mammals do generate the high pressures needed to raise blood without siphoning – one of the incentives for work on giraffes. Blood vessels in the head appear reinforced against conventional outward aneurysms rather than inward collapse. And the descending veins are all too collapsible, so blood commonly descends in boluses rather than a continuous stream. Of course one should not rule out the possibility that at least some siphoning still occurs under some circumstances, perhaps during vigorous aerobic activity.

At the same time, blood vessels in the legs must be strong enough to take both the higher cardiac pressure and the extra gravitational component. Which, not surprisingly, they are. In addition, the entire legs must be wrapped with an especially inextensible integument lest the extracellular space become oedematous. Which they are as well. In giraffes in particular, the vessels of head and neck need similar reinforcement, almost certainly important in preventing aneurysms when an animal lowers its head to drink.

Between their higher average blood pressures and lack of very tall extant members, birds should never hit an equivalent limit. One does wonder about giant moas, extinct for the past 800 years – the wall thickness of some miraculously preserved artery would probably allow reasonable estimation of their blood pressure. By contrast, reptiles (or “other reptiles” to some) present a much more interesting issue. Blood pressures run about a third of those of mammals, so the vertical limit line of figure 1 should occur at a third of the equivalent mammalian body height – about 0.57 rather than 1.7. Most extant reptiles are either small or lie low to the ground and should have no problem with gravitational pressure loss even so. Not all, though; in particular, some fairly long snakes climb trees and go over obstacles, making “fairly long” into “fairly tall.” In fact, the average heart-level blood pressures of long snakes vary widely, from about 3,300 Pa (25 mm Hg) in aquatic species to around 10,500 Pa (80 mm Hg) in terrestrial climbers. More remarkably, terrestrial climbers position their hearts substantially closer to their anterior ends— in a comparison of a python and a file snake of about equal length, about 25% of snout-vent distance versus 37%. In addition to these differences, climbers have reinforced body walls in their posterior regions and especially well-developed baroregulatory reflexes (Seymour and Arndt 2004).

No basic inferiority of reptilian heart muscle should rule out the giraffe’s trick. More likely, their basic lung-shunting scheme, dividing cardiac output between interconnected systemic and pulmonary circulations, presents a barrier. We mammals (and birds) have no such connection and an unalterably serial circulation. Volume flow (Q) through the lungs must exactly equal volume flow through the systemic circulation, depriving us of the ability to reduce pulmonary flow during, for instance, diving. But we gain the ability to run the pulmonary circuit at a different pressure (Δp) (typically a fifth or sixth) than that elsewhere. In effect, we keep the cost (ΔpQ) of pulmonary pumping low with a reduced Δp; reptiles keep the cost low with a reduced Q. (Crocodilian reptiles, with optional shunting, may have the best of both worlds; but they live in a severely horizontal world so the problem is moot.)

Extant reptiles may mainly keep their heads down, but one must wonder about dinosaurs, those famously tall reptiles. To take an extreme case, *Brachiosaurus* may have carried its head as much as 8 m above the heart, with an overall height of 12 m (Gunga *et al* 1995). A *GHI* limit of 1.7 suggests an average heart-level blood pressure of 73,000 Pa (550 mm Hg). Recognizing the atypically low heart and using (10,000 + ρgh) instead gives a pressure of 92,000 Pa (690 mm Hg). Either far exceeds that of a giraffe. One must assume that *Brachiosaurus* kept its head up – as Carrier *et al* (2001) pointed out, carrying a head so far in front of the center of gravity would have severely impeded turning, and the vertebrae certainly permit such posture. Still, we can imagine a variety of solutions or evasions. Brief cranial anoxia may have been tolerated. Or perhaps these creatures had subambient cranial blood pressures, driving flow by the pull of siphons rather than the push of pumps. A partial solution may not be especially obscure. Birds evolved from (or are) dinosaurs, and birds have fully serial circulatory systems. That dinosaurs did likewise thus involves no great stretch of any evolutionary scenario, according to one of their intimates, Kevin Padian (personal communication).

3. To walk or to run

Almost all our terrestrial vehicles move on rotating wheels. Occasionally we even use temporary, axle-less wheels, moving heavy objects on rollers by shifting them from rear to front as they emerge, one by one. Physics imposes no irreducible minimum cost – only imperfect stiffness of wheels and path, friction of wheel bearings, accelerations, slopes, and air resistance impede motion. Railroads, with metal wheels and level, metallic tracks, could provide economic transport with the inefficient steam engines of two centuries ago, long before road vehicles could shift from draft animals. Wheels, especially with axles, are splendid devices.

No terrestrial animal goes from place to place on wheels and axles. One can argue (as did Gould 1981) that evolutionary constraints preclude their appearance. Or one can argue (as did LaBarbera 1983) that we easily overrate the utility of wheels, that they lack versatility and, in particular, work badly on either soft or bumpy surfaces. That latter argument receives at least tacit endorsement by recent attention (mainly military) to legged robots for off-road use, emulating the general arrangements of animals such as ourselves.

The use of legs may be widespread but it cannot be described as energetically efficient. However many legs an animal uses, it faces a basic difficulty that rolling wheels circumvent. Legs work by reciprocating rather than rotating, which means that any leg of finite mass must waste work accelerating at the start of a cycle and then decelerating again at the end. Of course an evasion comes immediately to mind – bank the decelerative work for reuse in the subsequent acceleration. What kind of short-term battery, then, might store that work? Electrochemical storage could be used, like the regenerative brakes of some hybrid automobiles, but no natural examples have yet come to light. Or inertial storage might serve, as in a flywheel. Again we can point to no obvious natural case, although bicycles, passive locomotory prostheses, make some use of the scheme.

Two kinds of brief batteries do find widespread use – lifting and then lowering masses against and with gravity, and straining and then releasing springs. Interestingly, animals cannot be dichotomized by their use of one or the other of these fundamentally different ways to store energy. Instead, most legged terrestrial animals depend on both, shifting from one to the other at a specific speed. At low speeds, gravitational energy storage does the job in what we call walking gaits; at higher speeds elastic energy storage serves in the various running gaits. It would be a rare culture that lacks specific words for at least these two gaits, so obvious is the distinction.

Quite recent – surprisingly recent – is the recognition that this shift from gravitational to elastic energy storage underlies the abrupt transition. Traditionally, walking gaits have no fully aerial phase while running gaits include at least a brief aerial phase. True enough, except for elephants (at least), which trot without an entirely aerial phase, but that classic distinction holds far less prescience. The realizations both that the basic game consisted of offsetting the inefficiency of legged locomotion and of the role of gait shifting we owe to R McNeill Alexander and his associates (Alexander 1976; Alexander and Jayes 1983 and other papers and books). In addition they have done as much perhaps as everyone else put together in working out its implications. The crux of the matter takes few words. In walking gaits, whether bipedal or polypedal, gravitational storage does the job, and almost the entire body mass contributes to the functional weight. In running and hopping gaits (trotting, galloping, cantering, skipping, bounding, etc.) stretched tendon does most of the work of elastic storage, with substantially lesser contributions from muscle and bone.

How gravitational energy storage can ease a task can be easily demonstrated. Swing a lower leg back and forth while sitting on the edge of a desk and measure the period. Plug that time, *t*, into the standard equation for a pendulum,

$$t = 2\pi\sqrt{\frac{l}{g}}, \qquad (4)$$

and you get an effective length, *l*. My 1.1 s swing predicts a length of 30 cm, a reasonable measure of the distance from knee to the leg's center of gravity. The exercise is not entirely trivial – it illustrates the ease with which one's neuromuscular system phases its output to maintain that frequency. Put on a heavy shoe, and you swing with a longer period, again with no initial awkwardness. Try to change swinging frequency and you find yourself working a lot harder. Similarly, when you walk, you immediately adopt a 'natural' pace, increasing or decreasing speed as much by changing stride length as by changing frequency. A pendulum length for a normal adult pace of 1.4 s per stride is about 50 cm, not unreasonable for hip to center of gravity of a leg – ignoring some bias and complications from the constrained motion of a leg in contact with the ground. About the location of the pendulum, though, the extrapolation from leg swing to walking misdirects us.

Just how gravitational storage operates in walking gaits turns out to be less easily specified; indeed it operates in a distinctly odd manner – perhaps the reason it escaped analysis for so long. Were our walking to resemble the swinging of an ordinary pendulum, we would reach greatest speed and our centers of gravity would be lowest in mid-step. In fact, we are highest, not lowest, and slowest, not fastest, in mid-step, as we vault over relatively extended legs. In addition, as we walk, we sway slightly side to side at half the frequency at which we move up and down.

Walking, again whether bipedal or polypedal, is commonly described in terms of the motion of an inverted pendulum. The head and torso provide almost all the relevant mass whose center of gravity matters, rather than the mass of the legs, despite their more rapid motion. As shown in figure 2, head and torso travel in a series of arcs, convex upward rather than downward, as in a conventional pendulum. One gets some idea of the way kinetic and gravitational energies interchange by thinking of an egg rolling end-over-end down a slope – speed and height of center of gravity peak at opposite phases of its motion. While an inverted pendulum does not correspond to an intuitively obvious physical model, the analogy has proven analytically powerful.

One might think of walking as a process of lifting one's center of mass and then allowing it to fall forward, the combination forming an arc. Gravity then imposes a distinct limitation by setting the downward acceleration of that forward fall. That allowed Alexander and Jayes (1978) to estimate the maximum speed of walking, using only a few empirically-supported assumptions. First, in walking, at least one leg must always be on the ground – that is, the "duty factor" or temporal ground contact fraction cannot be less than 0.5. And second, relative stride length – stride length over hip-to-ground length – should peak at the same value for walkers of any size. Finally, the walkers should be similarly proportioned and walk with similar maximum arc angles for their strides. They predicted that the limit on downward acceleration would limit walking speeds to a value no more than about 0.4 or 0.5 times a particular dimensionless ratio, v^2/gh. The latter is the quotient of forward speed, *v*, squared to gravity times a height, *h*, taken as that of the hip joint from the ground.

The ratio happens to have the same arrangement of variables as that between kinetic energy and gravitational potential energy,

$$\frac{mv^2}{mgh} = \frac{v^2}{gh}, \qquad (5a)$$

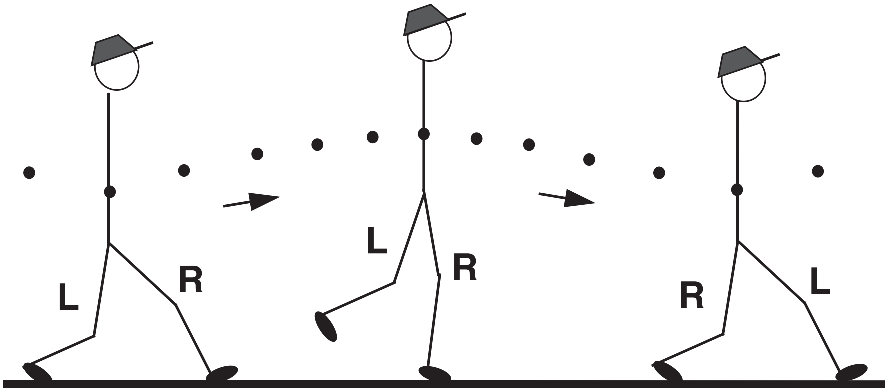

Figure 2. The motions of the body in a half-step of walking. At mid-stride the body is highest and the speed (indicated by distances between spots at standard intervals) is lowest. Changes in both have been exaggerated. (Adapted from Vogel 2003; see also Biewener 2003.)

where *m* is body mass and the factor of 2 in kinetic energy has been ignored. It also appears if inertial force is divided by gravitational force,

$$\frac{ma}{mg} = \frac{v^2}{gh} = Fr\ , \tag{5b}$$

noted as the Froude number, *Fr*, in the last essay (Vogel 2006) – the ratio introduced by William Froude in the 19th century as a scaling rule for models of the hulls of ships. In this last guise, it will reappear in the next essay. It can also be derived by ignoring the constant factor and squaring what is left of both sides of eq. (4), which itself can be obtained by simple dimensional analysis.

The ratio provides a specific rule for the relationship between animal size and maximum walking speeds, a rule with both explanatory and predictive value. And the rule works well for a very wide range of walkers, which hit maximum speeds at Froude numbers between 0.3 and 0.5 (Biewener 2003). Above that range of size-adjusted dimensionless speeds, animals switch to other gaits – we begin to jog, a dog begins to trot, a crow begins to hop. The greatest distance covered per unit energy expenditure occurs at about *Fr* = 0.25, the size-independent optimum walking speed. Had Alexander not pointed out Froude's precedence (albeit in relating at the wave lengths and speeds of surface waves), we would now be talking about the Alexander number. The diversity of organisms that follow the rule makes it a remarkable generalization. It stands as the classic illustration of how dimensionless ratios can serve biomechanics just as they serve mechanical (mostly fluids) engineering.

Animals of whatever size stress their bones to similar maxima when moving – about twice standing during walking and about five times standing in running – but do not exceed 50–100 MPa (Biewener 1990). With this range of maximal bone stress and the transition range of Froude numbers we can ask about the speeds of dinosaurs. The combination implies that the largest theropods such as *Tyrannosaurus* ran gingerly if at all (Alexander 1976; Hutchinson and Garcia 2002); conversely, they could walk exceedingly fast. And from the skeletal dimensions and trackways the walking speed of the 3-million year old Laetoli (Tanzania) hominids can be estimated. They were about a third shorter than modern humans and should have been slower by a similar factor (Alexander 1984).

We can also ask what might happen were the value of gravitational acceleration altered. Greater *g* should give a higher transition speed; lower *g* should give a lower transition speed. Humans on the moon, with a sixth of terrestrial *g*, found that hopping was a better way to get around than walking, which would have been (ignoring the effect of space suits) less than half as fast as on earth. Skipping, as done by children here on earth, was a useful gait as well (Minetti 2001). When walking on a (terrestrial) treadmill, partly supported by a traveling overhead harness, humans maintained the characteristic exchange of kinetic and potential energy of walking (Griffin *et al* 1999). And in brief exposures to truly altered gravity in maneuvering aircraft, maximum walking speed increased with the value of *g*, as expected from eq. (5) (Cavagna *et al* 2000).

One of the benefits of a rule is how it directs attention to apparent exceptions. Emperor penguins walk long distances at an especially high cost for their size. Their short legs mean that they are not geometrically similar to other birds – for their size, they make especially quick strides. That may preclude the usual arrangement for energy interchange, but they have another, side-to-side waddling. The high cost, then, does not come from abandonment of the interchange, but from the high rates at which the muscles running their short legs must generate force (Griffin and Kram 2000). Penguin walking appears to be close to a model developed by Coleman and Ruina (1998), a bipedal toy or robot (a "passive-dynamic walker") that goes down a slope with a side-to-side pendulum motion – a description of an easily-built model can be found at *http://ruina.tam.cornell.edu/research/topics/locomotion_and_robotics/*.

Bear in mind that on a level path, the entire cost of locomotion (ignoring drag) represents inefficiency. Although walking costs energy, the relative (mass specific) cost of body transport decreases as the size of animal increases. Most likely, its cost traces to a basic disability of muscle, the need to expend energy to produce force, even when moving nothing. The more rapidly we ask a muscle to develop force, the greater the cost, as just mentioned for penguins; the smaller the animal, the greater its stride frequency, and the greater the cost of level walking relative to its mass.

If the path slopes upward, walking incurs an additional cost, that of working against gravity, which scales with body mass. Combining the cost of level walking with the additional price of going upward explains a curious but familiar phenomenon. The relative difficulty of ascent depends on an animal's size. A horse walks more efficiently on the level than does a dog, but even a slight slope extracts a great fractional increase in demand for energy – quite familiar where animal-drawn vehicles provide transport. A small rodent handles slopes more easily than any dog, and those ants that construct roadways do so with magnificent indifference to slope, caring only about overall path length. Minetti (1995) applied treadmill data to predict the optimum slope of mountain paths, assuming a goal of gaining altitude cheaply. The slopes of paths in the Italian Alps corresponded nicely to the predictions, with switchbacks wherever the critical steepness would be exceeded. In theory, at least, one could predict the size of an unknown animal (perhaps a yeti) from the slopes of its paths.

4. To trot or to gallop

We bipeds have only a few variants on walking, such as flexed-leg rather than stiff-legged walking, race-walking and goose-stepping. To these we add several gaits that depend on elastic energy storage, such as running, hopping and skipping. Quadrupeds have a considerably wider range of possibilities for gaits that use elastic storage; of these the two most common are trotting and galloping. In trotting each of four legs strike the ground in a left-right symmetrical sequence – front-left plus hind-right, front-right plus hind-left. In galloping almost paired front and almost paired hind legs alternate, 'almost' because a leading side and thus some minor asymmetry is typical. Like trotting, galloping mainly stores energy from stride to stride as stretched tendon.

Several questions immediately occur. First, why gallop? Simply because by doing so an animal can go faster. Among other things, galloping permits recruitment of an additional mass of elastic in the back and elsewhere for energy storage (Alexander 1988). Moreover, after rising as trotting speed increases, cost relative to distance drops again following the shift to a gallop. The speeds of this second gait transition raise a second and more peculiar question. Among quadrupeds that gallop, the trot-to-gallop transition occurs within a fairly specific Froude number range, between 2 and 3 (Biewener 2003). Froude number, again, represents a ratio of inertial to gravitational force. In this second transition, oddly, both gaits use elastic energy storage and neither uses gravitational storage. So why should Froude number matter?

Perhaps we need to reverse the argument that explained the first transition. What determined that one was the upper practical speed for walking. Here, by contrast, what matters may not be an upper limit of trotting but a lower limit of galloping, a limit set by the maximum practical aerial period. Trotting has (elephants, again, excepted) only short periods when no foot makes contact with the ground, while galloping involves considerably longer aerial periods. And while airborne, an animal must fall earthward – with gravitational acceleration. Too long a fall, and an animal will not be easily able to position one or more feet on the ground beneath its torso. What can we make of that intuitively argument?

Assume an animal can fall a fixed fraction of leg length,

$$d \propto h \propto gt^2, \tag{6}$$

where d is distance fallen, h is leg length, and t is the time in free fall. What we need to know is how the speed at transition, v, varies with leg length. Heglund and Taylor (1988) report that it varies as one might expect, with leg length divided by stride time – basically all gallopers gallop in about the same way at the transition point. So

$$v \propto h/t\,. \tag{7}$$

Combining the two proportionalities to eliminate t and taking the reciprocal (if you are constant, so is your reciprocal) yields, in fact, the Froude number:

$$Fr = \frac{v^2}{gh}. \tag{8}$$

Can we go a step further and rationalize the particular value (or range) of Froude number at which transition occurs? We might assume that value and estimate the fraction of leg length that a galloper drops while airborne. Breaking speed into length per stride (l) and time per stride (t), we get

$$Fr = \frac{l^2}{hgt^2}. \tag{9}$$

Heglund *et al* (1974) reported a minimum galloping speed for a particular horse of 5.6 m s^{-1} at a frequency of 2.0 Hz. Alexander *et al* (1980) found that the stride length of a galloping horse is about 5 times its hip height. Adjusting that down from average to minimum speed (using the speeds of Heglund *et al* 1974 and Heglund and Taylor 1988) gives 3.4 times hip height, the later about 1 m (from a skeleton). The final item needed is the fractional duration of the airborne periods at minimum galloping speed. Here specific data seems lacking – people care far more about how rapidly than how slowly horses can gallop! I will assume two periods, each of 25% of stride duration, noting that relative time airborne will be at its lowest at minimum galloping speed.

These data give a stride duration of 0.69 s and thus airborne periods of 0.172 s each. During each period, gravity will make the horse fall 0.145 m, about 15% of the hip to ground distance. That does seem a practical maximum for getting feet positioned for the next stride, again noting the very rough character of the estimate.

5. The height of trees

Surely trees provide the paradigmatic examples of gravitationally responsive organisms. Each is a tall column that keeps a crown of photosynthetic structures elevated in the face of a gravitational force that would prefer otherwise. It does so to win access to sunlight in competition with other trees – greater height cannot bring it significantly nearer the sun. Each of the lineages in which tree-like organisms have evolved from shrubbier or herbaceous ancestors has used the same basic material, wood. In each tree or tree-like system, water must be extracted from the substratum and lifted to leaf level, typically through evaporation at the top and consequent suction below. Despite considerable structural and developmental diversity among the lineages, their tallest members have achieved about the same maximum heights, roughly 25 to 100 m (Niklas 1997). Explaining such consistency tests our understanding of the biological consequences of gravity.

Perhaps the column of stacked bricks invoked at the start of this essay might provide an instructive analogy. Wood has a compressive strength of about 50 MPa and a density of about 500 kg m^{-3} – better specifications than brick, incidentally. A column (as in figure 3a) could extend 10 km upward even with no taper, a hundred times the height of the tallest contemporary tree. Clearly resistance to compressive crushing imposes no limit.

But crushing mainly afflicts short, wide columns. A more likely failure mode is so-called Euler buckling, the sudden collapse that occurs when the middle of a column bows ever further outward (as in figure 3b). Elastic modulus, rather than compressive strength, now becomes the operative material property. For fresh wood we can assume a value of 5 GPa (Cannell and Morgan 1987), noting that the compressive moduli run slightly lower than the tensile (Young's) moduli but that trees compensate for the difference with some tensile prestressing. Trunk thickness becomes relevant because buckling stretches one side and compresses the other. The standard equation for Euler buckling (see Vogel 2003 or standard handbooks for mechanical engineers) gives a height of well over 100 m for a trunk diameter of 1 m. This assumes that the tree does not taper and that its entire weight is concentrated at the top – an unrealistically harsh scenario. Offsetting (at least in part) those biases, trunks are assumed straight and their bases firmly fixed. Even admitting the simplifications, though, it appears that gravitational loading through buckling imposes no practical limit.

We might look at the tree in yet another way, a simplified version of Greenhill's (1881) classic analysis. Consider a brief lateral perturbation near the top of a tree from wind or some other cause. That will move the center of gravity laterally, tending to make the tree topple. At the same time, it will generate an opposing elastic restoring force in the wood. In effect, this treats the tree as a self-loaded cantilever beam (as in figure 3c), albeit one extending upward rather than outward. If lever arm and restoring force scale linearly with deflection distance, then that distance drops out. Young's modulus also drops out since in practice it varies directly with the density of the wood. Again adopting standard equations and the standard relationship between Young's modulus and density, our 1 m tree can extend upward about 120 m before the wood of the tree reaches maximum tolerable stress. Again, trees rarely approach that value. And, again, more realistic assumptions would raise the limiting height – I have once more assumed that the tree does not taper, which raises the center of gravity, and I have assumed that it pivots at the bottom rather than bending, which moves its mass too far outward.

Still, while both views – a column subject to Euler buckling and a cantilever beam – give unrealistically great heights, both say that height will scale with diameter$^{2/3}$. The girth of the taller tree will be disproportionately great, something easily observed. Quite a few sources note that particular scaling rule, going back at least to Greenhill's (1881) prediction and including McMahon's (1973) compilation from data on 576 trees in the United States, each of either record height or record girth for its species. The arguments for the rule have become much more sophisticated, in particular accounting for taper and crown weight (Niklas 1992 has a good discussion).

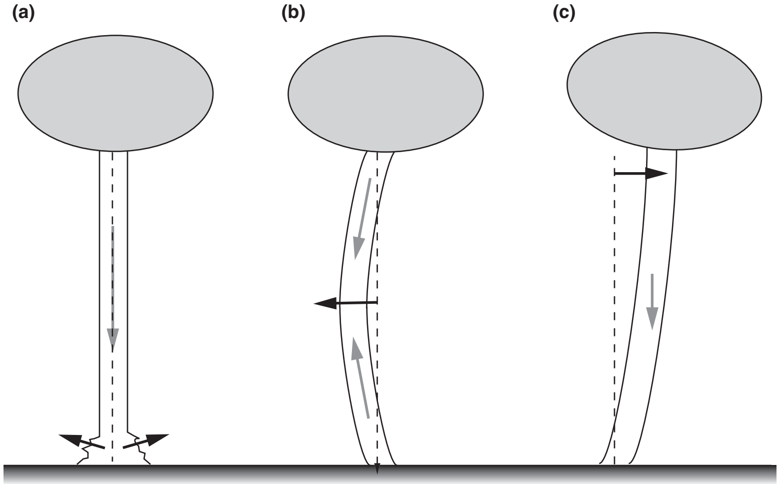

Figure 3. **(a)** A column failing by simple compressive failure – crushing. **(b)** A column, also end-loaded, failing by "Euler buckling," a mode in which, paradoxically, one side experiences tensile loading. **(c)** A column loaded sideways as if a cantilever beam, in which, once bent, its own weight generates a turning moment about the base.

That each of several starting assumptions yields the same scaling rule gives little help in choosing among them. Worse, we should not place great confidence in that exponent of 2/3, however often it gets cited. For few biological systems can we find so much data to test such a rule – for obvious reasons, books on practical silviculture, forest mensuration, and so forth pay great attention to the height and girth of trees. I tried a few regressions on published data and was rewarded with exponents ranging from about 1/3 to 4/3. Unsurprisingly, practical people concerned with timber production rely on other, more complex formulations – see, for instance, Johnson and Shifley (2002).

Moreover, many sources question the whole notion that the strength, density, and elastic moduli of wood determine the maximum heights and proportions of trees. The most common alternative views the limit as hydraulic, the problem of lifting water from the roots to such biologically prodigious heights. Our hearts develop systolic pressures during exercise of perhaps 25,000 Pa, and tall mammals when running probably approach twice that. Just working against gravity, a 100 m tree has to move water against a pressure difference of 1,000,000 Pa, 40 times better than our personal best. Worse, the main pump depends on suction from above rather than pushing from below, that is, on negative rather than positive pressure.

The main mechanism for raising water needs a few words, especially because at first encounter nearly every physical scientist expresses skepticism or outright incredulity. Evaporation across tiny interfaces in the feltwork of fibers of the cell walls of cells within leaves draws water out of the soil and up through a large number of small conduits (xylem) just beneath the bark. Surface tension at these interfaces (around 0.1 μm across) should have no trouble keeping air from being drawn in at the top – the surface tension of pure water can sustain a pressure difference of nearly 3,000,000 Pa, almost 30 atmospheres, across such a tiny interface (Nobel 1999).

But then things get decidedly unconventional. Atmospheric pressure can push water up to a maximal height, defined by eq. (1), that corresponds to a pressure difference of 101,000 Pa at sea level – the difference between that of the atmosphere and a full vacuum. For water (or xylem sap), with a density of 1,000 kg m^{-3}, that height is 10.3 m. Evacuate a vertical tube and place the open end in water, and the water will rise to that height, with a vacuum above. Of course if a clean pipe a bit longer than 10.3 m is initially fully filled with water containing little dissolved gas, one may have to bully the system a bit for the water level to drop and the vacuum to appear. In the interim, the water column will have developed a pressure below 0 Pa, a slight and brief negative pressure.

Even if water is freely available at ground level and can be raised without frictional loses, trees should be able to grow no higher than 10.3 m – unless they can capitalize to a fabulous degree on such negative pressure. Before taking offence at the notion of negative pressure, pause to observe that the water in question is liquid, not gaseous. The internal intermolecular cohesion that makes a liquid a liquid rather than a gas should render it perfectly capable of withstanding tension, the more sanitary term for negative pressure. The difficulty comes from containing a liquid while subjecting it to tensile stress. Not only must its intermolecular cohesion withstand the stress, but the adhesion of the liquid to the walls of the container must do the same – neither grip can fail or a vacuum will appear. In addition, very little gas or other impurities can be dissolved in the water, so ordinary soil water must be pre-processed before entering the main conduits.

Trees apparently meet these demanding conditions and raise sap despite severely negative pressures. A field-usable device (a so-called Scholander bomb – see Scholander *et al* 1965) makes possible routine measurements of negative pressures in plants by indicating the positive pressures required to counterbalance them. –1 or –2 MPa (–10 or –20 atm) pressures are common, and values as extreme (one hesitates to say ‘high’) as –12 MPa. (–120 atm) have been reported (Schlesinger *et al* 1982). In laboratory tests, macroscopic quantities of water have resisted tensile stresses of hundreds of atmospheres, so the picture does not rely solely on calculated intermolecular forces.

Other things being equal, the taller the tree, the more extreme the negative pressures. And the more extreme the pressures, the greater the danger that liquid within some conduit will cavitate, interrupting the process and putting that conduit out of action as if it were an unprimed pump. Cavitation does occur with some regularity – this is no hypothetical hazard – with a large fraction of the conduits in a normal tree sometimes embolized. In practice, the greater the diameter of the conduits running up the tree, the greater the likelihood of cavitation (Ellmore and Ewers 1986; Maherali *et al* 2006). But recent work (see, for instance Holbrook and Zwieniecki 1999 and other papers by each of these authors) has revealed specific devices to minimize the propagation of embolisms and to repair embolized conduits.

Trees face a curious balancing act. Their demands for water vary over a wide range, low in conifers, for instance, and high in many broad-leaved trees. Beyond the gravitational loss of 9,800 Pa m^{-1} (from eq. 1), making the water move raises another kind of loss, that due to the fluid-mechanical resistance of the conduits. The general rule for pressure drop per unit length ($\Delta p/l$) due to laminar flow in circular conduits is the Hagen-Poiseuille equation (here given in terms both of total flow, Q, and maximum, axial, flow speed, $v_{\max}$):

$$\frac{\Delta p}{l} = \frac{8\mu Q}{\pi r^4} = \frac{4\mu v_{\max}}{r^2}\,. \tag{10}$$

μ is the fluid's viscosity and r the radius of the conduit. Whether one considers total flow or flow speed, the smaller the conduit the worse the pressure drop. In addition, passage of sap between adjoining conduits entails additional losses – see, for instance, Lancashire and Ennos (2002). One might argue that a tree should move water in pipes large enough to keep the cost of flow low but not so large that embolizing becomes an excessive risk. And in enlarging pipes to reduce losses from flow, trees must meet diminishing returns – after all, that gravitational loss of 9,800 Pa m^{-1} remains.

Thus we expect conduit sizes will strike a balance, large enough to keep flow losses down to the same order as gravitational losses but not much larger. What do we find? Maximum flow speeds *in vivo* can be measured by heating a trunk locally and then timing the interval before a thermocouple located somewhat higher detects a temperature change. I calculated pressure drops per unit length for a variety of trees (and a liana) from a variety of sources, using measured averages of maximum speeds and conduit diameters from Milburn (1979), Zimmermann (1983), Gartner (1995) and Nobel (1999). The data cover a 10-fold range of diameters and a 100-fold range of speeds; the resulting pressure drops range from 1,300 to 20,000 Pa m^{-1}, that is, from 13% to 200% of the gravitational drop, with little evident regularity. But the data is highly heterogeneous, reflecting spread in conduit diameters within individual trees, uncertainty about which ones happen to be active and not embolized at a particular time, variation in flow speeds with time of day and wetness of season, and so forth.

Nonetheless, the values do not disagree with the notion that trees balance the diminishing returns and increasing risk of enlarging conduits, keeping a fairly fixed relationship between flow and gravitational losses. Put another way, why should a tree risk making conduits large enough to reduce flow loss much below the unavoidable gravitational pressure loss? At the same time, the values provide at least indirect support for the idea that the difficulty of lifting water imposes a general limitation on forest height.

That, though, is hard to reconcile with lots of data showing that gravitational pressure drops and the flow losses predicted from the Hagen-Poiseuille equation commonly do not represent the largest part of the overall negative pressures measured at tree-top heights. A further pressure drop come from extracting water from less-than-saturated soil ("matrix potential" sometimes), osmotic processes in roots, and (as noted) flow through the pits and plates that divide the ascending tubes of xylem. Trees 20 or 30 meters high often develop pressures of –2 MPa or more, far above a twice gravitational drop of –0.4 to –0.6 MPa. For that matter, the record of –12 MPa mentioned earlier comes from measurements on a desert shrub, not a tree, and mainly results from the scarcity of soil water. By contrast, Koch *et al* (2004) measured an extreme pressure of –1.8 MPa 4 M below the top (112 M) of the tallest known tree, a redwood (*Sequoia sempervirens*). They found in the laboratory that a pressure of –1.9 MPa imposes serious loss of hydraulic conductivity on such material and therefore argued that hydraulics limits height. The skeptic wonders if the closeness of those figures, –1.8 and –1.9 MPa, merely tells us that such trees conduct and utilize water no better than they have to.

We also face the awkward fact that especially wide conduits occur in woody vines (lianas), with diameters sometimes exceeding 300 μm. But vines, unlike trees, need not support themselves; their dry densities are concomitantly low. Xylem, we remind ourselves, is wood, both a conductive and a supportive tissue. One suspects that relaxation of their supportive function at least in part underlies the size of these conduits. And that suspicion points back to mechanical support as the main limitation on height.

Before dismissing hydraulics, though, we should note another way it might bear relevance. Recently Niklas and Spatz (2004) have related both maximum tree height and the basic 2/3-power scaling to the problem of supplying an ever-increasing overall leaf area with water – an argument based on supply rather than pumping cost. I like their rationale but remain bit skeptical. The quantities of water that trees raise and transpire are almost as impressive as the pressures against which they do so. But these quantities far exceed the amounts used in photosynthesis and vary widely. Nobel (1999) notes a 40-fold range in water use efficiency – rate of carbon fixation divided by rate of water use. Furthermore, just as with pressure, the most extreme values (here high ones) come from plants living in dry habitats rather than from especially tall trees.

In short, the original question remains without a satisfactory resolution. We may even be looking at the wrong variables. In trying to choose between two different routes through which gravity might affect tree height, we presumed a gravitational limit. Even that presumption may be suspect. First, healthy trees rarely fail by gravitationally driven mechanical collapse. (Occasional windless ice storms where I live do cause trees to fail gravitationally.) Second, the correspondence between conduit size and flow speed and acceptance of a considerable rate of cavitation suggests that still wider conduits could be tolerated – conduits such as those of lianas. Finally, the fact that negative pressures at tree top level exceed, sometimes by large factors, the sum of both gravitational and flow-induced pressure drops suggests that still greater losses from these latter quarters could be tolerated.

Perhaps the limit on height, paradoxically, might sometimes come from something other than gravity. Trees blow over in storms, most often by uprooting, less often by snapping of their stems near their bases, still less by shear-induced snapping higher up. Whichever way, failure most likely results from drag, acting on the crown; the taller the tree, the longer the lever arm and the greater the turning moment. In such a scenario, the lateral drag of the crown, mainly due to its leaves, imposes the critical disadvantage of height. Several structural features of leaves and trees (and bamboo culms, etc.) make functional sense as devices to reduce vulnerability to drag, often termed "wind throw", and their ubiquity argues that drag surpasses gravity as a hazard.

The commonness of uprooting, in particular, implies that much of the problem of a tree must come from a peculiarity of tree substratum, the limited resistance of soil to tensile forces. Shear and compression soil can resist, and its weight above buried roots may assist, but many trees may not be able to pull on the ground with particular effectiveness. At one time, perhaps somewhere still, large stumps were pulled directly upward by teams of horses solely with the aid of simple windlasses that could be moved from stump to stump.

Trees may stay upright in winds in several ways (Vogel 1996):

(i) With a long, stiff taproot that extends the trunk downward a tree can take advantage of the shear and compression resistance of soil. If lateral roots near ground level fix the location of the base of the tree, blowing the trunk one way asks that the taproot be forced the other way, compressing and shearing soil. The array of smaller, vertical 'sinker' roots from larger horizontal ones may work the same way, as well as providing significant resistance to uprooting tension through shear numbers and area covered. That combination of tap root, laterals, and sinkers seems to be central to the support system of many trees (pines, paradigmatically) of temperate and boreal forests, trees whose trunks obviously bend in winds.

(ii) Some tension resistance in the most superficial soil layer can come from the tangle of roots of surrounding vegetation, something many tropical trees take advantage of with large, thin, upwind buttresses. These act like diagonal cables from trunk to roots rather than the compression-resisting buttresses of Gothic architecture – the misleading linguistic analogy confused things until recently (Smith 1972; Ennos 1993). Again, sinker roots assist. Trees with such tensile buttresses tend to be thin relative to their heights.

(iii) Ground level lateral extensions of the trunks of many big temperate-zone broad-leaved trees are lower and thicker; they most likely work as conventional downwind buttresses that take advantage of soil's reliable compression resistance – as well as providing attachment points for sinker roots. Trees with these wide, heavy bases ('plates' sometimes) typically have thick trunks of dense wood that do not bend noticeably in winds. The arrangement comes into use as a tree matures and shifts from system (i).

The wide bases and stiff trunks of system (iii) may convey another message. I have argued that the vulnerability of trees to wind-throw shows that gravity need not always be the physical agency that limits height. Compressive buttressing and thick, stiff trunks suggest that gravity may at times operate on the other side of the equation, assisting a tree in staying erect. When trees such as large oaks blow over, the bases of the trunks often lie 1 or 2 m above the ground; by contrast, pine trunks lie directly on the ground. Thus in uprooting, compressively buttressed trees pivot around a horizontal axis well to the side of the axis of the trunk, as in figure 4. To make a tree uproot, the turning moment must exceed the stabilizing moment – the product of drag times the height of the center of the crown must

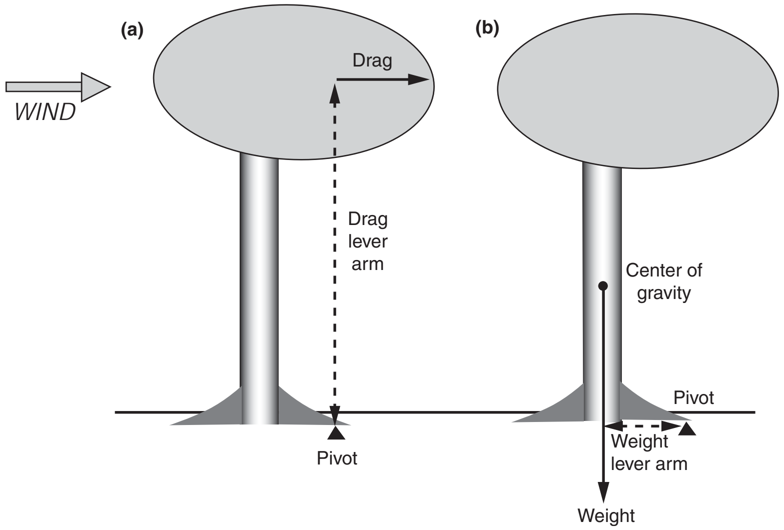

Figure 4. The drag of a wind loads a tree not as a column but as an end-loaded cantilever beam. A tree with stiff trunks and basal, compression-resisting buttresses, will suffer "wind-throw" when the turning moment from drag and the height of the crown **(a)** exceeds the opposing moment from its weight and the width of the buttressing **(b)**.

exceed the product of the weight of the tree times the distance from trunk axis to turning axis. That simple view ignores any contribution from soil around the roots, of sinker roots, and so forth. But it exposes the possibility that such a tree might use its weight to stay upright with its sinker roots to keep from sliding sideways.

Does such a model survive quantification? Consider a tree with 30 m of cylindrical trunk, 0.7 m in diameter, of a density of 1000 kg m^{-3}, a pivot point 1.5 m to one side of the trunk's vertical axis, an otherwise weightless basal plate, and a weightless, spherical crown of branches and leaves. Using symbols for the variables described in figure 4, the stabilizing moment will be

$$\rho_{tree}\ \pi r_{tree}{}^{2} h g r_{base}. \tag{11}$$

The tipping moment will be the drag of the crown times the height of the tree,

$$0.5 C_d\ \rho_{air}\ \pi r_{crown}{}^{2} v^{2} h. \tag{12}$$

Assuming a drag coefficient, C_d, of 0.1, appropriate for a large sphere in fast flow, an air density of 1.2 kg m^{-3}, and a speed of 35 m s^{-1}, we equate (11) and (12) and solve for the radius of the crown. It comes to almost 5 m, and thus a diameter of nearly 10 m. While perhaps a little smaller than one observes in nature, it comes close enough to suggest taking this model of an oddly detached tree seriously.

Still, I must emphasize its crudeness. We have distressingly little information on the real drag of this kind of broad-leaved tree in high winds. I did some work on the drag of individual leaves and small clusters (Vogel 1989), enough to undermine confidence in any extrapolation or estimate for whole crowns, something Ennos (1999) has reemphasized. Besides the obvious logistical problems, people who run sufficiently large wind tunnels do not take kindly to tests of items expected to fail by detaching pieces just upwind from valuable and vulnerable fans and motors.

Note, though, what the model says about the relevant variables. First, wind speed has a severe effect on the result. Second, height does not directly matter, since it equally affects the weight of the tree and the moment arm of its drag. Greater height does, though, require that the trunk be wider to have the additional flexural stiffness needed to minimize lateral movement of its center of gravity. Of course wider means heavier and thus gives further improvement of a tree's stability. Finally, gravity itself aids stability, as in eq. (11), so if gravity were greater, such a tree might be able to grow taller – unless, as suggested in the last essay (Vogel 2006) air density (and thus drag) were thereby also increased. But whatever the specific value of g, in this model the tree depends on gravity to stay erect.

Whatever the limitation on height, it must most often operate through the competitive interactions of individual trees. If height does scale with diameter to the 2/3 power and thus cross section to the 1/3 power, then successive increments in height demand making ever increasing amounts of wood. Better access to sunlight than one's peers extracts an ever increasing constructional penalty. Furthermore, growing significantly above canopy level should disproportionately increase peak wind speeds and thus drag. So any cost-benefit analysis ought to include competitive interactions and growth. And growth depends on a host of other factors; thus the dipterocarp forests of Southeast Asia, growing on rich, volcanic soils, achieve greater canopy height than tropical forests elsewhere on earth. Givnish (1995) expands on this kind of argument, noting the ever-decreasing ability of a tree in a forest to compensate for cost with increased leaf area.

I must admit some attachment to a picture that emphasizes the lateral force of wind, a bias stemming from my own interest in air flow and drag. So I hasten to remind the reader (and myself) of the old adage that when one's tool is a hammer, all problems resemble nails. It well may be a case, as said of raccoon- and opossum-hunting dogs in this part of the world, of barking up the wrong tree.

6. The diverse roles of gravity

In aerial systems, gravity impels dense bodies downward, with only the relationship between size and descent speed at all negotiable. In terrestrial systems gravity may be less insistently intrusive, but it plays a wider range of roles. Here we moved from cases where the role of gravity was straightforward to ones in which it played increasingly subtle roles – clearly important, but in ways that challenged our analyses. But I conclude with a mild caution, noting that many other cases might have been considered as well as the present ones, that this essay just scratches the surface. The present essay might have compared impact loading with gravitational loading in various forms of locomotion. It might have noted the shift in mammalian posture from flexed-legged to straight-legged, a likely consequence of the way body weight scaled with volume, while postural muscle force scaled with cross-section. Or it might have suggested that an alteration gravity's strength (or wood's strength-density relationship) would affect the length and taper of branches more than it would the overall height of trees.

In these essays I have made much of scaling rules and their particular exponents; the way blood pressure depends on body size illustrates one hazard of the approach – a real threshold effect that would be missed by the normal regression-based scaling analysis. For gait transitions we do have a scaling rule, based on Froude number, but here the rule itself applies to thresholds. For tree height, we examined the near constancy of forest heights over space and time, suggestive of mechanical (solid or hydraulic) limitation. Not only could we not pinpoint the limitation, but we could not either confirm or discredit a scaling rule – or even convince ourselves fully that gravity contributed to the limit.

Acknowledgements

Here, as in its predecessors, the scale and diversity of the relevant literature has been daunting; fortunately I received guidance at various points and times from R McNeill Alexander, Steve Churchill, Tim Griffin, Rob Jackson, Paul Manos, Andy Ruina and Kevin Padian – plus several of our university librarians.

References

Alexander R M 1976 Estimates of speeds of dinosaurs; *Nature (London)* **261** 129–130

Alexander R M 1984 Stride length and speed for adults, children, and fossil hominids; *Am. J. Phys. Anthropol.* **63** 23–27

Alexander R M 1988 Why mammals gallop; *Am. Zool*. **28** 237–245

Alexander R M and Jayes A S 1978 Optimum walking techniques for idealized animals; *J. Zool. (London)* **186** 61–81

Alexander R M and Jayes A S 1983 A dynamic similarity hypothesis for the gaits of quadrupedal mammals; *J. Zool. (London)* **201** 135–152

Alexander R M, Jayes A S and Ker R F 1980 Estimates of energy cost for quadrupedal running gaits; *J. Zool. (London)* **190** 155–192

Baudinette R V 1978 Scaling of heart rate during locomotion of mammals; *J. Comp. Physiol.* **B127** 337–342

Biewener A A 1990 Biomechanics of mammalian terrestrial locomotion; *Science* **250** 1097–1103

Biewener A A 2003 *Animal locomotion* (Oxford: Oxford University Press)

Calder W A 1984 *Size, function, and life history* (Cambridge, MA: Harvard University Press)

Cannell M G R and J Morgan 1987 Young's modulus of sections of living branches and tree trunks; *Tree Physiol*. **3** 355–364

Caro C G, Pedley T J, Schroter R C and Seed W A 1978 *The mechanics of the circulation* (Oxford: Oxford University Press)

Carrier D R, Walter R M and Lee D V 2001 Influence of rotational inertia on turning performance of theropod dinosaurs: clues from humans with increased rotational inertia; *J. Exp. Biol*. **204** 3917–3926

Cavagna G A, Willems P A and Heglund N C 2000 The role of gravity in human walking: pendular energy exchange, external work, and optimal speed; *J. Physiol*. **528** 657–668

Coleman M J and Ruina A 1998 An uncontrolled walking toy that cannot stand still; *Phys. Rev. Lett.* **80** 3658–3661

Dawson T H 2005 Modeling of vascular networks; *J. Exp. Biol.* **208** 1687–1694

Ellmore G S and Ewers F W 1986 Fluid flow in the outermost xylem increment of a ring-porous tree, *Ulmus americana*; *Am. J. Bot.* **73** 1771–1774

Ennos A R 1993 The function and formation of buttresses; *TREE* **8** 350–351

Ennos A R 1999 The aerodynamics and hydrodynamics of plants; *J. Exp. Biol.* **202** 3281–3284

Gartner B L 1995 Patterns of xylem variation within a tree and their hydraulic and mechanical consequences; in *Plant stems: physiology and functional morphology* (ed.) B L Gartner (San Diego, CA: Academic Press) pp 125–149

Givnish T J 1995 Plant stems: biomechanical adaptations for energy capture and influence on species distributions; in *Plant stems: physiology and functional morphology* (ed.) B L Gartner (San Diego, CA: Academic Press) pp 3–49

Gould S J 1981 Kingdoms without wheels; *Nat. Hist.* **90** 42–48

Greenhill A G 1881 Determination of the greatest height consistent with stability that a vertical pole or mast can be made, and of the greatest height to which a tree of given proportions can grow; *Cambridge Philos. Soc.* **4** 65–73

Griffin T M, Tolani N A and Kram R 1999 Walking in simulated reduced gravity: mechanical energy fluctuations and exchange; *J. Appl. Physiol*. **86** 383–390

Griffin T M and Kram R 2000 Penguin waddling is not wasteful; *Nature (London)* **408** 929

Grubb B 1983 Allometric relations of cardiovascular function in birds; *Am. J. Physiol.* **245** H567–H572

Gunga H -C, Kirsch K A, Baartz F, Röcker L, Heinrich W-D, Lisowski W, Wiedemann A and Albertz J 1995 New Data on the dimensions of *Brachiosaurus brancai* and their physiological implications; *Naturwissenschaften* **82** 190–192

Heglund N C, Taylor C R and McMahon T A 1974 Scaling stride frequency and gait to animal size: mice to horses; *Science* **186** 1112–1113

Heglund N C and Taylor C R 1988 Speed, stride frequency and energy cost per stride: how do they change with body size and gait?; *J. Exp. Biol.* **138** 301–318

Holbrook N M and Zwieniecki M A 1999 Embolism repair and xylem tension: do we need a miracle?; *Plant Physiol.* **120** 7–10

Hutchinson J R and Garcia M 2002 *Tyrannosaurus* was not a fast runner; *Nature (London)* **415** 1018–1021

Johnson P S and Shifley S R 2002 *The ecology and silviculture of oaks* (New York: CABI Publishing)

Koch G W, Sillett S C, Jennings G M and Davis S D 2004 The limit to tree height; *Nature (London)* **428** 851–854

LaBarbera M 1983 Why the wheels won't go; *Am. Nat.* **121** 395–408

Lancashire J R and Ennos A R 2002 Modelling the hydrodynamic resistance of bordered pits; *J. Exp. Bot.* **53** 1485–1493

Maherali H, Moura C F, Caldeira M C, Willson C J and Jackson R B 2006 Functional coordination between leaf gas exchange and vulnerability to xylem cavitation in temperate forest trees; *Plant, Cell Environ.* **29** 571–583

McMahon T A 1973 Size and shape in biology; *Science* **179** 1201–1202

Milburn J A 1979 *Water flow in plants* (London: Longmans)

Minetti A E 1995 Optimum gradient of mountain paths; *J. Appl. Physiol*. **79** 1698–1703

Minetti A E 2001 Walking on other planets; *Nature (London)* **409** 467–468

Niklas K J 1992 *Plant biomechanics: an engineering approach to plant form and function* (Chicago: University of Chicago Press)

Niklas K J 1997 *The evolutionary biology of plants* (Chicago: University of Chicago Press)

Niklas K J and Spatz H-C 2004 Growth and hydraulics (not mechanical) constraints govern the scaling of tree height and mass; *Proc. Natl. Acad. Sci. USA* **101** 15661–15663

Nobel P 1999 *Physicochemical and environmental plant physiology,* 2nd edition (New York: W H Freeman)

Nowak R M 1991 *Walker's mammals of the world,* 5th edition (Baltimore: Johns Hopkins University Press)

Pedley T J, Brook B S and Seymour R S 1996 Blood pressure and flow rate in the giraffe jugular vein; *Philos. Trans. R. Soc. London* **B351** 855–866

Schlesinger W H, Gray J T, Gill D S and Mahall B E 1982 *Ceanothus megacarpus* chaparral: a synthesis of ecosystem processes during development and animal growth; *Bot. Rev*. **48** 71–117

Schmidt-Nielsen K 1997 *Animal physiology: adaptation and environment,* 5th edition (Cambridge, UK: Cambridge University Press)

Scholander P F, Hammel H T, Bradstreet E D and Hemmingsen E A 1965 Sap pressure in vascular plants; *Science* **148** 339–346

Seymour R S and Blaylock A J 2000 The principle of Laplace and scaling of ventricular wall stress and blood pressure in mammals and birds; *Physiol. Biochem. Zool.* **73** 389–405

Seymour R S and Arndt J O 2004 Independent effects of heart-head distance and caudal pooling on pressure regulation in aquatic and terrestrial snakes; *J. Exp. Biol*. **207** 1305–1311

Smith A P 1972 Buttressing of tropical trees: a descriptive model and new hypotheses; *Am. Nat.* **106** 32–46

Vogel S 1989 Drag and reconfiguration of broad leaves in high winds; *J. Exp. Bot*. **40** 941–948

Vogel S 1996 Blowing in the wind: storm-resisting features of the design of trees; *J. Arboriculture* **22** 92–98

Vogel S 2003 *Comparative biomechanics* (Princeton NJ: Princeton University Press)

Vogel S 2006 Living in a physical world. VI. Gravity and life in the air; *J. Biosci.* **31** 13–25

Weisskopf V F 1975 Of atoms, mountains, and stars: a study in qualitative physics; *Science* **187** 605–612

Zimmerman M H 1983 *Xylem structure and the ascent of sap* (Berlin: Springer-Verlag)

Zweifach B W 1974 Quantitative studies of microcirculatory structure and function. I. Analysis of pressure distribution in the terminal vascular bed in cat mesentery; *Circ. Res*. **34** 843–857

*e*Publication: 5 May 2006

Living in a physical world
VIII.
Gravity and life in water

STEVEN VOGEL
Department of Biology, Duke University, Durham, NC 27708-0338, USA
(Fax, 919-660-7293; Email, svogel@duke.edu)

1. Introduction

Life was born in water, and aqueous habitats still hold most of life's diversity. The near-aqueous density of most organisms ensures something close to suspension by the surrounding water. A creature might be twice as dense as the medium but never, as on land or in the air, a thousand times as dense. Gravity? We might expect it to exert only a minimal impact on design and deportment. But that contrast in relative density between aquatic and non-aquatic life may mislead us.

As touched on in connection with the ascent of sap in trees in the last essay (Vogel 2006), gravity induces a change in hydrostatic pressure with height or depth of roughly 10,000 Pa m^{-1}, an atmosphere for every 10 m. How potent must that gradient be in oceanic water columns of hundreds or thousands of vertical meters! The hydrostatic squeeze will mercilessly compress a bubble of air or any other gas. Gases – pure or mixed – follow Boyle's law of 1662, the law that volume varies inversely with pressure. A bubble of air at a depth of, say, 10,000 m, that of deep ocean trenches, will have only about 0.1% of its volume at the surface – the 1000-fold pressure increase will result in a 1000-fold volume decrease. Unless the local water is air-saturated or the bubble is impermeably encapsulated, it will in short order redissolve, now a victim of Henry's law and Laplace's law, both of early 19th century origin. Henry's law declares that increased pressure leads to increased solubility of gases in liquids; Laplace's law (as we now know it) says that the smaller a bubble, the greater the internal pressure due to the squeeze of surface tension. Maintaining a gas under water thus bumps into the twin difficulties of depth-dependent volume (Boyle, augmented by Laplace) and dissolution rates (Henry).

That implies major effects of gravity on aquatic life. Again, we can easily be misled. What about a bubble of some liquid, perhaps a vacuole of lipid? Or a cell, separated from the ocean by a lipid membrane? Or some solid material such as bone or chitin? For liquids and solids, no analogous rule links pressure and volume, and their responses diverge dramatically from that of a gas. Pressure increase produces almost no volumetric change. The descriptive variable here (lacking a general rule) is the bulk modulus, K (or its reciprocal, the compressibility). K is the ratio of change in pressure, Δp, to change in volume, ΔV, relative to original volume, V_o:

$$K = \frac{\Delta p}{\Delta V / V_o}. \tag{1}$$

Most liquids and solids have very high bulk moduli. It is often said that water is incompressible, but the implied infinite modulus is an exaggeration. Fresh water has a bulk modulus of about 2.1 GPa, seawater about 5% more (sources vary on the next significant figure). So seawater is about 4% denser at the bottom of a deep ocean trench than at the surface. These are ordinary values for liquids – the bulk moduli of pure hydrocarbons (octane, for instance) run about half water's value, but such oils as cells might put in vacuoles (vegetable oil, in one tabulation) differ little from water. Solids run one to two orders of magnitude higher, which is to say that they compress even less easily – glass has a bulk modulus of about 40 GPa and steel about 160 GPa. Even allowing for some pressure-dependent variation of values, in its hydrostatic manifestation gravity should matter little to either liquids or solids.

Pressure exerts slightly more influence on chemistry. At a depth of 10,000 m, altered hydrogen bonding of water increases its dissociation constant, 2.5-fold at 20°C, for instance (Hills 1972). Thus at extreme depths life faces significant – but not overwhelming – changes in buffering, protein configurations, membrane permeabilities, and so forth. DNA is stable at up to ten times the 1000-atm pressure of that depth; while less barostable, proteins still denature only slightly at that pressure (Suzuki and Taniguchi 1972). By comparison, the fall in temperature with depth (and the lowered metabolic rates) causes substantially greater changes.

What about the surface of a pond or ocean, with air above and water below? Again, gravitational effects can range from profound to trivial. We know that a liquid's surface prefers to be horizontal and smooth. Gravity provides the main impetus for both, but smoothing involves another agency as well. Surface tension demands that work be done to create additional interfacial area, so it contributes another smoothing force. Disturbances of the smooth surface propagate as waves, and surface tension as well as gravity determines their behaviour. For water beneath air, surface tension sets the predominant rules for what we call "capillary waves," those with wavelengths below 17 mm. For instance, among these waves, shorter wavelength means faster rather than slower propagation. Before dismissing capillary waves as of relevance only to whirligigs and water striders, bear in mind that every big wave started small, with a wavelength at which surface tension ruled.

Nonetheless, most phenomena at the interface between sky and sea can be attributed to gravity. I mention their existence and note their importance before putting that interface aside for another occasion – here I mean to look only at what happens well beneath the water's surface. Even so, what follows must be recognized as an idiosyncratic selection of phenomena. No space will be given to density gradients caused by depth-dependent changes in temperature or salinity and thus to thermoclines and salt-wedges. Nor to the depth limitations of a chest-powered breathing snorkel, nor to the increasing effectiveness with depth of suckers such as those of an octopus. Attention will be limited to a few interrelated situations – problems of handling undissolved gases and of ballast and buoyancy control.

Of particular interest, as we will see, are the diverse instances in which submerged organisms maintain stores of air or other gases. They do so for either (on occasion, both) of two main reasons. For some, air breathers, what matters is the gaseous oxygen in the mix. Others use a gas to counteract body densities greater than that of the surrounding water – for flotation. Organisms may store gases internally, in cuttlebone and diverse bladders, or externally, as bubbles or body sheathing. Gas stores may be long-lasting or require periodic replenishment, the latter from secretion or transport downward from the surface. Only the shallowest and most turbulently moving water will be gas-saturated at pressures corresponding to their depths rather than contain gas equilibrated with the atmosphere above. Aquatic organisms containing gas can face Henry's law from either direction. Sometimes gas must be kept from disappearing into solution lest an air breather sink or asphyxiate; sometimes gas must be kept in just such solution lest it tear up tissue or impede circulation. However, one categorizes the schemes, the possibilities are many.

A final prefatory note – in moving from the aerial and surface worlds of the previous essays, we need to shift from forces and accelerations to changes in pressures, volumes and solubilities.

2. Using surface tension to extract gases from water

Surface tension can provide the functional equivalent of waterproofing, doing the job well enough to prevent bulk gas loss from a bubble and leaving dissolution as the remaining concern. In my youth, I was taught that one could knot each leg of a pair of pants, wet the fabric, and use it, held upside down, as a float – not that I ever knew anyone driven to do it. (A pillow case needs no knots and provides an easy test of the device.) Either air or water passes through the pores in the fabric with little resistance, but the interface between them cannot do so, at least if the fabric contains no trace of a laundry-day surfactant. At least one spider, *Argyroneta aquatica*, uses an analogous air store. It makes its web within the submerged vegetation of ponds. An obligate air-breather like other spiders, it fills a silken bell, analogous to an old-fashioned diving bell, with air that it carries down from the surface. Periodically it adds air to the bell to offset both oxygen use and dissolution.

How fine a mesh must the web have to prevent escape of air, assuming (as seems to be the case) high hydrophobicity? Surface tension (0.073 N m^{-1} in fresh water at 20ºC) keeps the air contained; hydrostatic pressure forces the air upward through the mesh. For a spherical shell whose radius of curvature is r and corresponding diameter is d, the pressure developed by surface tension (γ) is

$$\Delta p = \frac{2\gamma}{r} = \frac{4\gamma}{d}. \tag{2}$$

Hydrostatic pressure, of course, is simply

$$\Delta p = \rho g h, \tag{3}$$

where ρ is the density of water and h is the depth beneath the surface. Thus

$$d = \frac{4\gamma}{\rho g h}, \tag{4}$$

so the tighter the mesh (lower d), the deeper the spider can dwell. For depths up to 10 cm , the strands should be no more than 0.3 mm apart – not a particularly daunting requirement. [Alternatively one can set up a dimensionless ratio for the practicality of using surface tension to maintain air under water by dividing eq. (2) by eq. (3).] Neither its overall curvature nor the volume of air in the bell make a difference.

Thus for a given surface tension, density, and gravitational acceleration, depth and mesh size vary inversely with each other. Dissolution rate, of course, cares nothing about mesh size and will increase with depth; dissolution rate and the distance air must be transported downward probably limit the arrangement more than does web mesh. Schuetz and Taborsky (2003) note that these spiders seem bothered by excessive buoyancy when they bring air down to their bells, so they appear to pay for transport.

Many insects do much the same thing, if on a less impressive scale. Virtually all adult insects and many of their aquatic larvae, nymphs and pupae require access to gaseous as opposed to dissolved oxygen. Many hold air bubbles, periodically renewed by trips to the surface, in their various external irregularities – between body segments, beneath wings and elytra, and so forth.

A serendipitous physical phenomenon increases the persistence of such bubbles. Even in water that flows almost unnoticeably (Vogel 2004), solubility rather than diffusion coefficient determines how fast gases diffuse in or out of a bubble. And much less nitrogen than oxygen dissolves in a given volume of water (1.7% versus 3.5% by volume at 15ºC and 1 atm; Krogh 1941). A bubble thus loses nitrogen more slowly than oxygen. Since air is mostly nitrogen (about 80%), the presence of nitrogen thus increases the persistance of a bubble compared to one of pure oxygen. In a classic experiment of Ege, in 1918 (cited by Thorpe and Crisp 1947), water bugs with air-filled bubbles could manage 7 h submergences. By contrast, bugs with bubbles of pure oxygen lasted only 35 min – oxygen consumption paled beside its outward dissolution, which rapidly decreased bubble size and hence effective surface area.

Could the respiratory depletion of an external air store be offset by net inward diffusion or some other device instead of periodic renewal at the surface? Harpster (1941) and Brown (1987) suggest that photosynthetically-produced oxygen might be acquired from aquatic plants, but no quantitative investigation seems to have been done. Still, we do know of at least two ways to compensate for outward diffusion and respiration. One depends entirely on surface tension and was established in a series of papers by Thorpe and Crisp (1947), the first of them being of greatest present interest; see also Thorpe (1950). The other, more unusual, depends primarily on a hydrodynamic effect and is a consequence of Bernoulli's principle. Stride (1955) showed how it works.

3. Surface tension and plastrons

First, let us consider the role of surface tension. It has long been known that a thin film of air covers some submerged adult insects, a layer conspicuous as as a silvery sheen, much as one sees on a suddenly submerged leaf of nasturtium or lotus. Evidence for its respiratory role goes back at least to Comstock (1887), and Harpster (1944) proved that some insects could maintain the film without periodic replacement, at least in water with dissolved gases at atmospheric partial pressures.

The mechanism depends on the relationship expressed by eq. (2). Usually we apply the equation to bubbles convex on the outside – as they normally are. The smaller the bubble, the greater the component of pressure inside caused by surface tension. So tiny ones can spontaneously disappear as that pressure drives their gaseous contents into solution. For this thin film of air, or 'plastron' (Thorpe and Crisp

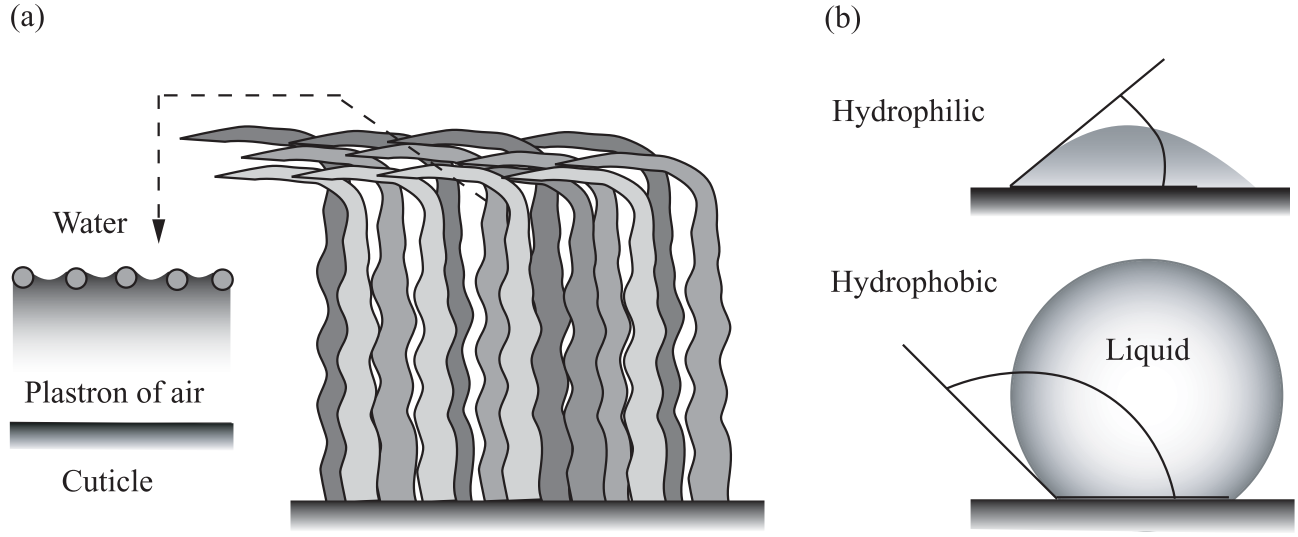

Figure 1. (**a**) Diagrammatic representation of a plastron; (**b**) contact angles for an aqueous liquid on hydrophilic and hydrophobic surfaces.

1947), one has to look at the implication of eq. (2) for a bubble of a seemingly unlikely geometry, one concave rather than convex on the outside. Smaller will now imply lower, not higher, internal pressure. With sufficiently small and numerous bubbles of this sort, enough oxygen could diffuse inward to supply an air-breathing animal's respiratory needs.

So submerged adult insects need a lot of tiny, concavely curved bubbles or else some kind of air layers characterized by interfaces with such curvature. They create such layers by coating themselves with air, forming air-water interfaces at the periphery of a dense layer of short, hydrophobic, cuticular hairs, as in figure 1a. In their initial report, Thorpe and Crisp (1947) estimated that *Aphelocheirus*, a naucorid bug, had about 2,000,000 hairs mm^{-2} – thus individual hairs less than 1 μm apart.

Using newer and better imaging equipment and the same species, Hinton (1976) revised that to 4,000,000 hairs mm^{-2}, with hairs tapering from 0.4 to 0.2 μm in diameter and extending 3 μm outward from the body. As in figure 1a, their distal portions are bent parallel to the surface, with little space between them. Treating the erect parts of the hairs as columns with the typical stiffness of arthropod cuticle (10 GPa) and assuming that a hair is vulnerable to Euler buckling, Hinton (1976) calculated that buckling such an array of hairs would require a pressure of about 40 atm. So the buckling strength of the hairs would impose a depth limit for use of a plastron of 400 m, which poses no problem for an entirely freshwater fauna. That hydrophobicity is less than perfect, so failure of surface tension restricts plastron use to lower hydrostatic pressures and thus shallower depths. Both Thorpe and Crisp (1947) and Hinton (1976) found that, in practice, plastrons break down through wetting at about 3 atm (above ambient pressure). Still, 3 atm corresponds to a depth of about 30 m, fairly deep by the freshwater standards of insects.

The relative role of the geometry of the outer part of the plastron and of its wettability remain uncertain. The usual measure of hydrophobicity is contact angle, that between the surface of a bubble and the surface it contacts, as in figure 1b; an angle of 180º would indicate perfect hydrophobicity. The hairs (and cuticle generally) cannot be perfectly hydrophobic – we know nothing that extreme. Contact angles for ordinary waxy coatings range between 105º to 110º, the range assumed in most calculations of plastron performance. Recent work, though, has shown that some biological systems achieve higher angles. The surfaces of the leaves of lotus and some other plants that have complexly sculptured waxy cuticles can reach 160º (Barthlott and Neinhuis 1997; Neinhuis and Barthlott 1997). A recently developed biomimetic super-hydrophobic coating, 'lotusan', (Sto Corp., Atlanta, GA, USA) is similarly and usefully self-cleaning. Of especial relevance here, Wagner *et al* (1996) reported angles as high as 155º on insect wings, and Gao and Jiang (2004) reported an angle of 168º for the legs of a water strider – an insect for which high hydrophobicity should be particularly advantageous. These high values depend on surfaces with roughness of the same scale (Feng *et al* 2002) as the conspicuous bumps on the plastron hairs in Hinton's (1976) scanning electron micrographs.

Plastrons turn out to be widespread among arthropods and have undoubtedly evolved many times, perhaps because a hydrophobic exoskeleton with minute outgrowths represents nothing out of the ordinary and because no more diverse group of air-breathers occurs elsewhere in nature. They occur among eggs and larvae that suffer occasional floods (Hinton 1976), and they appear as well in some millipedes, mites, and whip scorpions (Hebets and Chapman 2000). I wonder about plant leaves, most of which have hydrophobic outsides and many of which have fuzz as well as stomata on their undersides (or on both surfaces), but I know of no data indicating any analogous functional arrangement. I also wonder whether diffusion alone can adequately transport oxygen to the spiracles through a gas layer only 3 μm thick or whether some additional physical device awaits recognition. (An earlier essay, Vogel 2005, raised the possibility of an analogous transport limitation within leaves.) Diffusion-augmenting bulk gas motion within the plastron might be induced by movement of an insect through the water or by such things as local water pumping by hindleg motion, the latter as reported by Harpster (1941).

4. Using flow to extract gases from water

Far less common than plastrons as a way to maintain air under water is flow-induced local pressure reduction. Only one case has been well documented, a few others remain conjectural; as we will see, all too few locations meet its physical requirements.

In addition to hydrostatic pressure, the surface of an object in a flowing fluid feels the pressures of that flow. The specific pressure on a location on the surface depends on its location. At some upstream point pressure is maximal, with a value that corresponds to the local hydrostatic pressure plus a component from conversion of the flow's kinetic energy to pressure. Bernoulli's principle gives the pressure increase (over the local hydrostatic pressure) at that point as

$$\Delta p = \frac{\rho v^2}{2}, \tag{5}$$

where v is the speed of flow before slowing by the object. Downstream, pressures are inevitably lower, with specific values determined by location and the object's shape.

These downstream pressures (Δp's) are commonly expressed as their ratios to that maximum; the resulting dimensionless variable goes by the name "pressure coefficient," C_p:

$$C_p = \frac{2\,\Delta p}{\rho v^2}. \tag{6}$$

Thus a graph of pressure coefficient versus distance on the surface from upstream to downstream must always start at the y-axis with a value of $C_p = 1.0$, as in figure 2. In effect, pressure coefficients adjust pressures for the effects of speed and fluid density.

Not only does the pressure coefficient never reach 1.0 anywhere downstream, it drops below zero over much (usually most) of the rest of the body. The positive region turns out to be surprisingly limited, not even extending back to where the body is thickest. (For unstreamlined objects the pressure coefficient remains below zero back to the rear end, while for streamlined objects it gradually returns to positive territory, eventually approaching but not reaching 1.0.) Of present relevance, the overall pressure coefficient, integrated over the entire body, will almost always be negative; the particular value depends mainly on the body's shape. As a result, a bubble held stationary in a flow develops a net outward pressure coefficient. That coefficient ought to lie between about -0.1 and -0.3, with more negative values for broader and less tapered bubbles. In short, the pressure inside will drop below the pressure outside (Vogel 1994).

(By integratrating over the surface, taking local surface orientation into account, one can calculate drag, the downstream force on an object. Here we are concerned, instead, with transmural pressures.)

Here we see another way to extract gas from solution, a way whose operating range can be defined quite simply. Assume, as is common in rapidly moving, shallow water, saturation with air at atmospheric – or surface – pressure. If the flow-induced pressure drop in a bubble exceeds the local hydrostatic pressure increase due to depth, then the bubble should act as a gas extractor. The condition, then, for gas extraction, can be expressed quite simply by combining eq. (3) and eq. (6) as a ratio,

$$\frac{\Delta p_{flow}}{\Delta p_{hydrostatic}} = \frac{-C_p}{2}\frac{v^2}{gh} > 1.0\,. \tag{7}$$

(The minus sign on the right reflects the comparison of a pressure drop with a pressure increase.) Incidentally, the dimensional variables on the right form the Froude number, the ratio of inertial force to gravitational force, mentioned in

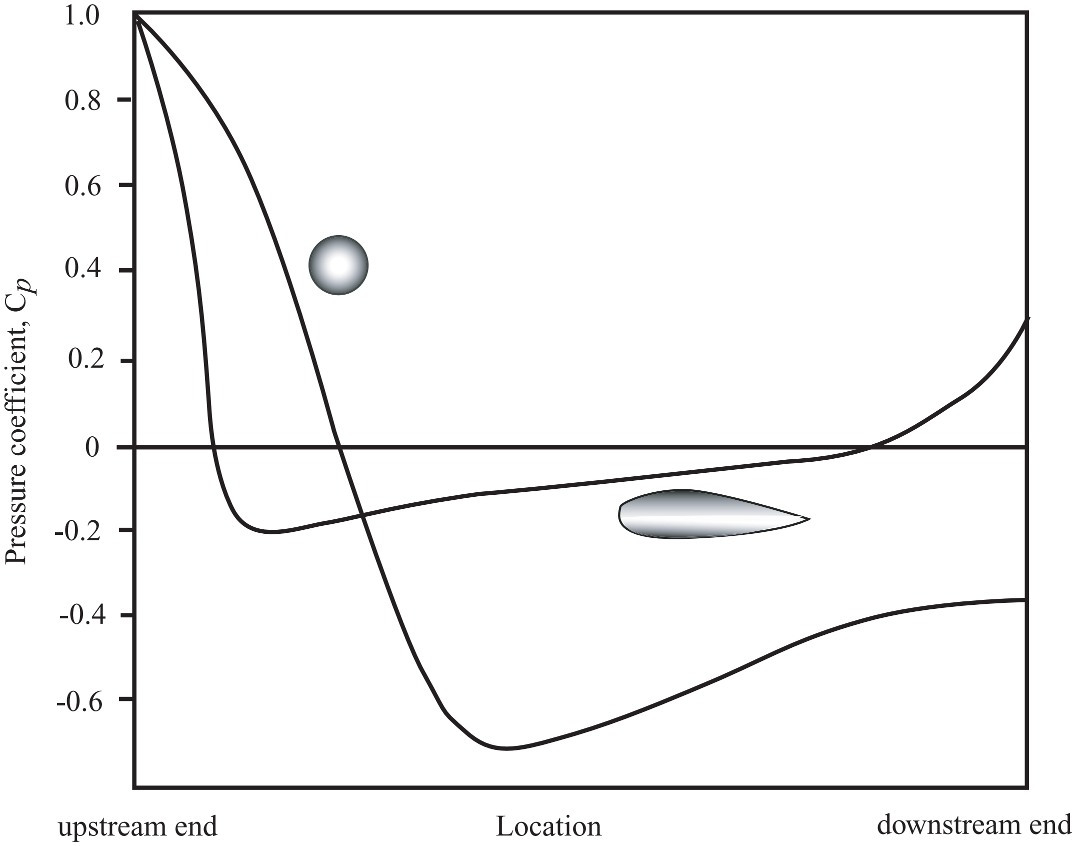

Figure 2. Pressure coefficient versus location on the surfaces of two objects – a sphere and a streamlined body of revolution. Both are about 40 mm in diameter. The data come from tests in air at 10 m s^{-1} and thus a Reynolds number (based on diameter) of 25,000.

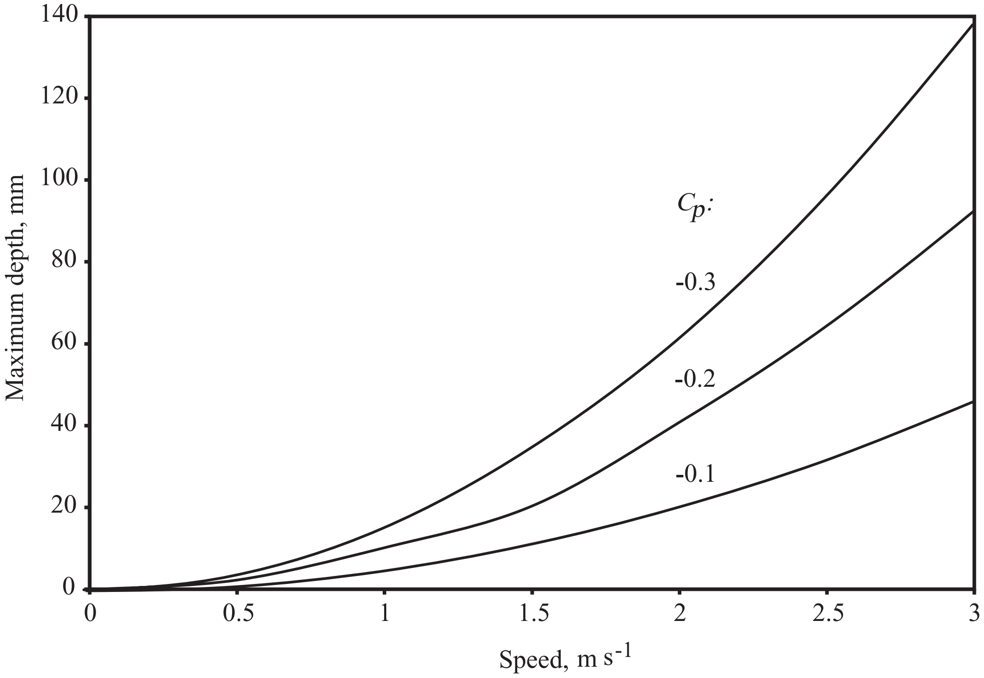

Figure 3. The maximum depth at which a bubble of air can be maintained by flow-induced gas dissolution for three different pressure coefficients. The water is assumed saturated with air at atmospheric pressure.

the previous two essays. Here it appears in the guise of flow force over hydrostatic force.

But eq. (7) defines a daunting condition. A bubble must be maintained despite the drag of the flow around it. And flow works far less effectively than surface tension as a pressure-reducer, at least when compared to surface tension acting across interfaces with minute radii of curvature. So the bubble must be held by an animal in the face of a substantial current, and it cannot be far from the surface. Figure 3 defines the limits for three possible pressure coefficients. It suggests that at a current speed of 1 m s^{-1} a bubble could persist at depths between about 5 and 15 mm. Worse, given the inevitable velocity gradients near surfaces, an organism holding onto some solid surface may be exposed to a local flow substantially slower than that of the mainstream. Were it not for one well-documented case, we might dismiss the scheme as creative but impractical. That case, investigated by Stride (1955), merits some attention.

Working in Ghana, Stride (1955) noticed that adults of a particular kind of elmid beetle, *Potamodytes tuberosus*, often "appeared to fly straight into the rushing water" of a rapid stream and then congregated on rocks just beneath the surface. Each faced upstream and carried "a large silvery air bubble". In the laboratory, its bubble persisted indefinitely if – and only if – rapid, shallow flow enveloped the beetle. With some difficulty, Stride managed to measure the pressure within bubbles on restrained beetles subjected to a range of flow speeds. At the test depth of about 10 mm, bubbles persisted at speeds above about 0.8 m s^{-1}.

I have reanalysed his data, extracting the flow-induced pressure reduction from the background hydrostatic pressure of 98 Pa (that 10 mm depth), and put his 34 points, in contemporary units, on figure 4. They do not correspond to a specific value of pressure coefficient: the faster the flow, the lower its apparent value. That must come in part from change in bubble shape with flow speed, as he notes; but changes in the air-water interface just above the beetles probably contribute as well. A linear regression nicely fits his data and as nicely misleads, implying, by extrapolation, an impossible flow-induced pressure drop of 69.3 Pa with no flow at all. Nonetheless, as the figure shows, the data do correspond to reasonable values of pressure coefficient, giving confidence in both his measurements and the present analysis.

Yes, the conditions may be daunting, but Stride noted that such beetles were common enough in his area, and Brown (1981) believes that beetles of the genera *Hispaniolara* and *Potamophilops* play the same game. Moreover, elmids are not the only kind of riffle beetle. Adult riffle beetles use plastron respiration, so they already have spiracular connections to an outer store of air (Brown 1987). Furthermore, sub-surface photosynthesis (or dissolution of entrained bubbles from local waterfalls) can raise the partial pressure of oxygen beyond atmospheric level, which would allow a bubble to be maintained to greater depth – although Stride excluded

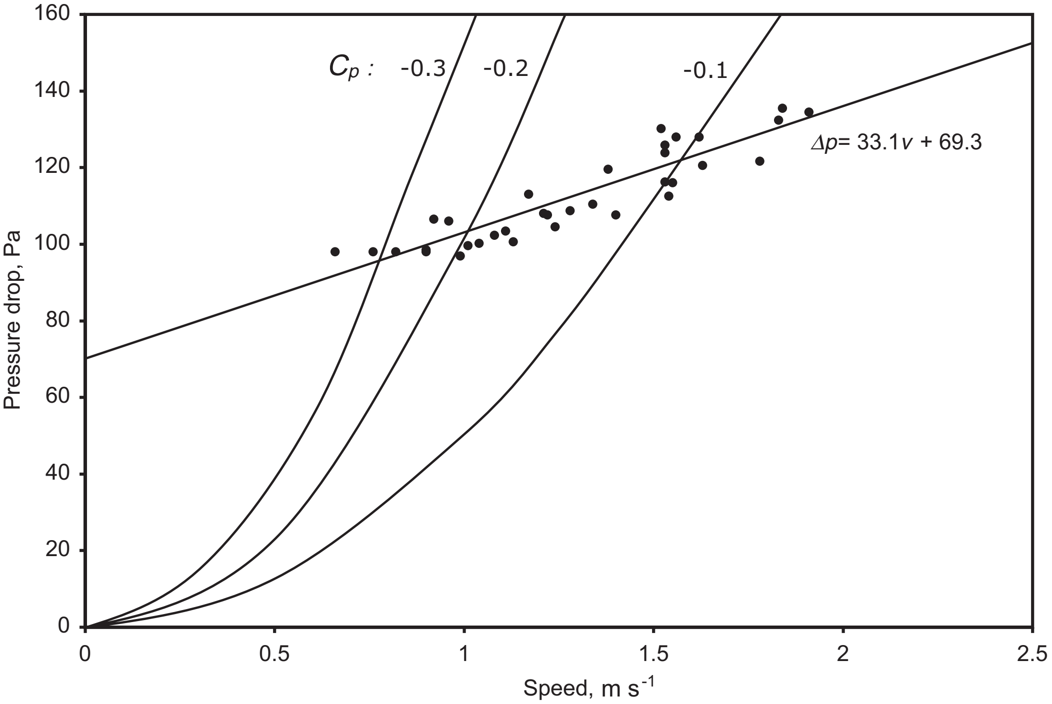

Figure 4. The pressure reduction caused by flow around a gas bubble – the predictions for three pressure coefficients and the measurements (points and fitted equation) by Stride (1955).

the possibility in his particular system. Finally, his beetles show no great anatomical adaptations, which implies both that such a trick can be done with behaviour alone and that lurking cases might easily be missed. Perhaps mere inattention explains why no others have surfaced in the intervening half-century. The composition of a diverse but globally consistent "torrential fauna" has been well-studied, although more with regard to who lives where than to any functional issues. The older accounts, especially Hora (1930) and Nielsen (1950), are still worth reading in any search for candidates.

Beyond the "plunge-and-grab-on" trick of *Potamodytes*, other ways of using current to maintain an underwater store of air must be possible. According to Pommen and Craig (1995), the plastrons of pupal net-winged midges (blephariceríd flies) produce bubbles that then persist in the low-pressure vortices behind their gills. Let me suggest yet another arrangement on the chance that it will either stimulate a specific investigation or consideration of still others.

The larval stages of many pyralid Lepidoptera (especially the Nymphulinae) are entirely aquatic, although pupae and adults typically live in air. In at least some genera – *Aulocodes* in northern India (Hora 1930) and *Elophila* and *Petrophila* (= *Parargyractis,* sometimes *Paragyractis*) in North and South America – the last larval instars spin tightly woven cocoons atop submerged rocks in rapids. The pupa then rests on a shelf within the cocoon, as in figure 5. The space above the shelf is consistently and persistently air-filled, air of unknown origin as put by Nielsen (1950) and Resh and Jamieson (1988). All descriptions mention upstream and downstream holes that allow ingress and egress of water and flow beneath the shelf.

Since the openings are small and close to the substratum, I doubt if water flow through the cocoon would be rapid enough for oxygen extraction. But flow across the top could do the job. Pressure will be locally reduced by the velocity increase (shown in figure 5b as compression of streamlines) needed to bring water over the cocoon. Animals do use the locally reduced pressure as fluid flows over a protrusion for a variety of purposes – it has been shown in systems such as sponges (increasing flow for suspension feeding), keyhole limpets (increasing respiratory water movement), and giant termite mounds (increasing respiratory air movement) (Vogel 1994). Surface tension between air and water within the outer wall of the cocoon should add the equivalent of a plastron, one in which the lower pressures would permit a coarser mesh and weaker structure. Unlike the situation in an uncontained bubble, in a plastron a solid structure helps offset inward fluid pressure. That gives relevance to Lloyd's (1914) comment that the outer walls of the cocoon "are spun of thick inflexible layers of silk." Perhaps these lepidopteran pupae combine two mechanisms, the surface tension of plastrons with flow-induced persistent gas bubbles.

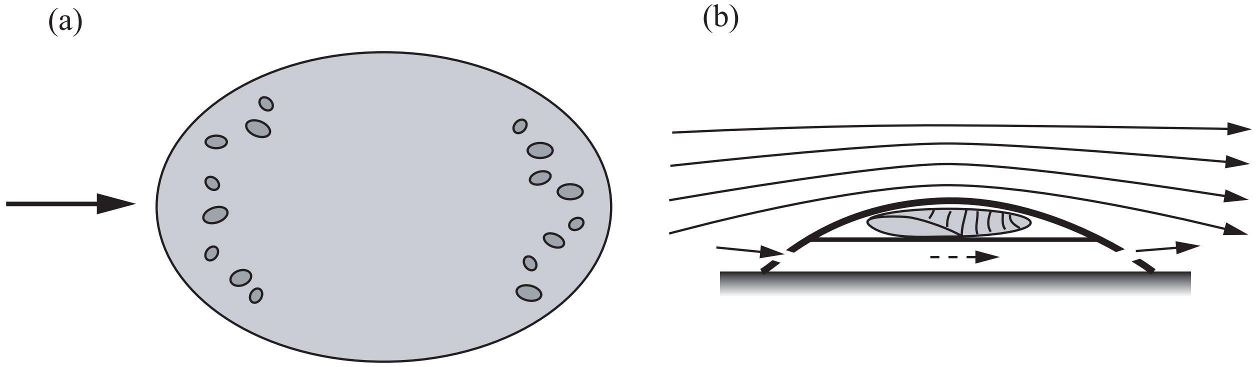

Figure 5. The cocoon (about 15 mm long) and pupa of a pyralid lepidopteran. (**a**) Top view of the cocoon, showing upstream and downstream openings. (**b**) A diagrammatic cross section of the cocoon, with the pupa resting on a shelf within an air-filled upper chamber (incorporating agreed-upon features from a variety of drawings and descriptions).

That raises the question of whether ordinary plastrons on insects that live in rapid flows take advantage of currents to reduce pressures still further than could surface tension alone – Pommen and Craig (1995) mention the possibility. My guess is that the effect, while inevitable, must be relatively minor except in very shallow water. After all, currents that an insect might encounter and withstand will generate at most a few hundred pascals of pressure, while unassisted plastrons can withstand pressures a thousand times greater.

5. Dealing with buoyancy

Gravity acts not so much on an organism's mass *per se*, as on the difference between that mass and the mass of the fluid that the organism displaces. While terrestrial organisms have densities much higher than that of the air around them, aquatic organisms differ little in density from water, and that displaced fluid cannot be ignored. Still, that difference varies from organism to organism and even, through both passive and active alteration, for a given organism over time.

Marine organisms without solid supporting materials have densities close to that of the surrounding fluid (around 1026 kg m^{-3}) and live in a weightless (although emphatically not massless) world. Minor adjustments in such variables as ionic composition can handle any residual weight (or buoyancy), often setting a slight downward bias. An equivalent freshwater organism can lower its density to that of the surroundings with a small body of one or another lipid, whose densities run around 900 kg m^{-3}. (Cholesterol, with a density of 1067 kg m^{-3}, cannot be used, while squalene, at 860 kg m^{-3} is especially effective.) Since lipids have about the same compressibility as water, buoyancy does not depend on depth.

Trouble can arise from either insufficient or excessive density. Thus the low densities of biological materials can hamper the ability of sedentary organisms to stay put as water flowing over them imposes both drag and lift or when insufficient density limits locomotion by surface purchase at the bottom. The problem of insufficient density can be simply remedied by adding biosynthesized or environmental stony material to the system. For instance, various insect larvae (trichopterans, most notably) incorporate tiny pebbles into their cases. They reportedly use larger pebbles in swifter flows, although that may just be a result of differences in what they have at hand (Pennak 1978).

Still, one can recognize clear cases of deliberate density increase. A wide variety of air-breathing aquatic vertebrates, both fossil and living, swallow and retain stones ('gastroliths') to offset the buoyancy that comes with air-filled lungs – plesiosaurs, some crocodilians, some pinniped mammals, some penguins, and others (Taylor 1993). Marine gastropod and bivalve mollusks commonly have thick shells of calcium carbonate that at least on some occasions must help them hold position in currents. Sessile adult bivalves that lack specific attachment devices (such as the byssus threads that tie mussels to rocks) tend to be the ones with the thickest shells. One specific case may be instructive – a tiny surf clam (*Donax variabilis*) depends on staying near the substratum as wave swash or backwash moves it up or down a beach. According to Ellers (1995), its relatively thick and dense shell gives it an overall density of 1650 kg m^{-3}; the densities of other bivalves, mostly larger, from the same beaches, range from 1170 to 1660 kg m^{-3}.

But density costs little in marine habitats, because sea water is usually either saturated or supersaturated with calcium, and because stones of greater density than calcium salts are common enough. The more aquatic mustelid mammals seem to have denser bones – but with ambiguous functional significance (Fish and Stein 1991). Good organisms for exploring functional increases in density might be the freshwater mussels – all shelled, diverse in size and flow speed preferences, and bottom-dwelling but not attached. Many, perhaps most, should be at some risk of dislodgement through sediment erosion during floods. One encounters anecdotal statements about thicker shells beneath more rapid waters (as in Pennak 1978), but no systematic study that rules out some simple scaling rule for size versus density or shell thickness seems to have been done (M C Barnhart, personal communication).

Alternatively, the presence of solid supportive systems can raise densities enough to make pelagic organisms sink. Compensatory adaptations for increasing buoyancy seem to be more common than those for reducing it, perhaps because the macroscopic supportive systems of big animals usually reduce their buoyancy. (Hydroskeletons are the most widespread exception.) Bones and stony corals have densities around 2000 kg m^{-3}, mollusk shells of calcium carbonate (as calcite or aragonite) around 2800 kg m^{-3}, and crustacean exoskeletons of calcified chitin about 1900 kg m^{-3} (Wainwright *et al* 1976). In most groups, species with larger individuals devote a greater fraction of body mass to support – even, if less dramatically, in aquatic groups – so the problem gets worse with increasing size.

Some animals take a "brute force" approach, producing sufficient lift to offset their negative buoyancy while swimming uninterruptedly. The paradigmatic examples are pelagic sharks. Lift comes from the combined action of asymmetrical caudal fins – larger lobes above than below– and a body pitched nose-upward. The fin asymmetry causes a downward tilt to the tail's rearward force (29º below horizontal in a leopard shark), generating lift as well as the thrust needed for swimming. Trouble from the posterior line of action of that lift is offset by additional lift from a flattened head and upward body pitch (11º in a leopard shark), which has an anterior line of action (Wilga and Lauder 2002). Still, sharks minimize the need for lift with skeletons that are much less calcified than those of bony fish of similar size – 'bony' recognizes just that difference. And squalene (with an especially low density of 860 kg m^{-3}, as noted) makes up a large fraction of their lipid.

Squid, likewise, lack flotation devices to compensate for negative buoyancy, and they also make do with a minimum of stiff material. Their main skeletal element is a light, thin lengthwise 'pen' that keeps the upstream (posterior) end from bending when the jet gives a forceful squirt. Since it can direct its jet downward, a squid need not make headway to maintain enough lift, yet it must still work – hovering costs about twice as much as does resting, and almost as much as does normal locomotion (Webber *et al* 2000).

A few bony fish, members of an obscure group that lacks swimbladders, take the shark game a step further. These Antarctic notothenioids have bones with only a trace of ossification – the ashed skeletons weigh only around 0.4% of body weight rather than a typical 2%. And they are full of lipid, mainly triglycerides of about 930 kg m^{-3}, located subdermally and in intermuscular sacs (DeVries and Eastman 1978; Eastman and DeVries 1982). With near-perfect neutral buoyancy, they need produce no hydrodynamic lift.

6. Using gases at local pressure for buoyancy

Gases underlie the most space-efficient schemes for buoyancy augmentation since at ordinary pressures the densest gas weighs much less than the least dense liquid. At atmospheric pressure, air gives 700 times more buoyancy per unit volume than the best lipid, squalene; even at a depth of 1000 m, air (or, very nearly, oxygen or nitrogen) will be about 7 times better. And storing air will normally require a lower metabolic investment than storing lipids, whose synthesis is especially costly.

Several problems, though, come with that high volumetric efficiency and low cost. As already mentioned, the solubility of gases in liquids varies with pressure, so a quick reduction in pressure may bring dissolved gases out of solution – the origin of "the bends", the name alluding to the stooped posture of human divers who surface suddenly: they suffer from painful gas bubbles trapped in their joints. We can equilibrate with local pressure at depths up to 100 m or so in a diving bell, caisson, or when using an aqualung. But we then need a slow ascent to allow time for dissolved gas to work its way out through the lungs rather than vaporize within our blood and other tissues. Of the gases in air, nitrogen makes the most trouble, partly because air contains so much of it – 80% by volume – and partly because of its substantial solubility in blood and tissues and its high solubility in body fat. Diving with helium works better since its lower solubility more than compensates for the rapidity of the diffusion of its smaller molecules into the body. Schmidt-Nielsen (1997) gives a particularly good account of the relevant physiology.

Nitrogen dissolution causes less trouble for diving animals than for us, mainly because they do not breath from tanks of air while deep underwater. So only gas already present in their respiratory passages can go into solution. And diving animals usually minimize that volume by exhaling before leaving the surface, tolerating the extra thoracic compression at depth that results. But not all do so. Penguins, conversely, inhale before diving; buoyancy demands that they work hard during the initial phase of descent. At least the buoyancy from air in their plumage and respiratory systems speeds upward gliding during the latter part of their ascents (Sato *et al* 2002).

That thoracic compression and the peculiar dynamics of descent and ascent in penguins brings us to the next problem. Gases compress all too readily, with volumes running almost exactly inverse with local pressure, in sharp contrast with the minor volume changes in other body constituents. Thus at only one depth can an organism containing air at local pressure be neutrally buoyant. Worse yet is the metastability of that neutrality. Ascend, and the gas volume and thus buoyancy increases; descend, and buoyancy decreases. Putting the gas in a rigid container, as within our submarines, would solve the problem, but the necessary stiffness for the container's wall limits that route. Inextensible gas-filled bladders do serve many kinds of surface-living or rooted aquatic organisms, but these need only positive buoyancy above some minimal value. And being functionally inextensible takes only tension-resisting materials, which are relatively cheap to make and light in weight. Withstanding compression well below the surface (as elsewhere) is less easy than resisting tension – columns and beams cost more than ropes.

Consider the poor diving duck. It carries air in its plumage, so just beneath the water's surface it must struggle against excessive buoyancy. Deeper dives cost less to sustain – but reaching greater depth takes more time and energy, both precious resources for an actively swimming air-breather (Lovvorn and Jones 1991). A few birds such as anhingas have hydrophilic plumage and need deal only with internal gases. At the surface, the less buoyant anhinga swims with only neck and head exposed (hence 'snakebird', one common name) and with, one presumes, greater locomotory cost. In addition, it loses most of the insulating value of the plumage, restricting it to warm waters, and immediately after emersion it cannot readily fly (Hennemann 1982).

Most bony fishes maintain near-neutral buoyancy with a gas-filled swimbladder, either gulping air at the surface or (more commonly) secreting gas from circulating blood. Freshwater fishes have swimbladders that make up 5.5 to 8.3% of body volume, while the bladders of those living in the sea occupy 3.1 to 5.6% (Alexander 1966). For comparison, a pair of our lungs (our homolog of a swimbladder, incidentally) averages about 4% of our body volumes – even with their fine volumetric efficiency swimbladders are not tiny organs.

One can demonstrate the problem of maintaining buoyancy by putting a small goldfish in a large glass jar such a 40-liter carboy. Aspirate the air above the water even slightly and the fish rises abruptly and only slowly readjusts to the lower pressure. It may even belch a bubble (which will help readjustment) – goldfish have a connection, the so-called pneumatic duct, between swimbladder and esophagus. A fish that has readjusted to a lower pressure will sink when atmospheric pressure is restored, readjusting again after a short time. Readjustment offsets about a meter of depth (10,000 Pa) per hour (Fänge 1983). If you try the demonstration, do not use just any fish – only some, such as salmon, carp (including goldfish), pickerel, and eels, have pneumatic ducts. Others risk a ruptured swimbladder, which cannot be pleasant.

To put a few numbers on the problem, consider a neutrally-buoyant fish that lives near the surface and whose swimbladder occupies 5% of its volume. If it descends to 10 m, pressure will double and the volume of the swimbladder will halve. As a result it will have a density about 2.5% greater than the surrounding water, enough for a bilaterally compressed body to sink further at an appreciable rate even when maintaining its long axis horizontal. If it descends to 90 m, pressure will go up 10-fold and swimbladder volume down by the same fraction, to about 0.5% of the body. Now 4.5% denser than the water, it will descend still faster. One need not consult a graph to recognize that the problem of depth metastability will be most severe near the surface. Abyssal fish should be able to ignore most depth-dependent volume change.

But how can a fish maintain a gas mixture in a bag at severely elevated pressures? The problem, the inverse of the outgassing of the bends, comes from the same high solubility of gases at high pressures. Blood with haemoglobin can transport a lot of oxygen for its volume, but it does so at a partial pressure no greater than that of the oxygen in the water that passes across the gills. Since the oxygen in deep waters has come either from the air above or from photosynthetic activities near the surface, it will be far below local saturation (partial) pressure. That strongly impels bladder oxygen to dissolve and diffuse into tissues and blood, thence to gills, and thence out to sea.

The solution (in both senses of the word) of fishes has two main components. First, fishes restrict the vascularization of the swimbladder to a tiny gas secreting gland, with thin layers of crystalline guanine that render the rest of the bladder almost completely impermeable to diffusing gas (Lapennas and Schmidt-Nielsen 1977). Second, they supply the gas gland with blood that has passed through a particularly efficient countercurrent exchanger – a device I described previously in connection with heat conservation (Vogel 2005). With such an exchanger, blood leaving the gland with dissolved gases at high partial pressures can lose gas, not to the gills and exterior, but to blood about to enter the gland. In effect, a bag containg oxygen (among other components) at very high partial pressure can be in diffusive contact with blood at nearly the same partial pressure rather than at the lower partial pressures of the ambient water, the gills, or elsewhere in the fish.

Nonetheless, some work does need to be done – a countercurrent exchanger can only minimize losses and secretory costs. The particular trick used by bony fishes consists of acidifying the blood in the gas gland with CO_2 and lactic acid. That reduces the haemoglobin's affinity for oxygen, driving oxygen into physical solution and increasing its partial pressure in the venous blood going from the gas gland into the exchanger. Even though the venous blood has less oxygen per unit volume than the arterial blood –some has passed into the swimbladder – its higher partial pressure means that net diffusion will move oxygen toward the arterial blood. So oxygen will head back toward the swim bladder even as the blood that formerly held it goes gillward. (Bear in mind that solutions, unlike gases, have partial pressures, sometimes called 'tensions', that depend on solubility as well as fractional composition.) This version of an exchanger has been called a "countercurrent multiplier." Schmidt-Nielsen (1997) again gives a succinct description, while Fänge (1983) supplies quantification and the details of the physiological chemistry. Again, the process still requires continuous work – clever machinery can only minimize the task.

7. Using gases at low pressure for buoyancy

In the ways they maintain buoyancy, as in so many respects, the cephalopod mollusks show us evolutionarily achievable alternatives to those of the animals of our own phylum. As noted earlier, squid, like sharks, make do with a minimum of stiff material and with continuous locomotory effort. More remarkable are the cuttlefish. They demonstrate that gas can be kept at pressures both much less than ambient and steady – circumventing both buoyancy loss due to gas compression and the metastability problem. And they do these things without sacrificing the ability to adjust gas volume.

While most bony fish put gas in a single chamber, cuttlefish put it in rigid foam – 'cuttlebone'. With its small and rigid chambers, each about 0.1 mm wide and 0.6 mm high, cuttlebone represents much more material than does a swimbladder. But extracting its main material, calcium carbonate, from saturated seawater should cost little. The compartmentalization allows the cuttlefish to make a nearly incompressible float of lower density, about 620 kg m^{-3}, than that of any lipid, if somewhat denser than the gas plus swimbladder wall of a bony fish.

In a series of now-classic papers, Denton and Gilpin-Brown (1961 *et seq*) and Denton *et al* (1961) worked out how the system operates. Neither gas gland nor countercurrent multiplier plays any role. Cuttlefish balance the hydrostatic pressure difference between the surrounding water and the interior of the cuttlebone with a liquid (within part of it) that has an osmotic pressure below that of the blood. Since the liquid has a lower salt concentration than blood, water is drawn out of the cuttlebone osmotically with the same pressure as it is forced into it hydrostatically. As with a swimbladder, maintaining the system takes work – here the osmotic work of extracting Na^+ and Cl^- to keep the fluid hypoosmotic.

The organ as a whole must (and, of course, does) withstand the local hydrostatic pressure. The measured collapsing pressure, 24 atm, comfortably exceeds the hydrostatic pressure, about 15 atm, at the depths at which the animals live. Pet stores sell pieces of dry cuttlebone for caged birds, who sharpen their beaks on it, so one can easily acquire a sample of this light, rigid, buoyancy tank. Its unusual mechanical properties have attracted attention from people interested in materials (Birchall and Thomas 1983; Gower and Vincent 1996).

We might compare different flotation media by calculating the fraction of body volume that each would require for a standard body density. If cuttlebone, with its density of 620 kg m^{-3}, makes up 9.3% of the volume of a cuttlefish swimming in seawater of 1026 kg m^{-3} (Denton and Gilpin-Brown 1961), then the rest of the body has a density of 1066 kg m^{-3}. Assuming that density for its body (excluding the gas in the swimbladder), a bony fish would need a swimbladder (containing essentially massless gas) of an internal volume of 3.9%, within the reported range. What if a creature used lipid for flotation – say fat or oil of 910 kg m^{-3}? It would have to devote 27% of its volume to this anti-ballast. Even the best non-gaseous flotation material, squalene, would still occupy 20% of overall volume. While these amounts seem plausible (some fishes do contain large amounts of oil and small cetaceans have thick layers of fatty blubber), they amount to an energy investment comparable to that of all other body components combined. Moreover the investment cannot be cashed in during starvation without additional locomotory effort to prevent sinking.

8. Cartesian divers

The depth metastability that bedevils a bony fish and that a cuttlefish evades underlies a wonderfully clever device once well-known to physiologists. This manometric apparatus could measure such things as the rate of oxygen consumption of invisibly small organisms or their parts. Physicists have long recognized a "Cartesian diver" – even if Descartes should not be given credit for it. A 'diver', a floating body of minimally positive buoyancy, can be made to sink by applying a small pressure to a container of water. A version can be assembled from the simplest of everyday items, as in figure 6a, and something like it graces science classes at diverse educational levels.

Because of its fish-like metastability, one cannot easily make a diver that, unattended, neither sinks to the bottom nor floats to the top. While not the usual point of the demonstration, its incarnation as a measuring device depends on that metastability. As originally described by Linderstrøm-Lang (1937) and shown in figure 6b, a glass 'diver' exposed to the local pressure contains a small volume of air (and a respiring bit of life in water) and a droplet of oil as a pressure-transmitting seal. A larger, closed container of liquid (usually ammonium sulphate or lithium chloride to reduce gas exchange) envelops the diver.

Initially the operator holds the diver at some arbitrary depth in the container by manipulating the pressure within the container. If the specimen withdraws gas from the air in

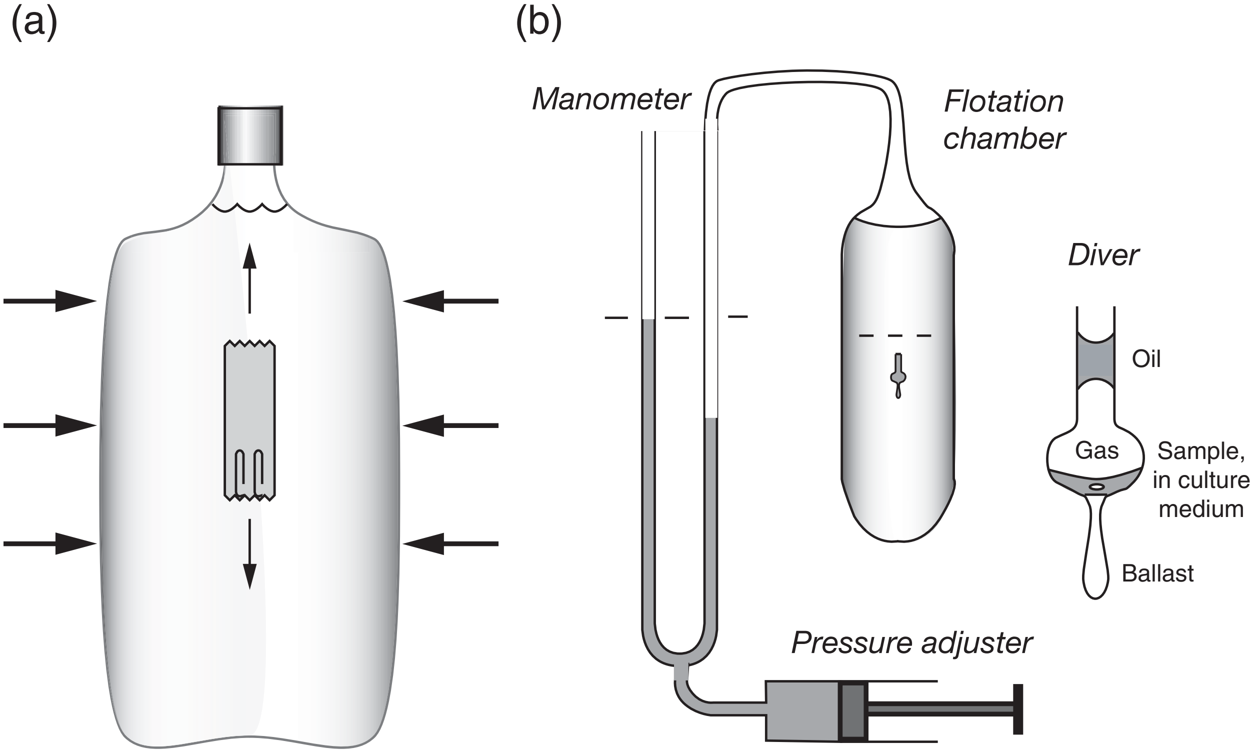

Figure 6. Cartesian divers. (**a**) One made from contemporary artifacts – a foil single-serving package of condiments ballasted just short of sinking with a few paper clips in a water-filled squeezable 2-liter plastic soft-drink bottle. (**b**) A diver used as a micro-respirometer, a simplified version of that described by Linderstrøm-Lang (1937). Today one would probably substitute an electronic manometer for the U-tube.

the diver, the diver becomes denser and plunges. Reducing the overall pressure will persuade the diver to return to the initial depth. Knowing that pressure and the initial volume of the gas within the diver then allows calculation of the volume of gas consumed. With a sensitive manometer, one indicating pressure differences of about 20 Pa, and with a good way to adjust pressure differences, a volume change of 2,000,000 μm^3 (2 x 10^{-6} cm^3) can be measured. Zeuthen (1943) managed to increase its sensitivity to 2,000 μm^3 and with it measured the oxygen consumption of single amoebae.

Some years later, Scholander *et al* (1952) improved the technique further, adding a reference diver for nulling the pressure and isolating the sample in a bubble whose volume could be measured with an ocular. This more complex version had yet greater sensitivity, about 200 μm^3 – the volume of a sphere about 7 μm across or of a typical animal cell – good enough to measure the metabolic rate of still smaller cells. Per Scholander, of course, was the great master of manometry, both macro and micro – recall his device for measuring the extreme negative pressures in the vessels of plants, described in the previous essay (Vogel 2006). These divers have not so much been superceded as shelved due to waning interest in what they could measure.

9. Gravity versus evolution

This extended consideration of the biology of gravity – three essays in all, with numerous facets left untouched – prompts one final note. Once established, we put theories or laws or definitive equations to two fairly distinct uses. Sometimes we ask that they explain phenomena in the world around us; sometimes we ask that they predict some future state of that world or the outcome of some deliberate manipulation. To explain is not necessarily to predict, and prediction need not depend on an intuitively satisfying explanation. Some theories do better at one task, some at the other.

Gravity, expressed as Newton's universal equation, does a splendid job of prediction. Our contemporary technology, especially in its larger manifestations, would be unthinkable without its reliable precision. But as an explanation, I think it serves us poorly. Taken at face value, it requires that every bit of matter in the universe have some sense of the existence of every other bit of matter. Modern physics (so I am reliably informed) does not rely on such a metaphysical assertion, but its alternative explanation lacks intuitively satisfying persuasiveness.

Evolution, as defined by the concept of natural selection, has quite the opposite virtue. It does a fine job of explaining both the large- and the small-scale phenomena of life, including many subtle and even counterintuitive observations. It does make predictions, but their precision never approaches that of Newton's simple and succinct equation. The real phenomena whose futures we want to predict involve too many players, too much contingency, too much amplification of insignificant perturbations. This contrast plays a role both in determining the relative status of biology among the sciences and in fueling the criticism of evolution from some quarters outside the sciences.

Acknowledgements

This essay has grown in large part out of suggestions and information from Chris Barnhart, John Havel, Dan Livingstone, Fred Nijhout, Howard Reisner, Knut Schmidt-Nielsen, and Rick Searles.

References

Alexander R M 1966 Physical aspects of swimbladder function; *Biol. Rev.* **41** 141–176

Barthlott W and Neinhuis C 1997 Purity of the sacred lotus, or escape from contamination in biological surfaces; *Planta* **202** 1–8

Birchall J D and Thomas N L 1983 On the architecture and function of cuttlefish bone; *J. Mater. Sci.* **18** 2081–2086

Brown H P 1981 Key to the world genera of Larinae (Coleoptera, Dryopoidea, Elmidae), with descriptions of new genera from Hispaniola, Colombia, Australia, and New Guinea; *Pan-Pacific Entomol.* **57** 76–104

Brown H P 1987 Biology of riffle beetles; *Annu. Rev. Entomol.* **23** 253–273

Comstock J H 1887 Note on the respiration of aquatic bugs; *Am. Nat.* **21** 577–578

Denton E J and Gilpin-Brown J B 1961 The buoyancy of the cuttlefish, *Sepia officinalis* (L.); *J. Mar. Biol. Assoc. UK* **41** 319–342

Denton E J, Gilpin-Brown J B and Howarth J V 1961 The osmotic mechanism of the cuttlebone; *J. Mar. Biol. Assoc. UK* **41** 351–364

DeVries A L and Eastman J T 1978 Lipid sacs as a buoyancy adaptation in an Antarctic fish; *Nature(London)* **271** 352–353

Eastman J T and DeVries A L 1982 Buoyancy studies of notothenioid fishes in McMurdo Sound, Antarctica; *Copeia* **1982** 385–393

Ellers O 1995 Form and motion of *Donax variabilis* in flow; *Biol. Bull.* **189** 138–147

Fänge R 1983 Gas exchange in fish swim bladder; *Rev. Physiol. Biochem. Pharmacol.* **97** 111–158

Feng L, Li S, Li Y, Li H, Zhang L, Zhai J, Song Y, Liu B, Jiang L and Zhu D 2002 Super-hydrophobic surfaces: from natural to artificial; *Adv. Mat.* **14** 1857–1860

Fish F E and Stein B R 1991 Functional correlates of differences in bone density among terrestrial and aquatic genera in the family Mustelidae (Mammalia); *Zoomorphology* **110** 339–345

Gao X and Jiang L 2004 Water-repellent legs of water striders; *Nature (London)* **432** 36

Gower D and Vincent J F V 1996 The mechanical design of the cuttlebone and its bathymetric implications; *Biomimetics* **4** 37–57

Harpster H T 1941 An investigation of the gaseous plastron as a respiratory mechanism in *Helichus striatus* Leconte (Dryopidae); *Trans. Am. Microsc. Soc.* **60** 329–358

Harpster H T 1944 The gaseous plastron as a respiratory mechanism in *Stenelmis quadrimaculata* Horn (Dryopidae); *Trans. Am. Microsc. Soc.* **63** 1–26

Hebets E H and Chapman R F 2000 Surviving the flood: plastron respiration in the non-tracheate arthropod *Phrynus marginemaculatus* (Amblypygi: Arachnida); *J. Insect Physiol.* **46** 13–19

Hennemann W W III 1982 Energetics and spread-winged behavior of anhingas in Florida; *Condor* **84** 91–92

Hills G J 1972 The physics and chemistry of high pressures; *Symp. Soc. Exp. Biol.* **26** 1–26

Hinton H E 1976 Plastron respiration in bugs and beetles; *J. Insect Physiol.* **22** 1529–1550

Hora S L 1930 Ecology, bionomics and evolution of the torrential fauna, with special reference to the organs of attachment; *Philos. Trans. R. Soc.* **B218** 171–182

Krogh A 1941 *The comparative physiology of respiratory mechanisms* (Philadelphia: University of Pennsylvania Press)

Lapennas G N and Schmidt-Nielsen K 1977 Swimbladder permeability to oxygen; *J. Exp. Biol.* **67** 175–196

Linderstrøm-Lang K 1937 Principle of the Cartesian diver applied to gasometric technique; *Nature (London)* **140** 1080

Lloyd J T 1914 Lepidopterous larvae from rapid streams; *J. N.Y. Entomol. Soc.* **22** 145–152

Lovvorn J R and Jones D R 1991 Effects of body size, body fat, and changes in pressure with depth on buoyancy and cost of diving in ducks (*Aythya* spp.); *Can. J. Zool.* **69** 2879–2887

Neinhuis C and Barthlott W 1997 Characterization and distribution of water-repellent, self-cleaning plant surfaces; *Ann. Bot.* **79** 667–677

Nielsen A 1950 The torrential invertebrate fauna; *Oikos* **2** 176–196

Pennak R W 1978 *Fresh-water invertebrates of the United States* 2nd edition (New York: John Wiley)

Pommen G D W and Craig D A 1995 Flow patterns around gills of pupal net-winged midges (Diptera: Blephariceeridae): possible implications for respiration; *Can. J. Zool.* **73** 373–382

Resh V H and Jamieson W 1988 Parasitism of the aquatic moth *Petrophila confusalis* (Lepidoptera: Pyralidae) by the aquatic wasp *Tanychela pilosa* (Hymenoptera: Ichneumonidae); *Entomol. News* **99** 185–188

Sato K, Naito Y, Kato A, Niizuma Y, Watanuki Y, Charrassin J B, Bost C-A, Handrich Y and LeMaho Y 2002 Buoyancy and maximal diving depth in penguins: do they control inhaling air volume?; *J. Exp. Biol.* **205** 1189–1197

Schmidt-Nielsen K 1997 *Animal physiology: adaptation and environment* 5th edition (Cambridge: Cambridge University Press)

Scholander P F, Claff C L and Sveinsson S L 1952 Respiratory studies of single cells. I. Methods; *Biol. Bull.* **102** 157–177

Schuetz D and Taborsky M 2003 Adaptations to an aquatic life may be responsible for the reversed sexual size dimorphism in the water spider, *Argyroneta aquatica*; *Ecol. Evol. Res.* **5** 105–117

Stride G O 1955 On the respiration of an aquatic African beetle, *Potamodytes tuberosus* Hinton; *Ann. Entomol. Soc. Am.* **48** 344–351

Suzuki K and Taniguchi Y 1972 Effect of pressure on biopolymers and model systems; *Symp. Soc. Exp. Biol.* **26** 103–124

Taylor M A 1993 Stomach stones for feeding or buoyancy? The occurrence and function of gastroliths in marine tetrapods; *Philos. Trans. R. Soc. London* **B341** 163–175

Thorpe W H 1950 Plastron respiration in aquatic insects; *Biol. Rev.* **25** 344–390

Thorpe W H and Crisp D J 1947 Studies on plastron respiration. I. The biology of *Aphelocheirus* [Hemiptera, Aphelocheiridae (Naucoridae)] and the mechanism of plastron retention; *J. Exp. Biol.* **24** 227–269

Vogel S 1994 *Life in moving fluids,* 2nd edition (Princeton NJ: Princeton University Press)

Vogel S 2004 Living in a physical world. I. Two ways to move material; *J. Biosci.* **29** 392–397

Vogel S 2005 Living in a physical world. V. Maintaining temperature; *J. Biosci.* **30** 581–590

Vogel S 2006 Living in a physical world. VII. Gravity and life on the ground; *J. Biosci.* **31** 201–214

Wagner T, Neinhuis C and Barthlott W 1996 Wettability and contaminability of insect wings as a function of their surface sculpture; *Acta Zool.* **77** 213–225

Wainwright S A, Biggs W D, Currey J D and Gosline J M 1976 *Mechanical design in organisms* (New York: John Wiley)

Webber D M, Aitkin J P and O'Dor R K 2000 Costs of locomotion and vertical dynamics of cephalopods and fish; *Physiol. Biochem. Zool.* **73** 651

Wilga C D and Lauder G V 2002 Function of the heterocercal tail in sharks: quantitative wake dynamics during steady horizontal swimming and vertical maneuvering; *J. Exp. Biol.* **205** 2365–2374

Zeuthen E 1943 A Cartesian diver micro respirometer with a gas volume of 0.1 μl; *C. R. Trav. Lab. Carlsberg, ser. chim.* **24** 479–518

*e*Publication: 31 July 2006

Living in a physical world
X.
Pumping fluids through conduits

Steven Vogel
Department of Biology, Duke University, Durham, NC 27708-0338, USA
(Fax, 919-660-7293; Email, svogel@duke.edu)

1. Introduction

The first of these essays argued that, because diffusion is ineffective over all but minute distances, an organism larger than a typical cell must move fluid to move material (Vogel 2004). By whatever name, internal bulk fluid movement absorbs energy, if nothing else as a consequence of that universal fluid property, viscosity. Supplying that energy requires provision of some manner of pump. Such a pump may also accelerate fluid or lift it against gravity, but neither of these rise to quite the same level of inevitabi-lity. Not that pumping must incur metabolic cost – sometimes an external agency can be co-opted to do the requisite work.

The diversity of circumstances under which organisms pump fluids, the phylogenetic diversity of the organisms, and the structural diversity of the pumps themselves all militate against treating biological pumps as a single class of functional devices. Separate books, or at least separate chapters, deal with the ascent of sap in a tree, the suction of blood by a mosquito, and the suspension feeding of a clam. Here I want to explore generalizations that might emerge from considering pumps in all of that functional, phylogenetic and structural diversity.

As well, pumps have been ubiquitous components of human technology since the first fields were irrigated with water that gravity alone could not supply, that is, since water was first hoisted from lake, river, or well. A remarkably wide range of simply-constructed yet effective devices remain in use where industrial products have yet to reach agrarian cultures. Among our machines, only electric motors may exceed them in range of sizes, applications, and designs. Engineers, designing pumps for diverse applications, have worried about variation in operating conditions, efficiency, and other parameters. In analysing and classifying nature's pumps, we biologists thus come late to the game, and we ought to take full advantage of all that earlier attention.

[In part, this essay extends the general reasoning of a predecessor (Vogel 1995), one that a reviewer of its manuscript rightly noted did not turn out quite as satisfactorily as the author had hoped. I will, of course, have to reiterate some of the points of that paper.]

2. The relevant variables

An insect, most famously an ant, can lift many times its own body weight – but it cannot lift it far. Similarly, a tree can draw sap upwards with pressures of tens of atmospheres, millions of pascals – but it does so very slowly. We may be overly impressed by the spectacularly high forces and pressures that organisms can produce and insufficiently mindful of constraints on distances and volume flows. Processes such as lifting weights or forcing fluids through pipes involve three variables, and these may operate in a wide range of combinations. Force, distance, and power define a lifting task, with power just the product of force and rate of change of distance. Similarly, pressure, volume flow, and power define the task of a pump, with the last again the product of the first two. A bivalve mollusk can pump its own volume of water in, across its gills, and out again every few seconds, but it does so against only a few pascals of pressure – a volume flow no less impressive than the pressure generated by a sap-lifting tree.

A particular graphic representation of pump performance has enjoyed long usage in engineering and is shown, without specific data, in figure 1. With only a slight modification (incorporated here, as will be noted shortly) it should work as well for biological pumps, even if we usually cannot

Keywords. Flow; pressure; pumps

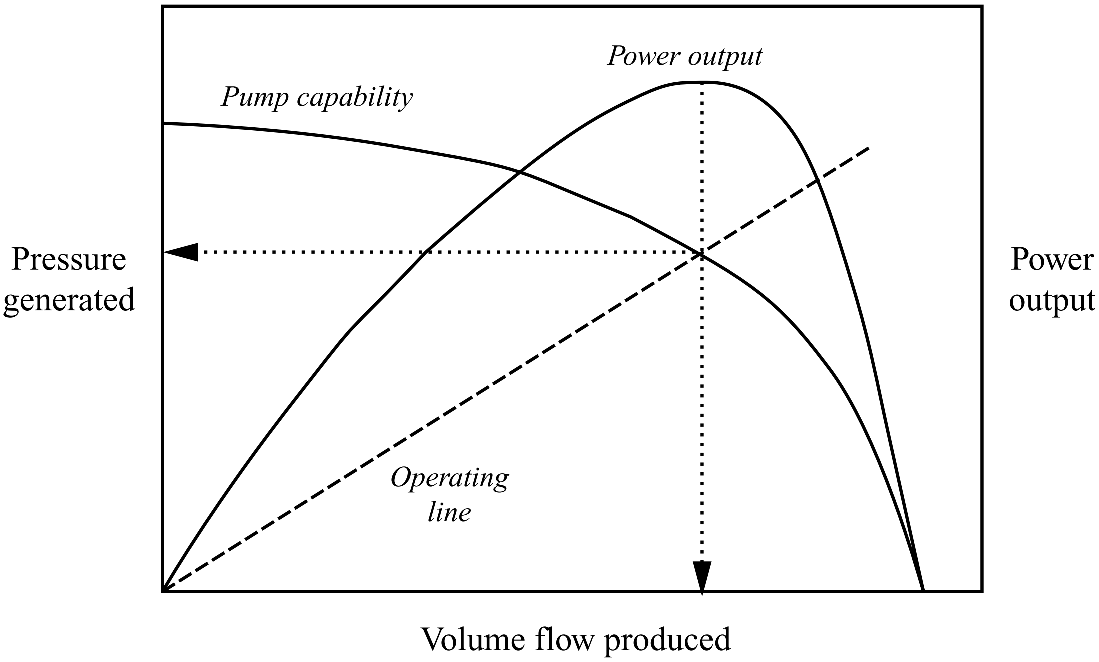

Figure 1. A pump performance graph. Most pumps yield plots similar to this one; they differ mainly in the scaling of the axes. The axis-to-axis outer curve ("pump capability") marks the limits of the pump under any operating condition. The dashed line corresponds to the operating condition set by the particular load imposed on this pump, while the dotted lines give the pressure and volume flow maxima for that particular load. This example shows an optimally efficient combination of pump and application–the intersection of operating and pump capability lines lies just beneath the point of maximum power output.

measure pump performance quite as far from normal operating conditions and must extrapolate to get the end points of their basic operating lines.

The maximum pressures most pumps can produce occur under conditions of zero volume flow, while their maximal volume flows happen when the opposing pressure or pressure they produce is zero. So a curved line from one axis to the other, the "pump capability" line, along with the axes themselves, defines a potential area of operation – possible combinations of pressure (Δp) and volume flow (Q). In a actual application, operating conditions are constrained by the resistance of the load, defined by how much volume flow corresponded to each value of pumping pressure. For most technological pumps, the pressure needed varies with the square of the volume flow it has to produce. Thus the "operating line" forms a parabolic curve extending upward from the origin. For virtually all biological pumps, Δp will vary almost directly with Q, so the line from the origin will be straight rather than parabolic. In either case, the intersection of this operating line with the pump capability line marks the maximum output of the particular pump in a particular application.

Size underlies that difference between technological and biological pumps. For the relatively small sizes and low speeds of the latter, most flows will be laminar. Thus the Hagen-Poiseuille equation or something analogous applies; it defines a resistance ($\Delta p/Q$) that is ohmically constant over the range of pressures and flows.

The power (P) a pump puts out equals the product of its volume flow and the pressure increase of the fluid passing through it – $Q\Delta p$. In practice, the graph indicates power on a second ordinate. Its curve begins and ends at the abscissa since zero values of either pressure increase or volume flow mean zero power output.

For an ideal match of pump to task, the peak of the power output curve should lie just above (or below, since the ordinates have different scales) the point of intersection of the operating line and the pump capability line. The graph thus gives a valuable view of that coupling. If peak power output occurs well to the left of the intersection, the pump is one designed more as a pressure producer and less as a volume impeller than would be best. Its peak power output will not be reached, wherever on the operating line it works. If, conversely, peak power output occurs to the right of that intersection, the pump produces too much volume flow and too little pressure; again its maximum power output will not be realized. This second mismatch can have a particularly serious consequence if, as when lifting from a well or from ground to top of tree, producing any useful output demands some minimum (here gravitationally-determined) pressure – that is, a pressure threshold must be exceeded to get any flow.

In short, specifying the power output of a pump, even power output plus energy conversion efficiency, may not indicate whether a particular pump will give proper service when harnessed to a particular task. Pumps vary widely in the mixes of pressures and volume flows they can produce – in the specific shapes and positions of their operating lines. While all graphs of the kind shown in figure 1 may look similar, the scales on their axes will be anything but.

3. A functional classification of pumps

The literature on pumps for technological uses (for instance, Karassik *et al* 2000) recognizes two general categories, with diverse implementations of each. Only a few devices fail to fit comfortably into one class or the other. (The specific names of each, though, differ somewhat from source to source, leading to occasional awkwardness for on-line searching or using indices.) A few words about the devices in each category might stimulate recognition of biological equivalents beyond ones long obvious.

Those in one category are most often called "displacement" or "positive displacement" pumps. In many of these, fluid is drawn into a chamber and then persuaded (as by reducing the chamber's size) to leave by a different route. Most familiar are ones with pistons that move back and forth in chambers, plus valves to ensure unidirectional, if pulsatile, flow – for instance the ones with which we hand-inflate pneumatic tires. A less common version, diaphragm pumps, change chamber volume with periodic pushes against a flexible element that forms one of the chamber's walls; it exchanges the problem of a closely fitted piston for that of a non-rigid element. Other displacement pumps work by translocating the functional chamber itself. In the commonest versions – gear, screw, vane and lobe pumps – multiple moving components carry fluid along as they themselves move. In another displacement pump, the so-called air-lift (or gas-lift) pump, bubbles of gas rising through a narrow vertical tube of liquid carry liquid upward in the moving chambers of liquid formed between each pair of bubbles. Similar to these latter two types, and of especial biological relevance, is the peristaltic pump, with its traveling constrictions of flexible tubes. It eliminates contact between fluid and pump housing and tolerates flows of variable viscosity and fluids with suspended solids. But the technological versions perform inefficiently and are not particularly reliable, so they have remained uncommon. The flows produced by displacement pumps range from nearly steady to severely pulsatile unless paired with some external buffer.

Those in the other category are called "dynamic," "fluid dynamic" or "rotodynamic" pumps, with this last name recognizing their ordinarily rotational operation. All depend on fluid dynamics rather than fluid statics. The commonest types drive fluids with axial or centrifugal fans; most familiar are ones driving air with either propellers or squirrel-cage blowers. Another type is the jet pump, in which one liquid is squirted into a channel or duct of another liquid through a jetting orifice or "eductor." While typically lower in efficiency than are rotary pumps, jet pumps need no moving solid parts. The vacuum pumps that we attach to tap water outlets are the most familiar examples. Related to jet pumps are other devices in which one flow induces another – old-fashioned carburetors drew in gasoline this way, and a variety of buildings, old and new, achieve ventilation by using ambient wind to draw air through themselves.

The present focus on the distinction between the two categories grows out of one particular generalization. Displacement pumps work best at higher operating pressures than do dynamic pumps – and, of course, vice versa. No sharp value of pressure marks the transition, though, with the simplicity of displacement pumps making them preferable for some low-pressure applications and the smooth operation of dynamic pumps lying behind their use, often with multiple stages, to produce fairly high pressures. Precisely this same distinction applies to the pumps found in organisms. Despite the imperfect dichotomization, the mechanical and operational distinctions between the categories help us understand the particular distribution of pump types we observe in nature. Thus the distinction provides an analytic and functional categorization nicely complementary to our traditional phylogenetic viewpoint.

We will begin with a brief look at the way past and present agrarian societies pump water, taking their simple devices as illustrative of pump types, before dealing with the greater complexity and diversity of biological pumps. In engineering textbooks, fluid statics precedes fluid dynamics because of its (at least superficially) relative simplicity. Similarly, displacement pumps have preceded dynamic pumps in adoption by humans, and they remain more common among less industrialized technologies. Still, they have long been a diverse lot. One gets a good view of the different types in recent and present use from the collection of pictures collected by Thorkild Schiφler and posted on the website of the Experimentarium, of Hellerup, Denmark (*www.experimentarium.dk/uk/naturvidenskab_og_teknik/schiolers/*); additional material can be downloaded from *www.timsmills.info/URL-S/Animal%20Powered%20Systems.pdf*.

Devices immediately recognizable as displacement pumps range from simple buckets and scoops to more complex pot chains, dragon-bone chains, swinging canoes (dhoons), shadufs, saquiyas (or saqias, etc.), and hoists (delous). One that, unusual for displacement pumps, avoids pulsatile or intermittent operation is the Archimedean screw (figure 2), a helical screw either turning in a tube or fixed within a cylindrical housing that turns. A Roman fresco at

Pompeii shows an Archimedean screw pump operated as a treadmill by a person on the outside; in modern versions the operator typically turns a crank at one end. (Cranks, at least in Europe, are medieval and later). The shallow incline of the tube allows pockets of water to form; turning then raises the pockets. They still find occasional and even large-scale use at air-water interfaces – if fully submerged, pockets do not form, so they then become inefficient viscosity-dependent dynamic pumps.

The only fluid dynamic pump apparently at all common in preindustrial societies is the noria (figure 2), and it relies on displacement as well as dynamics. A flowing stream turns an undershot waterwheel (the dynamic part); water-holders attached to the periphery of the same wheel fill from the same stream and raise water to an elevated spillway (the displacement part). As we will see – and one reason to start with old human technology – nature seems to face the same (or at least an analogous) difficulty in devising

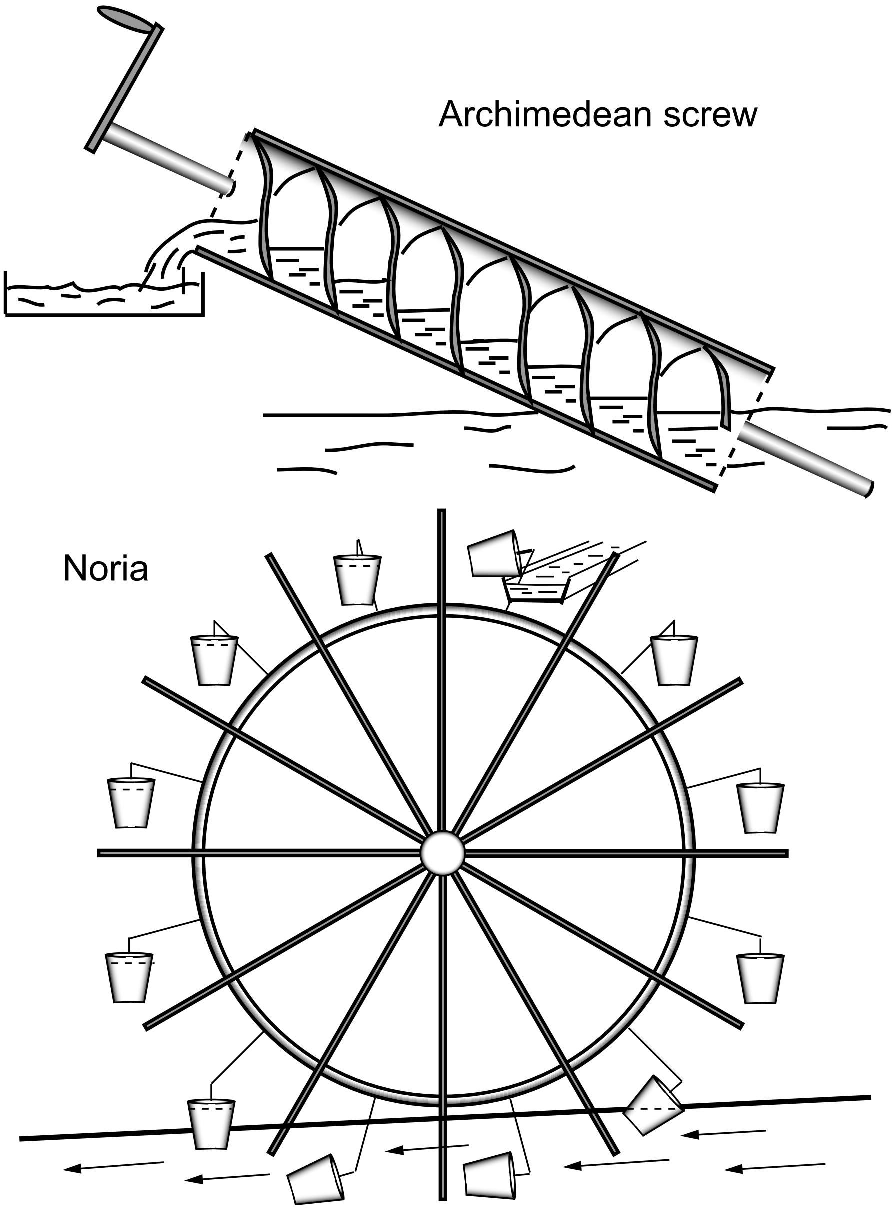

Figure 2. Two ancient devices for lifting water, an Archimedean screw and a noria; both are at least in part dynamic pumps. We still use a large-scale, modern version of the Archimedean screw for some low-lift, high volume applications. But the noria has become anachronistic, most directly replaced by the hydraulic ram, which became practical in the 19th century with the advent of inexpensive metallic components of decent precision.

dynamic pumps. Indeed, judging from our technological versions, nature may face an additional obstacle. All but a few dynamic pumps (such as jet pumps) employ rotating elements – hence the common name "rotodynamic." Such continuously rotating wheel-and-axle devices, of course, do not occur in eucaryotic organisms.

In looking at the pumps organisms employ, we will limit our purview to those that move liquids, that is, water plus aqueous solutions and suspensions. As a less obvious circumscription, we will for the most part exclude devices where the objective is movement of pumper rather than pumped – despite the artificiality of a line between, say, paddle-based locomotion and paddle pumps.

4. Living displacement pumps

The displacement pumps of organisms range from ones with close technological analogs to others that, although not fundamentally novel, have limited appeal for humans – either we have more attractive alternatives or they work in ways awkward for our materials, machines, and applications.

(i) *Valve-and-chamber pumps*: A single chamber whose volume can be changed, together with a pair of valves, satisfies the minimal requirements for such a pump. Our hearts, paradigmatic examples, have four chambers and six valves and operate as a pair of pumps, each with a two-stage pressure booster. Additional valve-and-chamber pumps return both blood from the veins of our legs and lymph from our tissues to our hearts – routine contractions of our skeletal muscles squeeze the chambers, these no more than the lengths of vessel between adjacent valves.

Valve-and-chamber hearts are widespread among the metazoa, particularly among vertebrates and mollusks. The requisite machinery demands only ordinary additions – just valves and muscle – to a flexible tubular element. Even single chambers can produce pressures of over 20,000 Pa, as do the left ventricles of tall mammals. The lymph hearts of fishes, amphibians and reptiles consist of contractile vessels with valves (Prosser 1973, Ottaviani and Tazzi 1977), and so do the lateral hearts of the giant earthworm, *Glossoscolex* (Johansen and Martin 1965). Some insect hearts may also operate as valve-and-chamber hearts (Jones 1977). Nor do all such pumps occur in circulatory systems. Valves on the inputs to the mantle chambers of jet-propelled cephalopods together with self-valving output funnels amount to the same kind of chamber-plus-paired-valves. Jetting scallops use their mantles as valves in an analogous manner. Similarly, the mouth, oral and branchial chambers, and opercula of fishes that pump water over their gills operate in this fashion (Lauder 1980).

(ii) *Valveless chamber and piston pumps:* Where pumps need produce only single pulses of fluid or reciprocating flows, valves become superfluous. That happens in many systems and appears in so many guises that the underlying commonality can easily escape notice. Our urethral pumps (see Glemain *et al* 1990) work this way. Most jet propulsors, from those of jellyfish (DeMont and Gosline 1988) to the anal jets of dragonfly nymphs, are valveless chamber pumps. Anal jets can produce both single jet pulses (Hughes 1958) and repetitive respiratory flows (Pickard and Mill 1974), with the same equipment serving both functions. Similarly, most injectors make use of valveless chambers, including the venom injectors and squirters of rattlesnakes and cobras (Kardong and Lavin-Murcio 1993, Young *et al* 2003, 2004), of the toxic snail, *Conus*, (Schulz *et al* 2004), and of spiders (Yigit *et al* 2004). So does the branchial chamber of jetting fishes, another system that can alternatively serve for respiration (Brainerd *et al* 1997).

And valveless chambers underlie most suckers, including both blood- and nectar-sucking insects (Kingsolver and Daniel 1995). Even aphids, which can use the considerable hydrostatic pressures of plant phloem to drive fluids in through their stylets, retain the capacity to generate suction in this way (Kingsolver and Daniel 1995). Most often sucking chambers depend on expansion tied to the elastic recoil of some muscularly-stressed material. The pressures sucking insects produce can be well below ambient, even, in the bug *Rhodnius*, subzero (Bennet-Clark 1963).

Less common than valveless chambers are piston pumps, perhaps because the physical arrangement represents something unusual in nature – however ordinary it might be in human technology. Several kinds of infaunal marine worms have been described as irrigating their burrows by acting as piston pumps, in particular the clam worm *Nereis,* and the parchment worm *Chaetopterus*. In both cases that may oversimplify the well-coordinated movements of appendages as well as body walls (Riisgård and Larsen 1995).

(iii) *Valveless moving chamber (peristaltic) pumps:* These typically produce pressures lower than the preceding two types, and they are likely to be lower in energetic efficiency; like them they should be easy to evolve from a basic muscle-enclosed tube. Our intestines and our esophagi, of course, depend on peristalsis, as do the hearts of most annelids, holothurians (sea cucumbers) and arthropods (Martin 1974). Burrow irrigation in *Nereis*, noted above, seems to involve some peristaltic body wall movement as well as piston action. Inasmuch as earthworms locomote in an essentially peristaltic mode, we might expect similar peristaltic pumping among burrow-dwelling aquatic oligochaete as well as polychaete annelids.

Besides their undoubted ease of evolution, peristaltic pumps have functional advantages. Pumping liquids of high viscosity or with a lot of suspended solids presents no great problem. Peristaltic action provides mixing as well as lateral transport, offsetting the laminarity of low-speed flows in small pipes. That should be of some consequence (as in intestines) where absorption or exchange across pipe walls accompanies lengthwise transport – if not in, say, ureters. And pumping direction can be reversed with nothing more than a minor shift in neuromuscular coordination. Both intestines and esophagi can go either way. Cud-chewing bovids routinely reverse their esophageal pumping, and insect hearts often switch directions (Jones 1977).

(iv) *Osmotic pumps:* These uniquely (but not necessarily) biological devices operate not by decreasing the size of a compartment but by increasing the volume of what the compartment must contain. Organisms rarely engage in active transport of water; instead they move ions or small molecules, with water following by passive osmosis. Thus the driving force for osmotic pumps comes indirectly from such transport or else from local increase in osmotic strength as a result of dimer, oligomer, or polymer hydrolysis.

Most osmotic pumps are small; indeed these are the predominant pumps that drive bulk fluid flow in unicellular systems. They may move low volumes, but they can develop high pressures, fortuitously complementary with the peculiar ability of small systems to resist great pressure differences. A mere molar difference in solute concentration (assuming a non-electrolyte) across a membrane produces about 2.2 MPa of pressure. To cite a specific case, the fungus *Gibberella* (Trail *et al* 2005), about which more below, produces a peak pressure of about 1.5 MPa. Nonetheless, these pumps suffer from several drawbacks. For one thing, they act at surfaces, so scaling up three-dimensionally takes a disproportionate increase in surface area, either with folds or villi or by proliferation of the basic units. For another, pumping ordinarily consumes the osmolyte, not as fuel, but by simple dilution – a water-pump that works by hydrolyzing starch into osmotically active mono- or disaccharide will find that those products have been diluted and then carried away in the flow it produces. So resynthesis may require more than mere metabolic reversal. While countercurrent or other such devices may help, the basic problem cannot be entirely evaded.

Osmotic pumps figure in at least two of the schemes for throwing fungal projectiles that were described in the third essay (Vogel 2005), those of *Pilobolus* and *Gibberella*. In both, osmotic engines power hydraulic ejection both by providing hyphae with liquid and by stretching their elastic walls; *Gibberella* does so by transporting potassium, with chloride coming along as counterion. Another osmotically charged hydraulic engine closes the Venus flytrap (Forterre *et al* 2005). One can point as well to the excretory organs of animals, varying from partially osmotic to fully osmotic ones such as the aglomerular kidneys of some marine teleost fishes. (Our own kidneys capitalize on arterial blood pressure and thus on our hearts to drive their initial glomerular ultrafilter.)

The protonephridia of some acoelomate invertebrates, the best known being the flame cells of planaria, are a peculiar case – or perhaps one should say possible case. Ducts opening to the exterior remove excess water, as required by these fresh-water creatures. Presumably water transport follows some osmolyte secretion that gets reabsorbed. One wonders about the role of the one cilium (in solenocytes) or the tuft of cilia (in flame cells) at the blind ends of these ducts (Schmidt-Nielsen 1997). Osmotic pumps work at high pressures and low flow rates, while ciliary pumps, as fluid dynamic devices, do best when called on for the opposite service. I have seen no suggestion about what good – perhaps a bit of stirring – one or a few cilia can do under such circumstances. Our own renal tubules may have cilia, but no analogy can be drawn. Ours lack central microtubules and cannot propel fluid; instead they appear to work in the opposite mode, as generators, more specifically as flow sensors (Yokoyama 2004).

Osmotic pumps play major parts in two large-scale fluid transport systems, although in both instances the details of their mechanisms have yet to be fully elucidated. If you cut the top off a well-watered herbaceous plant, sap oozes out from specific places, the xylem elements, on the cut surface. Water is absorbed from the soil and then pumped up the stem by so-called root pressure. Herbaceous stems may provide the obvious expression of the phenomenon, but it occurs in some large, woody plants as well. Pickard (2003a, b) provides a good view of the present knowns and unknowns surrounding root pressure.

While flow in xylem depends mainly on evaporative pull from the top, osmotic pumping seems mainly responsible for driving flow in the complementary tissue, phloem. Again, the details have given trouble. The classic Munch hypothesis from the 1930s invokes osmotic forces, and that they play some role has not been contentious. But once again the details still confuse us; here one encounters a daunting diversity of structures, flow pathways, and chemistry. A look at, for instance, van Bel (1993) or Thompson and Holbrook (2003) will give some sense of the problems involved.

An osmotic pump should be designed to get as much passive water movement for a given amount of osmolyte transport as possible. That underlies a feature common among such pumps. Instead of secreting osmolyte into some large external (or extracellular) spaces, they discharge it into restricted areas, isolated to some extent from those larger volumes. Thus its concentration is (and, for a time remains) higher. Depending on the system, osmolyte may be ultimately lost downstream or actively reabsorbed for reuse. The loops of Henle of mammalian kidneys play a particularly fancy version of this game, with a countercurrent multiplier isolating a region of high osmolarity. The basic arrangement seems to have been first recognized by Curran (1960) in rat intestines and by Curran and MacIntosh (1962) as a general phenomenon. It was later demonstrated in the water-ejecting invaginations of gall bladder cells by Diamond and Bossert (1967), who gave it the nicely descriptive name, "standing gradient osmotic flow." As noted by Tyree and Zimmermann (2002) (and at least tacitly by Nobel 2005), in higher plants such standing gradient osmotic devices serve crucial roles in generating root pressure and in loading and propelling flow in phloem.

(v) *Evaporative pumps:* In one sense evaporative pumps work in the opposite manner from osmotic pumps. Instead of generating positive pressure by transport of osmolytes and wate into a compartment, they generate negative pressure by removing liquid from a compartment. Like osmotic pumping, evaporative pumping requires no macroscopic moving machinery, preadapting it for use by plants. Osmotic pumping depends on differentially permeable membranes, biologically ubiquitous; evaporative pumping requires an air-water interface, limiting its applications to terrestrial or semiterrestrial organisms. So evaporative pumps should be less widely distributed. They should be limited, as well, by a peculiar asymmetry between positive and negative pressures. At least in physical models, pressures can be increased without intrinsic limit, while pressures cannot easily be decreased much below zero. Thus one might expect pressure drops to be limited to whatever amounts offset ambient pressures. Another possible constraint is that evaporation must occur across a surface that can, at the same time, tolerate the pressure difference that the pump generates.

Despite both evident and possible limitations, evaporative pumping probably moves more liquid through organisms than do all other macroscopic pumps combined. It does, one might say, the heavy lifting in drawing water from soil and raising it to the photosynthetic structures of terrestrial plants, which in some sense pay with such water loss to obtain rare and precious atmospheric CO_2. These plants manage to evade the zero pressure limit, not trivially but monumentally, generating tensions in water as low as – 120 atmospheres (12 MPa), far below zero. They manage to create interfaces that withstand such pressure differences without either collapsing or restricting evaporation; they do so by taking advantage of the fine-scale cellulose meshwork of their cell walls and the high surface tension of water in contact with air. The seventh essay (Vogel 2006) described this remarkable scheme; let me just note here that the evaporative pumps of terrestrial plants generate the most extreme pressures of any biological pumps, and that such pumps are rare elsewhere in either natural or human technology.

5. Nature's dynamic pumps

For these perhaps we should retain the additionally qualified name one sometimes sees, *fluid dynamic pumps*, since another group of biological pumps depends on the dynamics of solid materials. Compared to the analogous devices in our technology, the fluid dynamic pumps of organisms appear both less diverse and more distant in appearance. That greater distance emerges from two basic differences between the two technologies – not only nature's inability to make macroscopic rotational machinery, but also our lack of anything much like cilia or ciliated and thus wall-pumping tubes.

(i) *Drag-based paddles:* In our quest for efficient propulsion, propellers, which move blades normal to flow, have largely replaced paddles, which move them parallel to flow. One must go back to the noria to find a fluid dynamic pump based on the drag of broad blades in flow. Similarly, nature makes only limited use of pumps based on paddling. Foster-Smith (1978) recognized such a pump in the amphipod crustacean, *Corophium volutator*, which burrows in mud and propels water by beating its pleopods. I suspect that members of the infaunal shrimp genera *Upogebia* and *Callianassa* do likewise. But they could do so only occasionally, since they normally live in blind pockets branching off their U-shaped burrows, and since they seem to depend on flow induced by asymmetry of the burrow apertures. Foster-Smith found that *Corophium* could achieve pressures only about 4% as high as those made by the piston pumpers *Nereis* and *Arenicola*, although for its size it could drive considerably greater volume flows.

(ii) *Lift-based propellers:* My search for liquid-propelling pumps, *sensu strictu*, that use propellers has come up nearly dry. Some fishes do ventilate egg masses by tail beating while stationary, but I found no specific performance data. In air, at least, one can point to the hive-ventilation system used by honeybees. One or a series of honeybees beat their wings while standing just beyond the entrance to their hive. Hertel (1966) points out that a line of bees constitutes a multi-stage axial compressor analogous to that used in the jet engines of aircraft; one should be aware that the photograph he provides has been inappropriately retouched. Southwick and Moritz (1987) claim that hives "breathe" as the bees alternately pump it out and allow it to inhale elastically. The present discussion of pumps suggests otherwise – even a line of bees should form a high-volume, low-pressure pump, and beehives do not feel as if they have the required low elastic modulus and high resiliency.

(iii) *Ciliated surfaces and chambers:* By contrast with both the previous fluid dynamic pumps, these abound in nature. Muscle must be persuaded to move fluid with some form of transducing equipment; cilia do so as their basic *modus operandi*. Cilia may be far slower in operation than muscle, but a collecting manifold with a decreasing aggregate cross-sectional area can raise the output velocity of a ciliary or flagellar pump. For instance, with their flagellar pumps, sponges eject water at about 0.2 m s^{-1}, a far higher speed than any flagellum can generate directly. Ciliary pumps find wide use for low pressure, high volume applications such as suspension feeding.

But they have several drawbacks. Cilia are microscopic and work at that scale. So scaling up a pipe with ciliated walls encounters the problem of a pumped cross section that increases faster than the pumping circumference. In addition, the cilia-lined pipe cannot have the gently parabolic velocity gradient of a remotely pumped pipe. The entire gradient from the mandatory zero speed at the wall to the peak speed of the pipe cannot span as much as the length of a cilium, so it becomes severe even at modest maximum flow speeds. Since viscous energy loss depends on the steepness of that velocity gradient, ciliary pumping suffers from an intrinsically low efficiency in all but the narrowest pipes and channels. Still, for surfaces across which organisms exchange material or heat, that steep velocity gradient can be advantageous. Thus ciliated surfaces serve admirably for organs such as gills – as on the gills of most gastropod mollusks, where they pump water for respiration, and on those of bivalve mollusks, where they play a central role in suspension feeding (Vogel 2004).

One wonders whether this inauspicious scaling explains their absence on the gills of fish and whether the gills of aquatic arthropods would find them useful were motile cilia known to that phylum. Less puzzling is their absence as pumps in our capillaries. Velocity presents no problem, since blood in our capillaries flows at speeds that cilia can produce. But effective operation is precluded by the relatively high resistances of circulatory systems – the steep slopes of lines from the origin on graphs such as that of figure 1 – together with the low positions of ciliary pump capability lines on the *y*- or pressure axes. We might wish for circulatory systems in which well-disseminated ciliated capillaries make our fallible hearts unnecessary. But the low pressure-generating capability of ciliary pumps rules them out, at least where blood volumes remain under 10% of body volumes, as in both vertebrates and cephalopods. (LaBarbera and Vogel 1982 failed to consider the need to match pump performance to system resistance and mistakenly attributed the choice between ciliary and muscular pumping solely to ancestry.) For the same reason, and as noted earlier, one suspects that pumping cannot be the primary function of the cilia of flame cells and solenocytes.

(iv) *Capillary (surface-tension) pumps:* Inasmuch as it lifts water against gravity, the capillary rise of water in a narrow hydrophilic tube constitutes a proper pump. As does evaporative pumping, such surface-tension pumping works only with an air-water interface; so, similarly, it lies solely in the domain of terrestrial and semi-aquatic organisms. Even with a considerably greater range of cases than evaporative pumping, its role remains by comparison a modest one.

From time to time one runs across statements (by non-biologists) asserting that sap rises in trees as a result of capillarity – simply by the ascent of an aqueous fluid in a hydrophilic tube. That cannot be the case, as repeatedly pointed out (see, for instance, Nobel 2005), because the conduits are just too wide. The capillary rise of water in a circular vertical tube, *h*, is

$$h = \frac{2\gamma\cos\theta}{\rho g r}, \tag{1}$$

where γ is its surface tension, 0.073 N m^{-1}; ρ is its density, 1000 kg m^{-3}; θ is the contact angle (0° for perfect wetting); *g* is gravitational acceleration; and *r* is tube radius. Even under ideal circumstances, water will rise only 1.5 m in a small tracheid, one 20 μm in diameter. In a xylem vessel of 200 μm, as in an oak, the rise will be ten times less. Even with perfect wetting, sustaining a 50 m column of water by capillarity would require a tube less than 0.6 μm across.

Capillarity does matter in a few cases, for the most part situations involving narrow tubes of no great length. Rehydration of dry stems and leaves of the resurrection plant *Myrothamnus* depends on it, but conduit diameters are of the order of 2 μm (Schneider *et al* 2000, Tyree 2001). Some insects, most notably orchid bees, draw in nectar through their probosci at least in part by capillarity (Kingsolver and Daniel 1995, Borrell 2003). At least two kinds of birds use the mechanism, hummingbirds to draw in nectar (Kingsolver and Daniel 1983) and phalaropes to raise small quantities of water with edible plankton up a vertical bill whose tip has been dipped in a body of water (Rubega and Obst 1993).

Capillarity, in the guise of wicking, can move liquid upward on the outside of sufficiently wettable surfaces. A few cases have been described, not surprisingly, in amphibians. Lungless plethodontid salamanders (at least the genus *Desmognathus*) breathe through wet skin and can stay wet by wicking water upward as well as by exuding body water through their exceptionally permeable skin (Lillywhite 2006). Toads (genus *Bufo*), which lack skin mucus, can stay moist by wicking as well (Licht and Lillywhite 1974).

(iv) *Flow inducers:* These may be more common in nature than in modern human technology. While our jet pumps have little in the way of immediate natural analogs, nature capitalizes on both the elevated pressure of oncoming flows and the reduced pressure due to flow over an orifice opening normal to flow. The low pressures they produce impose the main limitation on flow inducing pumps. Pressure cannot deviate either upward or downward from ambient by significantly more than the dynamic pressure difference, Δp, defined by Bernoulli's principle,

$$\Delta p = \frac{\rho v^2}{2}, \tag{2}$$

where v is the speed of flow. (For small-scale and thus low Reynolds number flows, pressures will be still lower.) A fish swimming steadily at 0.5 m s^{-1} can generate only about 125 Pa, or 1/800th atm. A suspension-feeder working at 0.1 m s^{-1} has available a mere 5 Pa, or 1/20,000th atm. Still, a baleen whale swimming with open mouth at 3 m s^{-1} can take advantage of a more substantial 4500 Pa, or 1/22nd atm. I gave considerable space to flow inducing schemes earlier (Vogel 1994) and will do little more than mention representative cases here.

Ram ventilation in fishes is the best-known case of pressure elevation at an orifice facing into a flow. Its use varies fish to fish, with a trivial role for it in some, a role only at high swimming speeds (where respiratory needs are greatest, of course) in others, and a total dependence on it in some large, fast fish that consequently must either swim or asphyxiate (Steffensen 1985). Analogous pumps with upstream-facing inputs drive the suspension feeding systems of some ascidians and caddisfly larvae; they probably drive fluid through the olfactory passages of many fish as well.

Still lower pressures are available for drawing fluid out of an elevated orifice. The arrangement, though, finds use by keyhole limpets and abalones to draw water across their gills for respiration, by some sand dollars to draw food-laden water from underlying sediments up past their oral surfaces and through their slots, and perhaps by the shrimp genera mentioned earlier to irrigate their U-shaped burrows. Sponges take advantage of both elevated pressures on their upstream facing (and indirectly on their other) ostia and the reduced pressures at the oscula through which they discharge water.

(v) *Temperature gradient pumps:* Flows can be induced in several ways by spatial variations in temperature. Some of these were noted in essay 4 (Vogel 2005b); most such pumps move air rather than water. Free convection, the most obvious, drives the internal circulation of some giant African termite mounds (Turner 2000) as well as providing some cooling currents around sunlit trees during periods of unusually low wind. Evaporative pumps, especially sap lifters, were considered earlier as displacement devices. A related arrangement, evaporation in one place and condensation in another – as in heat pipes – might better be regarded as a dynamic pump. But at this point its use by organisms is no more than a suspicion. Finally (and more clearly dynamic) is Marangoni pumping, flow driven by surface tension gradients that follow temperature gradients – again, it remains something to be kept in mind as a distinct possibility for organisms. All of these pumps develop very low pressures.

6. An index for pump performance

Figure 3 gathers data for pressure boost and volume flow for 53 pumps, 37 displacement and 16 dynamic. They were chosen for their diversity in function and the range of values of the two variables they represented; they include sap lifters, hearts, blood suckers, jets, projectile ejectors, gill irrigators, and suspension feeders. The graph appears to confirm the generalization that in nature as in human technology displacement pumps for the most part work at higher pressures and dynamic pumps at lower pressures; a *t*-test of the data gives a significance level for that distinction of about $P = 0.05$. The data support considerably less well the notion that displacement pumps work at lower volume flows, either by inspection of the graph or by another *t*-test.

Can we contrive a single parameter that encapsulates our picture of how biological pumps sort out? One might calculate a ratio of the two variables, $\Delta p/Q$, for each of the pumps on the graph – or for any others. But, like the data for volume flow, the distinction between our two general pump types fails statistical test by a clear margin ($P > 0.2$), even though the mean $\Delta p/Q$ of the displacement pumps comes out higher than that of the dynamic pumps. On that score alone, the ratio has little if any value for biological pumps. One suspects trouble from the huge size range of these pumps together with the intrinsic size- and speed-sensitivity of the ratio. Thus dimensional manipulation of the ratio gives $\rho v/S$, where S is conduit cross section. Assuming consistent density, faster flowing systems will be biased higher values, while larger systems tend toward lower values. A dimensionless ratio might be more informative, assuming one can be found with unambiguously defined and easily determined variables.

One possible ratio is the pressure coefficient, C_p, long used in fluid mechanics to describe pressure distributions around bodies in flow. It divides pressure by dynamic pressure, the pressure that would be generated were the moving fluid to be suddenly halted, as described by Bernoulli's principle. Specifically,

$$C_p = \frac{2\Delta p}{\rho v^2}, \tag{3}$$

in effect a dimensionless form of eq (2). v represents the highest speed in the system, most often at the output of the pump. Pressure change appears, as we would like, in the numerator, while flow speed stands in for volume flow in the denominator.

In practice, though, the pressure coefficient does almost as poor a job sorting out pump types as a pressure-volume flow ratio. Applying it to the 39 of the previous 53 pumps for which I found adequate data yields a distinctly odd ranking. For instance, both the lymph hearts of toads and our own lymphatic vessels, which we mainly pump with our skeletal muscles, have values up with the xylem of trees, a bias attributable to their very low speeds. And the sporangium of the fungus, *Pilobolus*, an osmotic engine, gives the lowest value of all; its very high speed overcompensates

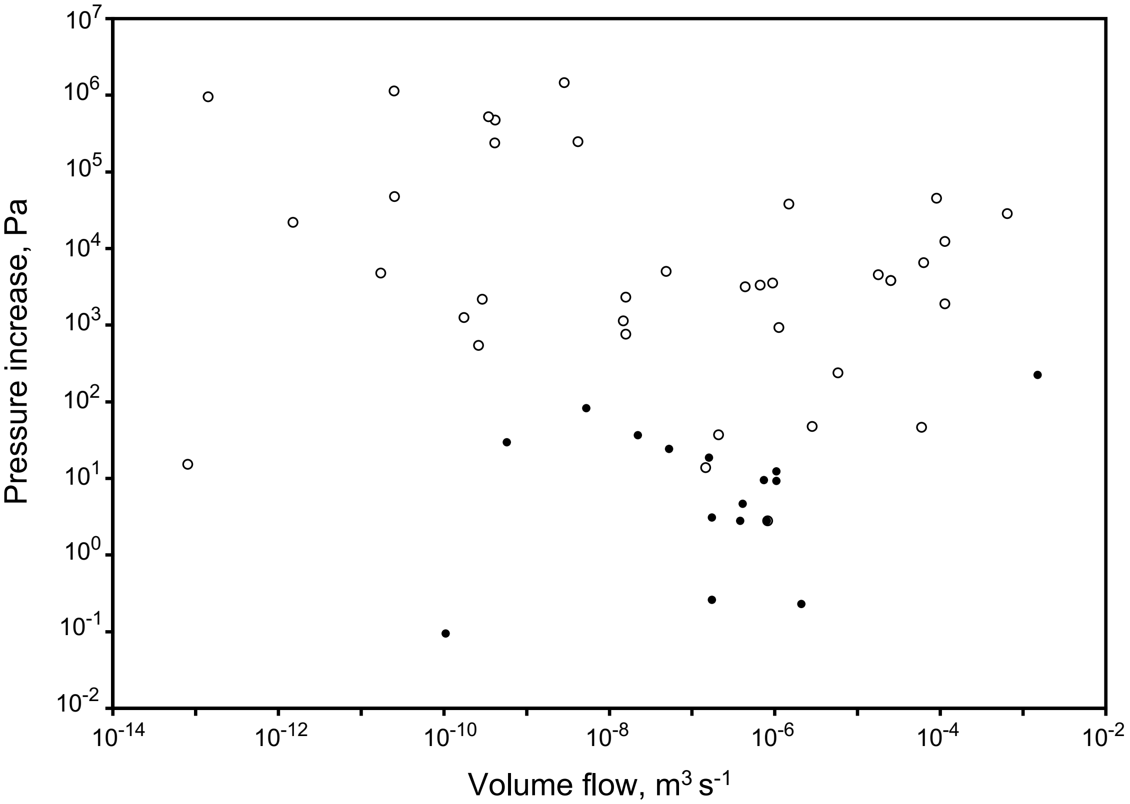

Figure 3. Pressure produced versus the resulting volume flow for a collection of different pumps; all of those described in figure 4 are included, as well as some additional examples (of the same general types) for which I found no data for either radius or flow speed.

for its substantial pressure. At least the ratio puts the very high pressure, low flow xylem of pine, oak, and the vine *Entadopsis* near the top. Its oddities most likely result from the tacit presumption in the formula of flow at high Reynolds numbers – relatively large, fast, and turbulent – rather than at biologically appropriate low Reynolds numbers and laminar flows. (And so one might add a *t* for "turbulent" and designate this pressure coefficient C_{pt}.) Put another way, its denominator reflects an inertial energy loss rather than a more relevant viscous loss.

In my earlier look at pumps (Vogel 1995) the coefficient was for just this reason replaced by one that presumed viscous rather than inertial pressure loss,

$$I = \frac{\Delta pt}{\mu}, \tag{4}$$

where t is time and μ is viscosity. Density has deferred to viscosity, as usually happens in low-Reynolds-number formulas. Among a set of pumps more limited than the present one, xylem and hearts came out at the top, as we think they should. But two displacement pumps produced the lowest indices, the jet of the jellyfish *Polyorchis* and the blood sucker of the bug *Rhodnius*. This last generates the greatest pressure difference known in any animal. The index has a practical problem as well, the interpretation of t, a kind of length-less inverse velocity. Without great conviction, I took it as the transit time for a bit of fluid to pass the part of the system with the greatest resistance. Not only does it take more guesswork than one would prefer but it cannot escape ambiguity for the tapering pipes so common among organisms.

A dozen years later, I offer an alternative dimensionless ratio. This one divides the pressure force, pressure times cross section, by viscous resistive force. The latter, the product of viscosity, flow speed, and vessel radius, comes either from Stokes's law for the drag of a sphere or from an equation (eq 13.17 in Vogel 1994) for the pressure drop of flow through a circular orifice. We might call it the "pressure coefficient for laminar flow" to draw an analogy with the well-established (turbulent) pressure coefficient of eq (3). Specifically,

$$C_{pl} = \frac{\Delta pr}{\mu\nu}, \tag{5}$$

To get the ratio from the most appropriate source, the Hagen-Poiseuille equation for laminar flow through a circular pipe, one has to assume an isometry in which pipe length can be replaced by pipe radius. As we'll see, that assumption occasionally generates peculiar values, and it must be borne in mind when drawing inferences from values of the ratio.

Figure 4 gives values for the 39 pumps previously mentioned, 30 displacement and 9 dynamic. What can we make of these numbers?

While the overall range of values varies 2.5-million-fold, functionally homogeneous groups cluster satisfyingly. The xylem pumps, the jets, and the blood suckers each span ranges of about 10-fold, while the hearts (including both

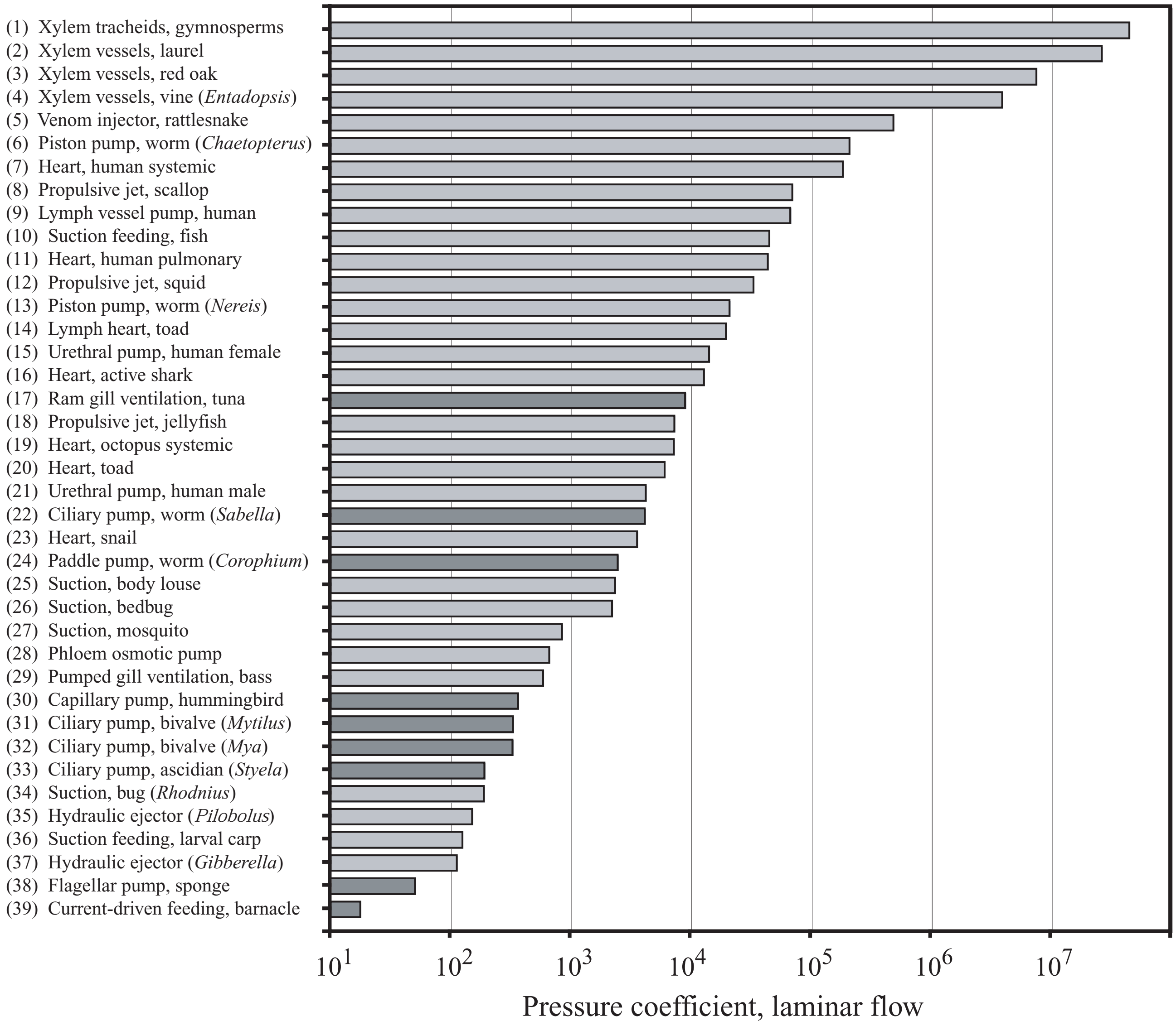

Figure 4. Values of an index of pump performance, a pressure coefficient that applies to laminar flows. Darker bars mark the dynamic pumps. [Sources: (1) Cermák *et al* 1992, Pittermann and Sperry 2003, (2) Cermák *et al* 1992, (3) Kramer 1959, Zimmermann 1971, (4) Fichtner and Schulze 1990, (5) Kardong and Levin-Murcio 1993, Young *et al* 2003, (6, 13, 22, 31, 32, 33) Riisgård and Larson 1995, (7, 11) textbook values, (8) Cheng *et al* 1996, (9) Milnor 1990, (10) Alexander 1969, (12) Shadwick 1994, (14) Müller 1833, Jones *et al* 1992, (15, 21) Glemain *et al* 1990, (16) Lai *et al* 1990, (17) Stevens and Lightfoot 1986, (18) DeMont and Gosline 1988, (19) Wells 1987, Agnisola 1990, (20) Gibbons and Shadwick 1991, (23) Jones 1983, (24) Foster-Smith 1978, (25, 26, 27) Daniel and Kingsolver 1983, (28) Nobel 2005, (29) Lauder 1984, (30) Kingsolver and Daniel 1983, (34) Bennet-Clark 1963, (35) Vogel 2005a, (36) Drost *et al* 1988, (37) Trail *et al* 2005, (38) Bidder 1923, Vogel 1978, (39) Trager *et al* 2000.]

single and dual-stage pumps) vary less than 50-fold. Ciliary and flagellar pumps vary about 6.6-fold, with the flagellar one of a sponge not unexpectedly the lowest. Thus the ratio provides expectations for pumps not yet analyzed – it appears to have predictive value.

The evaporative pumps of xylem come out at the top, with the highest value for the narrower tracheids of gymnosperms – they generate comparable pressures but get less flow from them than the wider vessels of broad-leafed trees. A vine, with (as is typical) the widest vessels, gives the lowest value in the group. r may be larger than for other vessels, but v increases by a greater factor.

Flows over the gills of fish span a wider range, 360-fold, but the distribution of cases within that range looks quite reasonable. Suction feeding, with its necessity for rapid, impulsive flow, yields the highest value. The ram-ventilating tuna also has a high ratio, the highest for any dynamic pump. But it swims exceptionally fast, and so has access to the greatest driving pressure. We also note a low value for pumped gill ventilation, whose respiratory function demands continuous, low cost flow, and suspect that the low value preadapts the system to take advantage of ram ventilation during faster swimming.

Two dynamic pumps, both of suspension-feeders, give values that look anomalously high, the paddle pump of the amphipod *Corophium* and the ciliary pump of the polychaete *Sabella*. But we may be miscategorizing the infaunal *Corophium* as a dynamic pump. Since the paddles operate within its tube they make what is more like a set of moving compartments, as in the human-powered "dragon-bone" pumps in China that move water from one rice paddy to another – and less like the serial paddles on say, a rowed trireme or galley. *Sabella*'s ratio may draw attention to the limitation of the ratio alluded to earlier. It erects a fan of ciliated tentacles normal to flow, in effect a huge number of ciliary pumps operating in parallel. That parallel array may not increase pressure at all or velocity all that much, but it makes use of the radius of the entire array, as done here, at least questionable. Using the distance between individual ciliated elements reduces the value down to the level of the other ciliary pumps, those of two bivalves and an ascidian. I have not altered either the categorization of *Corophium* or the value for *Sabella*, in part to preserve them as illustrative examples and in part to avoid such conscience-troubling *post hoc* adjustments.

A look at refilling in a sea anemone points up the predictive value of the ratio. When disturbed, *Metridium* can contract its body wall musculature and collapse down to form an inconspicuous flat blob, largely by expelling almost all the water in its central gastrovascular cavity through its mouth (Batham and Pantin 1950). It then slowly reinflates, according to textbook accounts, by pumping water back in through ciliated tubes, the siphonoglyphs, while it keeps its mouth closed. Batham and Pantin (1950) measured reinflation pressures around 25 Pa. Using the dimensions of their animals and an estimate (from various sources and personal observations) of an hour for the process, I calculate a pressure coefficient for laminar flow of 1.0×10^5. The value lies between those of our systemic hearts and of the jets of scallops, about 250 times higher than even the exceptionally high value for the ciliary pump of *Sabella*. Perhaps the usual accounts incorrectly assume that pumping by the ciliated tube does most of the work. One need not look far for another player. The viscoelasticity of the body wall of *Metridium* has about the right elastic modulus and temporal behavior for the task, judging from the measurements of Alexander (1969), Gosline (1971) and Koehl (1977).

Several final notes on this measure of pump performance. Notice that the ratio contains the product of pressure and radius. Since for a given material the tolerable pressure varies inversely with the radius of the pipe, this nicely offsets any scaling relationship that relationship might impose. The ratio does less well in correcting for the effects of collecting or expanding manifolds – constricting a flow will increase its speed, both lowering the numerator and raising the denominator.

Unexpectedly, perhaps, whether or not the pumping activity is sustained makes little difference. Venom injection by rattlesnakes and spore ejection by fungi are the quickest of mechanical processes, but their widely different values span almost the entire range of long-acting pumps. And while some blood suckers (like *Rhodnius* and lice) remain painlessly attached to their hosts for long periods, others (like mosquitoes) get the job done quickly lest they be swatted; nonetheless, their ratios differ little. Such indifferences suggest that efficiency, energy and power play, at most, secondary roles in determining the match of pump type to application. That parallels what I noted for the scaling of ballistics, where force (there in the guise of stress) appeared more critical than work or efficiency (Vogel 2005a).

7. Yet another kind of pump

A brief comment earlier noted the imperfection of the dichotomy between displacement and dynamic pumps in human technology. We saw a case of that same imperfection in nature's pumps when categorizing the pump of an amphipod, which used paddles, ordinarily dynamic, to form moving compartments, as in some displacement pumps. Using the two categories to distinguish among the pumps of organisms may run into a more general difficulty, the exclusion of other categories of devices. Perhaps we should bear in mind that biomechanics usually recognizes functional devices in nature by their similarities to devices in human technology. Clearly not everything in our technology devices has a natural equivalent. Less obviously, devices in nature may lack technological analogs – less obviously because that lack makes them easier to overlook.

One of the differences between the two technologies is their relative reliances on rigid versus flexible structural materials. Thus we speak of artifacts that change shape under load as "deformed" – a word whose pejorative sense reflects our preference for rigidity. Think back – only two of the technological pumps mentioned make significant use of flexible materials. Peristaltic pumps squeeze tubing; while inefficient, they avoid possible contamination of fluid with pump parts. Diaphragm pumps pulse a (periodically replaced) rubber or neoprene membrane to change the volume of a compartment; they have some advantage when, as in sewage systems, pumping slurries of suspended solids. But in nature, flexibility appears to be the default condition, with rigid materials used only where functionally mandatory. Can we recognize ones other than peristaltic pumps that depend on flexible materials?

When visiting the laboratory of Mory Gharib a few years ago, I was shown a more refined version of the device pictured in figure 5, one in which a very flexible element completes a circuit of otherwise inflexible conduits. Repeatedly compressing the flexible tubing near one of its ends produces an impressively strong unidirectional flow with the aid of neither check-valves nor peristalsis. Thus this valveless pump can be run in either direction, depending on where the flexible tubing is compressed. How it works seems clear enough – compressing the tube forces fluid in both directions, but one responds by expansion of the flexible tubing rather than by sending the fluid on through the rest of the circuit. Crude models tolerate a wide range of sizes, tube flexibilities, and circuit resistances.

The hearts of ascidians may work the same way as do these models. These hearts have long been known to reverse periodically, which they do by changing the end at which a pacemaker triggers constriction (Martin 1974). They seem not to work by reversible peristalsis but to depend on some other direction-determining arrangement. By cardiac standards, ascidian hearts produce only modest pressures, about 300 Pa in a large one (Goddard 1972), implying only modest resistance levels in their circulations. While insect hearts, also valveless low-pressure devices, commonly reverse, they have usually been described as peristaltic (Sláma 2003). Still, given the great diversity of the arthropods, this kind of valve-less, non-peristaltic, reversing pump might well occur somewhere among them.

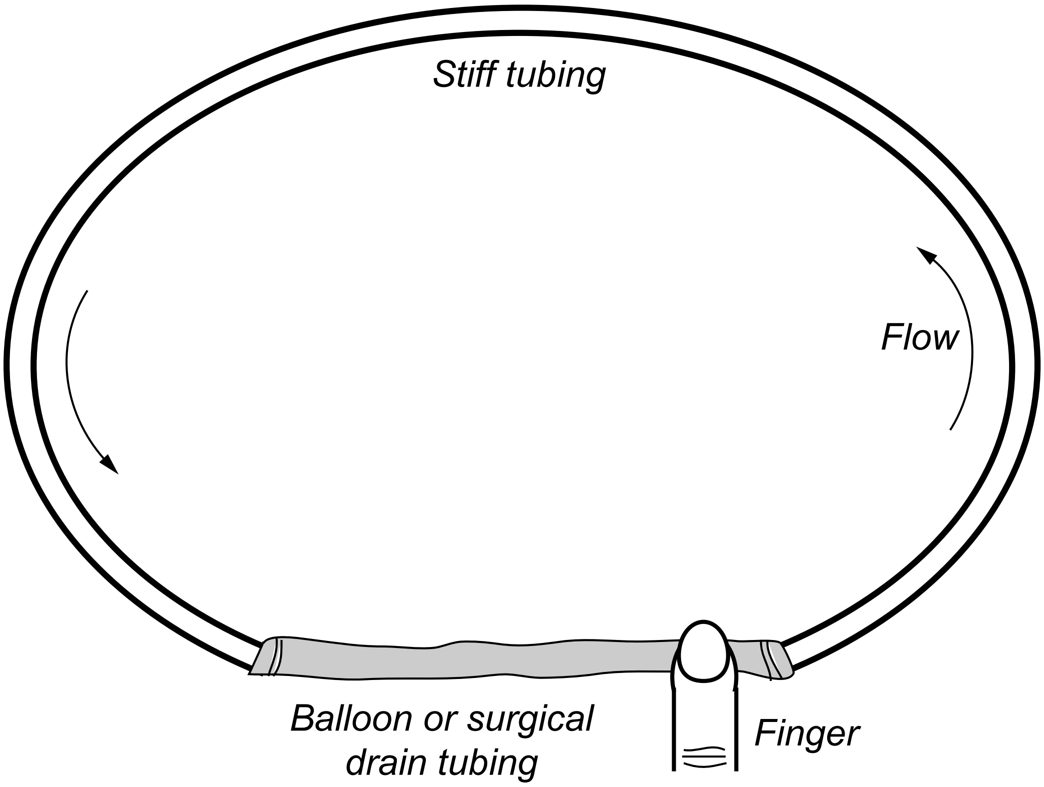

Figure 5. Perhaps the simplest possible solid dynamic liquid pump. Pressing repeatedly near the right end of the flexible tube, as here, drives the fluid counterclockwise around the circuit–as one can see with either a few trapped bubbles or some suspended matter. The pump tolerates addition of a substantial resistive element in the stiff tubing, and even a very crude one can generate 2500 Pa.

Forouhar *et al* (2006) have recently described something analogous to this valveless pump, one also not likely to be limited to a small group of organisms. As they point out, an embryonic vertebrate heart begins pumping well before it develops valves. Working with zebrafish embryos, they showed that such a tubular heart need not depend on peristalsis. Instead, they propose that pumping results from the suction generated by propagation of an elastic wave in the wall of the heart.

Perhaps we should entertain the idea that a third general category of pumps occurs among organisms, one that might be called "solid-dynamic pumps," together with explicit use of the qualifier "fluid" for the dynamic pumps described in earlier sections here. Such solid-dynamic pumps would likely be associated with quite specific tunings of the multi-dimensional properties of flexible biological materials. We vertebrates and cephalopods certainly come close to using such a pump as we buffer the radical pressure fluctuations of our hearts with flexible arterial walls whose stress-strain curves match the requirements set by our various blood pressures.

8. Perspectives and speculations

Two final items. First, what we see here are curious and varied combinations of functional and phylogenetic constraints. Higher plants must do with pumps that need no moving, macroscopic solid parts, which largely limits their options to evaporative and osmotic pumps. Both such displacement pumps can generate impressively high pressures but neither does well when volume flow is the measure. Induction by external flows remains an option, but in terrestrial systems it will move air rather than water. A few kinds of animals such as sponges can make no macroscopic pumping machinery either, but at least cilia and flagella give them reasonable low-pressure, high-flow options. Arthropods know nothing of motile cilia, relying mainly on the movements of rake-like appendages for suspension feeding and on peristaltic and hydraulic pumping for internal bulk transport. Yes, the pumps of nature appear well-chosen for their assigned tasks, but, no, no creature has anything approaching a free choice from a comprehensive catalog.

Second, whether analysing locomotion, photosynthesis or foraging, we biologists have given considerable attention to energetic efficiency as an index of performance. This essay, for instance, implied early on that a maximal product of pressure generated and volume moved per unit time, power output, marked a pump as well matched to its task. Most often, energetic efficiency can be unambiguously defined, and it accords well with prejudices from our physics courses, the physical devices in our lives, and our fuel bills.

But I am skeptical about whether efficiency provides a unique or even a particularly good comparative measure of devices of such disparate function as the pumps that move aqueous liquids through organisms. For one thing, all too often quantity of water moved may not adequately represent useful output. A suspension feeder may prefer to move less water if by doing so it can increase the fraction of edible material it extracts. Similarly the cost relative to oxygen extracted by a gill may be minimized at a different flow rate than one that minimizes cost relative to the volume of water pumped.

For another, pressure ordinarily represents what we might call an unavoidable evil. It may be something with which a system finds itself stuck from first principles, as with the gravitational loss of sap ascent. Or it may reflect some trade-off, with as circulatory vessel size versus the effectiveness of transmural diffusive exchange, or blood volume versus speed of flow. Only in hydraulic systems such as the fungal projectile ejectors does pressure matter as much as flow.

Finally, for many pumps metabolic cost must be the least immediate of considerations – what fraction of its overall output does a rattlesnake devote to squirting venom? Should a mosquito suck more slowly to minimize the cost of getting its dinner? We might assert (admitting the rare exception) that energetic efficiency will matter, if at all, for pumps that operate steadily rather than for those that give an occasional pulse.

Nor, for that matter, need pumping incur any metabolic cost at all. Somewhat paradoxically, the only initial construction and maintenance impose any cost on both the highest and the lowest pressure pumps – solar energy powers the evaporative pump that lifts sap while the energy of fluid moving with respect to a surface powers both ram ventilation and the variously-employed current-induced flows.

In short, a look at pumps may inject some valuable doubt about whether we can find in the living world a straight-forward measure of utility comparable to power or energy in mechanical technology, information in telecommunications and computing, or money in economics.

Acknowledgements

Mory Gharib started me thinking about valveless peristaltic hearts, Ron Grunwald brought up capillarity on the skin of plethodontid salamanders, Fred Nijhout straightened me out on insect hearts, Tim Pedley pointed out the insufficiency of my first index for pump performance, Amy Zanne headed me in the right direction on flow in the xylem of trees, and Melvin Tyree drew my attention to the diversity of instances of standing gradient osmotic flows

References

Agnisola C 1990 Functional morphology of the coronary supply of the systemic heart of *Octopus vulgaris*; *Physiol. Zool.* **63** 3–11

Alexander R M 1962 Visco-elastic properties of the body wall of sea anemones; *J. Exp. Biol.* **39** 373–386

Alexander R M 1969 Mechanics of the feeding action of a cyprinid fish; *J. Zool. London* **159** 1–15

Batham E J and Pantin C F A 1950 Muscular and hydrostatic action in the sea anemone, *Metridium senile* (L.); *J. Exp. Biol.* **27** 264–289

Bennet-Clark H C 1963 Negative pressures produced in the pharyngeal pump of the blood-sucking bug, *Rhodnius prolixus*; *J. Exp. Biol.* **40** 223–229

Bidder G P 1923 The relationship of the form of a sponge to its currents; *Q. J. Microscop. Soc.* **67** 292–323

Borrell B J 2004 Suction feeding in orchid bees (Apidae: Euglossini); *Proc. R. Soc. London* (Supp4) **B271** 164–166

Brainerd E L, Page B N and Fish F E 1997 Opercular jetting during fast starts by flatfishes; *J. Exp. Biol*. **200** 1179–1188

Cermák J, Cienciala E, Kucera J, Lindroth A and Hällgren J E 1992 Radial velocity profiles of water flow in stems of Norway spruce and oak and the response of spruce to severing; *Tree Physiol*. **10** 367–380

Cheng J-Y, Davison I G and DeMont M E 1996 Dynamics and energetics of scallop locomotion; *J. Exp. Biol.* **199** 1931–1946

Curran P F 1960 Na, Cl, and water transport by rat ileum *in vitro*; *J. Gen. Physiol.* **43** 1137–1148

Curran P F and MacIntosh J R 1962 A model system for biological water transport; *Nature* (*London*) **193** 347–348

Daniel T L and Kingsolver 1983 Feeding strategy and the mechanics of blood sucking in insects; *J. Theor. Biol.* **105** 661–672

DeMont M E and Gosline J M 1988 Mechanics of jet propulsion in the hydromedusan jellyfish, *Polyorchis penicillatus*. II. Energetics of the jet cycle; *J. Exp. Biol.* **134** 333–345

Diamond J M and Bossert W H 1967 Standing-gradient osmotic flow; *J. Gen. Physiol.* **50** 2061–2082

Drost M R, Muller M and Osse J W M 1988 A quantitative hydrodynamical model of suction feeding in larval fishes: the role of frictional forces; *Proc. R. Soc. London* **B234** 263–281

Fichtner K and Schulze E-D 1990 Xylem water flow in tropical vines as measured by a steady state heating method; *Oecologia* **82** 355–361

Forouhar A S, Liebling M, Hickerson A, Nasiraei-Moghaddam A, Tsai H-J, Hove J R, Fraser S E, Dickinson M E and Gharib M 2006 The embryonic vertebrate heart is a dynamic suction tube; *Science* **312** 751–753

Forterre Y, Skotheim J M, Dumais J and Mahadevan L 2005 How the Venus flytrap snaps; *Nature (London)* **443** 421–425

Foster-Smith R L 1978 An analysis of water flow in tube-living animals; *J. Exp. Mar. Biol. Ecol.* **34** 73–95

Gibbons C A and Shadwick R E 1991 Circulatory mechanics in the toad *Bufo marinus*. II. Hemodyamics of the arterial windkessel; *J. Exp. Biol*. **158** 291–306

Glemain J, Cordonnier J P, LeNormand L and Buzelin J M 1990 Urodynamic consequences of urethral stenosis: hydrodynamic study with a theoretical model; *J. d'Urol*. **96** 271–277

Goddard C K 1972 Structure of the heart of the ascidian *Pyura praeputialis*; *Aust. J. Biol. Sci.* **25** 645–647

Gosline J M 1971 Connective tissue mechanics of *Metridium senile*. II. Viscoelastic properties and a macromolecular model; *J. Exp. Biol.* **55** 775–795

Hertel H 1966 *Structure, form and movement* (New York: Reinhold)

Hughes G M 1958 The co-ordination of insect movements; *J. Exp. Biol*. **35** 567–583

Johansen K and Martin A W 1965 Circulation in a giant earthworm, *Glossoscolex giganteus*. I. Contractile processes and pressure gradients in the large vessels; *J. Exp. Biol*. **43** 333–347

Jones H D 1983 Circulatory systems of gastropods and bivalves; in *The mollusca, v5* (ed.) K M Wilbur (New York: Academic Press) pp 189–238

Jones, J C 1977 *The circulatory system of insects* (Springfield, IL: Charles C Thomas)

Jones J M, Wentzell L A and Toews D P 1992 Posterior lymph heart pressure and rate and lymph flow in the toad *Bufo marinus* in response to hydrated and dehydrated conditions; *J. Exp. Biol.* **169** 207–220

Karassik I J, Messina J P, Cooper P and Heald C C 2000 *Pump handbook* 3rd edition (New York: McGraw Hill Professional)

Kardong K and Lavin-Murcio P 1993 Venom delivery of snakes as high-pressure and low-pressure systems; *Copeia* **1993** 644–650

Kingsolver J G and Daniel T L 1983 Mechanical determinants of nectar feeding in hummingbirds: energetics, tongue morphology, and licking behavior; *Oecologia* **60** 214–226

Kingsolver J G and Daniel T L 1995 Mechanics of food handling by fluid-feeding insects; in *Regulatory mechanisms in insect feeding* (eds) R F Chapman and G deBoer (New York: Chapman and Hall) pp 32–73

Koehl M A R 1977 Mechanical diversity of connective tissue of the body wall of sea anemones; *J. Exp. Biol.* **69** 107–125

Kramer P J 1959 Transpiration and the water economy of plants; in *Plant physiology v2* (ed.) F C Steward (New York: Academic Press) pp 607–726

LaBarbera M and Vogel S 1982 The design of fluid transport systems in organisms; *Am. Sci.* **70** 54–60

Lai N, Shabetai R, Graham J B, Hoit B D, Sunnerhagen K S and Bhargava V 1990 Cardiac function of the leopard shark, *Triakis semifasciata*; *J. Comp. Physiol*. **B160** 259–268

Lauder G V 1980 The suction feeding mechanism in sunfishes (*Lepomis*): an experimental analysis; *J. Exp. Biol*. **88** 49–72

Lauder G V 1984 Pressure and water flow patterns in the respiratory tract of the bass (*Micropterus salmoides*); *J. Exp. Biol.* **113** 151–164

Lillywhite H B 2006 Water relations of tetrapod integument; *J. Exp. Biol.* **209** 202–226

Lillywhite H B and Licht P 1974 Movement of water over toad skin: functional role of epidermal sculpturing; *Copeia* **1974** 165–171

Martin A W 1974 Circulation in invertebrates; *Annu. Rev. Physiol.* **36** 171–186

Milnor W R 1990 *Cardiovascular physiology* (New York: Oxford University Press)

Müller J 1833 On the existence of four distinct hearts, having regular pulsations, connected with the lymphatic system, in certain amphibious animals; *Philos. Trans. R. Soc. London* **123** 89–94

Nobel P S 2005 *Physicochemical and environmental plant physiology* (Burlington MA: Elsevier)

Ottaviani G and Tazzi A 1977 The lymphatic system; in *Biology of the reptilia, v6* (ed.) C Gans (London: Academic Press) pp 315–462

Pickard, R S and Mill P J 1974 Ventilatory movements of the abdomen and branchial apparatus in dragonfly larvae (Odonata: Anisoptera); *J. Zool. London* **174** 23–40

Pickard W F 2003a The riddle of root pressure. I. Putting Maxwell's demon to rest; *Func. Pl. Biol*. **30** 121–134

Pickard, W F 2003b The riddle of root pressure. II. Root exudation at extreme osmolalities; *Func. Pl. Biol*. **30** 135–141

Pittermann J and Sperry J S 2003 Tracheid diameter determines the extent of freeze-thaw induced cavitation in conifers; *Tree Physiol.* **23** 907–914

Prosser, C L 1973 *Comparative animal physiology,* 3rd edition (Philadelphia: W B Saunders)

Riisgård H U and Larsen P S 1995 Filter-feeding in marine macro-invertebrates: pump characteristics, modeling and energy cost; *Biol. Rev*. **70** 67–106

Rubega M A and Obst B S 1993 Surface-tension feeding in phalaropes: discovery of a novel feeding mechanism; *Auk* **110** 169–178

Schmidt-Nielsen K 1997 *Animal physiology: adaptation and environment* 5th edition (Cambridge: Cambridge University Press)

Schneider H, Wistuba N, Wagner H-J, Thürmer F and Zimmermann U 2000 Water rise kinetics in refilling xylem after desiccation in a resurrection plant; *New Phytol*. **148** 221–238

Schulz J R, Norton A G and Gilly W F 2004 The projectile tooth of a fish-hunting snail: *Conus catus* injects venom into fish prey using a high-speed ballistic mechanism; *Biol. Bull.* **207** 77–79

Shadwick R E 1994 Mechanical organization of the mantle and circulatory system of cephalopods; *Mar. Fresh. Behav. Physiol.* **25** 69–85

Sláma K 2003 Mechanical aspects of heartbeat reversal in pupae of *Manduca sexta*; *J. Insect. Physiol*. **49** 645–657

Southwick E E and Moritz R F A 1987 Social control of air ventilation in colonies of honey bees, *Apis mellifera*; *J. Insect Physiol*. **33** 623–626

Steffensen J F 1985 The transition between branchial pumping and ram ventilation in fishes: energetic consequences and dependence on water oxygen tension; *J. Exp. Biol*. **114** 141–150

Stevens E D and Lightfoot E N 1986 Hydrodynamics of water flow in front of and through the gills of skipjack tuna; *Comp. Biochem. Physiol.* **A83** 255–259

Thompson M V and Holbrook N M 2003 Scaling phloem transport: water potential equilibrium and osmoregulatory flow; *Plant Cell Environ*. **26** 1561–1577

Trager G C, Hwang J-S and Strickler J R 1990 Barnacle suspension feeding in variable flow; *Mar. Biol.* **105** 117–128

Trail F, Gaffoor I and Vogel S 2005 Ejection mechanics and trajectories of the ascospores of *Gibberella zeae* (anamorph *Fusarium graminearum*); *Fungal Genet. Biol.* **42** 528–533

Turner J S 2000 *The extended organism: the physiology of animal-built structures* (Cambridge MA: Harvard University Press)

Tyree M T 2001 Capillarity and sap ascent in a resurrection plant: does theory fit the facts? *New Phytol* **150** 9–11

Tyree M T and Zimmermann M H 2002 *Xylem structure and the ascent of sap* (Berlin: Springer-Verlag)

van Bel A 1993 Strategies of phloem loading; *Annu. Rev. Pl. Physiol. Pl. Mol. Biol.* **44** 253–281

Vogel S 1978 Evidence for one-way valves in the water-flow system of sponges; *J. Exp. Biol.* **76** 137–148

Vogel S 1994 *Life in moving fluids* (Princeton NJ: Princeton University Press)

Vogel S 1995 Pressure versus flow in biological pumps; in *Biological fluid dynamics: symposium of the society for experimental biology* (eds) C P Ellington and T J Pedley (Cambridge, UK: The Company of Biologists) pp 297–304

Vogel S 2004 Living in a physical world. I. Two ways to move material; *J. Biosci.* **29** 391–397

Vogel S 2005a Living in a physical world. III. Getting up to speed; *J. Biosci.* **30** 303–312

Vogel S 2005b Living in a physical world. IV. Moving heat around; *J. Biosci*. **30** 449–460

Vogel S 2006 Living in a physical world. VII. Gravity and life on the ground; *J. Biosci.* **31** 201–214

Wells M J 1987 The performance of the octopus circulatory system: a triumph of engineering over design; *Experientia* **43** 487–99

Yigit N, Güven T, Bayram A and Cavusoglu K 2004 A morphological study on the venom apparatus of the spider *Agelena labyrinthica* (Araneae, Agelenidae); *Turk. J. Zool*. **28** 149–153

Yokoyama, T 2004 Motor or sensor: a new aspect of primary cilia function; *Anat. Sci. Int.* **79** 47–54

Young B A, Dunlap K, Koenig K and Singer M 2004 The buccal buckle: the functional morphology of venom spitting in cobras; *J. Exp. Biol*. **207** 1383–1394

Young B A, Phelan M, Morain M, Ommundsen M and Kurt R 2003 Venom injection by rattlesnakes (*Crotalus atrox*): peripheral resistance and the pressure balance hypothesis; *Can. J. Zool*. **81** 313–320

Zimmermann M H 1971 Transport in the xylem; in *Trees: Structure and function* (eds) M H Zimmermann and C L Brown (New York: Springer-Verlag) pp 169–220

*e*Publication: 1 February 2007

Series

Living in a physical world
XI.
To twist or bend when stressed

STEVEN VOGEL
Department of Biology, Duke University, Durham, NC 27708-0338, USA
(Fax, 919-660-7293; Email, svogel@duke.edu)

1. Introduction

Several themes underlie the essay that follows. While any or all might be deferred to a final encapsulation, perhaps they are better borne in mind while reading it.

For all its arcane and counterintuitive phenomena, fluid mechanics builds its bioportentous aspects on only a few material properties of gases and liquids – density, viscosity and sometimes surface tension. Interest centers for the most part on just two substances, air and water. Solid mechanics, however greater its intuitive familiarity, encompasses a daunting host of potentially significant material properties – three elastic moduli and strengths, corresponding to tensile, compressive and shearing loads; extensibility and compressibility; strain energy storage; work of fracture; up to six Poisson's ratios; hardness; and yet others. In addition, and of similar relevance, it invokes structural properties such as flexural and torsional stiffness. But for any given application, biological or technological, only a few properties must bear directly on functional success. Often a biomechanical investigation must begin with a decision – or guess – as to which properties might matter.

As pointed out by an engineer, the late James E Gordon, humans most often design structures to a criterion of adequate stiffness (plus, of course, a safety factor). He noted that nature, by contrast, appears to design for adequate strength (again plus some margin), a criterion that ordinarily demands less material. More than that material economy, what matters for the present discussion is the implication that a fundamentally different philosophy of design might depend on different suites of properties. We probably need not need devise novel ways to describe materials, but we might well need to make different selections among those properties long established by mechanical engineers.

While structural properties depend on material properties, they also depend on geometric properties—on shape. Put another way, they depend both on what goes into the structure and on its arrangement. Relying on pure materials, simple composites, or preexisting natural materials, we humans typically alter structural properties by tinkering with geometry. Nature is adept at making materials whose properties vary from place to place within individual structures, and she appears to play as much with material as with geometry. Central to her structural technology are anisotropic composite materials, that is, ones whose properties depend on load direction and location within structures. Perhaps the difference can ultimately be traced to the contrast between Nature's tiny factories, cells and their components, much smaller than the structures they produce, and our large factories, which produce smaller products. But whatever the underlying cause, the availability of location-tuned, complex composites must reflect itself in structural designs.

Extending the argument takes a brief introduction to materials and structures. Materials can be stressed in three distinct ways – they can be stretched (tension), squeezed (compression), or sheared. Structures, in addition to these, can be bent, that is, loaded flexurally, or twisted, meaning loaded torsionally. Each of these structural loadings combines several material stresses. Thus a beam extending outward from a vertical support responds to downward bending of its free end by developing tension along and below its top surface, compression along and above its bottom surface, and shear in between, as in figure 1a. A shaft extending outward responds to twisting of its free

Keywords. Bending; elastic moduli; rigidity; stiffness; torsion

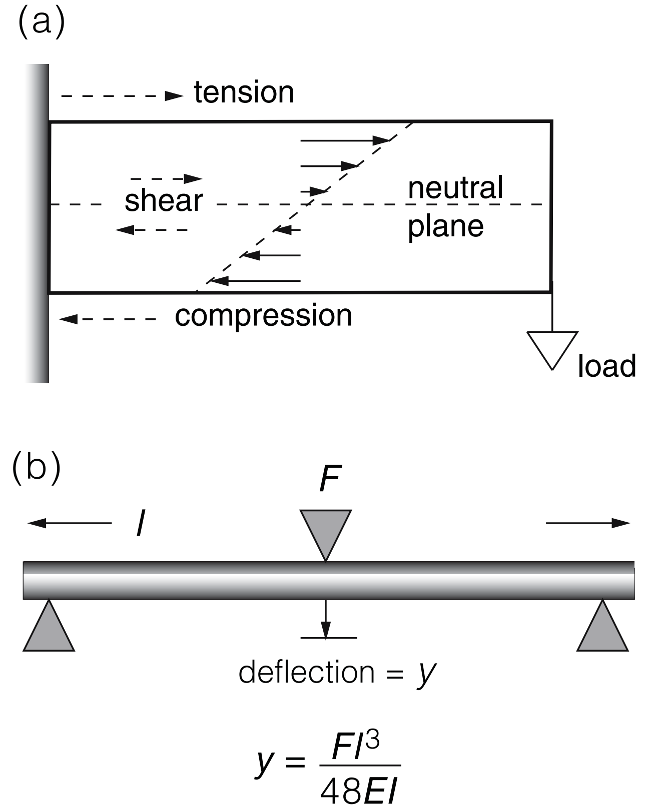

Figure 1. **(a)** The internal forces that result from a normal force on a beam. **(b)** A simple apparatus for measuring *EI* as a composite variable with what is known as three point bending.

end with tension and shear on and near its outer surface while being compressed in the middle, as in figure 2a – we use that compression when wringing out a wet cloth by twisting it. Textbooks first introduce tension, in practice the simplest stress, then compression, and finally shear. When considering structures, they look first at flexion and then at torsion. Perhaps as a result, shear and torsion often get short shrift; perhaps as further consequence, we too easily ignore what appeared as afterthoughts.

In part as a reaction to this underemphasis, I want to focus on torsion. The purpose here goes beyond any sense of social fairness or a physical balance of forces. I worry that we may overlook or misinterpret widespread cases in which torsional stresses and their resulting strains play adaptationally significant roles.

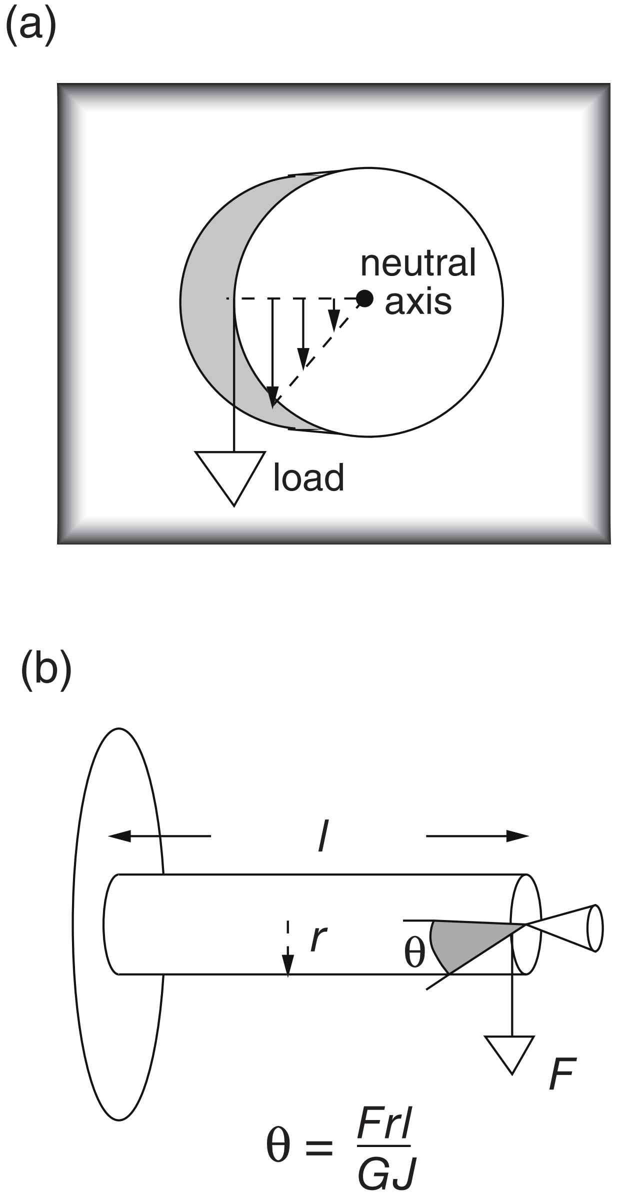

Figure 2. **(a)** The internal shear forces that result from a torsional load on a beam; compression occurs as well–think of how one squeezes water out of wet fabric by twisting it. **(b)** An apparatus, perhaps a metal lathe equipped with locking headstock and live center, for measuring *GJ* as a composite variable.

In examining such cases, we encounter the familiar problems of standards and scale. Compilations of values of shear modulus or torsional stiffness provide poor bases for evaluations, and they serve still less well for comparisons. Values, *per se*, provide neither an obvious frame of reference for the biologist nor convenient bases for relevant comparisons. Moduli refer to materials with no consideration of shape, and structural stiffnesses depend intrinsically on size. Looking, for materials, at stiffness when sheared relative to stiffness when stretched, or, for structures, stiffness when twisted relative to stiffness when bent solves both problems. With equivalent non-torsion-related properties as proper bases for comparisons one gets ratios that are conveniently dimensionless and therefore much less affected by size.

2. Defining the variables

First the variables in some such ratios, beginning with three properties of materials and then, to put the materials into structures, two geometric ones. The material properties come from measurements of how something changes shape under an applied force. To shift from a specific sample being tested to a material, one divides force by cross section to get *stress*. For a tensile test, the complementary variable, *strain*, is extension divided by original length. For shear, strain is the angle turned (as a rectangular solid, say, shifts to the equivalent parallelogram) as a result of a force applied over an area. Stress has dimensions of force over area; strain is dimensionless.

The *Young's modulus of elasticity*, or tensile modulus, is the slope of a graph of tensile stress (ordinate) against tensile strain (abscissa). Since the slope cannot be assumed constant for biological materials (by contrast with metals in particular), one should specify whether a datum gives initial modulus, tangent modulus, or something else. We will consider initial moduli here, bearing in mind that where stress-strain curves are non-linear and loads cause severe shape change, we might be overlooking errors two-fold or worse. The *shear modulus* is the equivalent slope for a graph of shear stress against shear strain, again specifying the place on the graph of the particular slope. Like tensile stress, shear stress is force divided by area, but here the area runs parallel to the force. As angular deformation, shear strain needs no correction for original form. Again, we will assume initial moduli and tolerate some resulting error.

Stretching an object usually makes it shrink in directions normal to that of the stretch. But the amount of shrinkage relative to the stretch varies from material to material. The common measure of that relationship is *Poisson's ratio*, compressive strain normal to the load relative to tensile strain in the load direction. It must be emphasized that the very concept of a single Poisson's ratio for a material presumes isometry, that a material behaves in the same way whatever the direction of the stress. No single ratio can truly describe an anisotropic material. That means just about any living or once living material – all but a few are anisotropic, multi-component composites. Material can be stretched in any of three orthogonal directions, with compensatory shrinkage in the two other directions. Thus a metal may have one Poisson's ratio while wood or bone will have no fewer than six.

One commonly encounters a formula in which Poisson's ratio, ν (lowercase nu), sets a relationship between the Young's modulus, E, and the shear modulus, G, of a material:

$$G = \frac{E}{2(1+\nu)}. \tag{1}$$

Thus something that retains its original volume (as do many biological materials) should have a Poisson's ratio of 0.5 and an E/G ratio of 3.0. I mention the formula and that value in order to assert their total unreliability for biological materials. The shadows of history, practicality, and embedded assumptions afflict the formulas we borrow from engineers even more strongly than those we get from physicists. Equation (1) assumes, again, an isotropic material, one that responds in the same way to loads from any quarter, a condition that the materials of organisms almost never meet. The *inapplicability* of the equation will be crucial in most of what follows.

E/G, though, provides a properly dimensionless ratio, a much better one than ν, for comparing several important properties of materials. That last word, *materials*, points up its main drawback. For it to apply to structures, all those being compared must have the same shape—although not size. To extend it to less homogeneous structures, we need a complementary pair of geometric variables, ones that take account of the ways stresses vary within loaded structures.

When bending (flexing) a structure, both stress and strain will vary with distance from the central plane of bending. For a material with a linear stress-vs.-strain line (again the usual simplification) both stress and strain distributions will take the form shown in figure 1a. Therefore material contributes to stress resistance in proportion to the square of its distance from that central "neutral" plane. We define our geometric variable as the sum of the squares of the distances (y's) of each element of cross section from that neutral plane, with each square multiplied by that unit of cross section, dA. This *second moment of area* (sometimes, ambiguously, "moment of inertia"), I, is thus

$$I = \int y^2 dA. \tag{2}$$

For a solid circular cylinder, for instance,

$$I = \frac{\pi r^4}{4}. \tag{3}$$

Flexural stiffness (sometimes "flexural rigidity"), the resistance of a structure to bending, is just the product of the material factor, E, and the geometric factor, I. In practice, it is usually convenient to measure the composite variable, EI, in a single operation. Not only can one resort to a particularly simple test (as in figure 1b), but – importantly – calculated and effective I's may differ considerably.

For twisting loads, material also contributes to stress resistance, now shear stress, in proportion to the square of its distance from a central element – again assuming (and thus limiting accuracy of the result) a linearly elastic material The element, though, is now a neutral axis rather than a neutral plane, as in figure 2a. So the corresponding geometric variable, J, the *second polar moment of area*, has

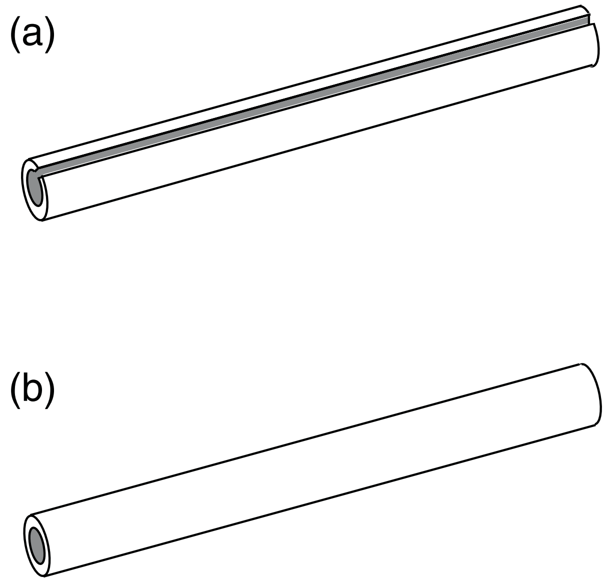

Figure 3. **(a)** An especially twistable structure, one for which conventional calculation of *I* gives a mechanically unrealistic value. **(b)** A similar structure that lacks the lengthwise slit and that resists torsion much more strongly–as calculated from *I*. The pair can be made from a piece of cardboard or plastic pipe for a dramatic illustration of the effect of permitting lengthwise shear.

a nearly identical formula:

$$J = \int r^2 dA. \tag{4}$$

For a solid circular cylinder, as before,

$$J = \frac{\pi r^4}{2}. \tag{5}$$

Similarly, *torsional stiffness* (or "torsional rigidity"), the resistance of a structure to twisting, is the product of the material factor, *G*, and the corresponding geometrical factor, *J*. And, again, measuring the composite variable, *GJ*, as in figure 2b, proves simpler and more reliable than dealing with its elements separately.

In a sense, none of our four key variables can be regarded as ideally tidy and law-abiding. Either modulus, strictly, works only for one location on a stress-strain graph, as already mentioned. *I* and *J* give trouble as well. Figure 3a shows a structure, a hollow cylinder with a lengthwise slit, for which calculated *J*'s greatly overestimate measured (and thus functional) *GJ*'s. Measuring the composite variables is not just simpler but is also less likely to mislead us.

We now have the elements of what was asked earlier – measures of behavior that, by incorporating *I* and *J*, apply to structures rather than merely to materials. Expressing them in a ratio gives what we sought, an adequate comparative basis and size-correcting dimensionlessness. That ratio of torsional to flexural stiffness, then, is *GJ*/*EI*. I prefer to take an arbitrary further step and shift to its reciprocal, flexural to torsional stiffness, or *EI*/*GJ*.

In effect, the inversion tacitly shifts from the world of the engineer to that of the biologist. A technology that values rigidity represents its variables as resistance to deformation, hence flexural and torsional *stiffnesses*. *GJ*/*EI* thus gives twist resistance relative to bend resistance. A biologist looking at natural design, where achieving rigidity seems less often a primary goal, does better with ease of twisting relative to ease of bending. That means, in effect, adopting a "twistiness-to-bendiness" ratio, in conventional terms *EI*/*GJ*. It amounts to a shift in thinking from terms of stiffnesses to ones of compliances. As we will see, *EI*/*GJ* has the additional advantage of yielding values most often above 1.0. Etnier (2003) has used just this ratio to define a "stiffness mechanospace" for elongate biological structures.

Combining the value of 3.0 given earlier for the *E*/*G* of a circular cylinder of an isotropic, isovolumetric material with equations (3) and (5) for *I* and *J* gives us a convenient base line for evaluating values of this twistiness-to-bendiness ratio, For such a cylinder, *EI*/*GJ* will be 1.50. The corresponding values for square and equilaterally triangular sections are 1.77 and 2.49. When comparing cylindrical structures, I will most often cite or calculate values of *EI*/*GJ* rather than *E*/*G* – mainly to anticipate discussion further along of shape effects. Where data exist only for the moduli, whatever the shape, the conversion will assume a circular cylinder and $\nu = 0.5$, so *EI*/*GJ* = *E*/2*G*.

Not only should the ratio of flexural to torsional stiffness, our twistiness-to-bendiness ratio, help assess the role of torsion, but it may provide a simple index of the degree of functional anisometry of a material or a structure. We will return to this role near the end of the essay.

Table 1 provides a summary of what will be a large number of different values of the *EI*/*GJ* ratio.

3. For example, metal versus wood

Adopting the formulations of the mechanical engineers allows us to use data from their handbooks as context for a look at biological structures. Their classic structural materials – although rarely used in quantity before the 19th century – are, of course, metals. The treatment just presented dates from that era; assumptions such as that of linear stress-strain plots (and thus strain-independent moduli) retain the odor of those metallocentric concerns.

Data for both flexural and torsional stiffness abound, the former critical for building large structures, the latter important for choosing rotating shaftwork in order to transmit power. *E*/*G* values range between about 2.4 and 2.8, corresponding to Poisson's ratios of 0.2 to 0.4, significantly below that isovolumetric 0.5. When pulled upon, metals, it

Table 1. Representative values of the twistiness-to-bendiness ratio, *EI*/*GJ*

Circular and assumed circular structures	
Isotropic, isovolumetric cylinder	1.50
Steel shaft	1.3
Commerical (dry) wood (99)	7.15
Tree trunks (5)	7.34
Mature woody vines (5)	3.13
Woody roots (1)	2.34
Long bones (femurs) (2)	2.86
Primate mandibles (2)	1.55
Circular petioles (1)	2.8
Gorgonian corals (13)	3.9
Jointed beams (5)	4.3
Sunflower shoot (1)	1.4
Structures with non-circular cross sections	
Grooved or flat petioles (3)	5.9
Daffodil stems (1)	13.3
Sedge stem (1)	~36
Banana petiole (1)	~75
Locust hind tibia (1)	6.4
Feather shaft (1)	4.8

Parentheses give number of species from which values are averaged. Details, additional values and references are in the text.

turns out, expand volumetrically as they extend lengthwise. We just use them at such low strains that we rarely notice (or have reason to care about) the volume change. For our paradigmatic circular cylinders, that range of *E*/*G*-values gives twistiness-to-bendiness ratios, *EI*/*GJ*s, of 1.2 to 1.4. By altering shape, one cannot easily push those values downward, but they can be elevated to essentially any level by using less and less cylindrical sections – cross shapes, I-shapes and so forth.

As important as metals, and in wide use by humans far earlier, is wood – as dried and shaped pieces of tree rather than the living biomaterial. Wood is anything but isotropic. As every woodworker knows, nothing rivals the direction of the grain in determining its mechanical properties. We glue pieces broadside with crossed grain, forming sheets of plywood and particle board that behave similarly in at least two directions – edge-glued strips gain little more than size and aesthetic advantages. That anisotropy gives it ratios of *E* to *G* far higher than those of metals; once again referring to the *EI*/*GJ* ratios for circular cylinders, I calculate an average value of 7.15 for the 99 kinds of wood tabulated by Bodig and Jayne (1982). Of course one has to specify direction for any anisotropic material – here *E* is for a longitudinal pull and *G* for a longitudinal-radial plane. Despite a fairly wide range of values of *E* and *G*, conspicuous differences in microstructure, and striking variation in practical performance, the ratio changes little from wood to wood, with a standard deviation of 1.21–16.9% of the mean.

The ratio varies less than does the density of the woods, the latter with a standard deviation of 21.1% of the mean value. It also varies a little less than both *E* (20.5%) and *G* (20.8%), which is to say that the two covary and that equation (1) still casts its shadow. Still, applying it as given yields a Poisson's ratio of 12.3, which would imply a fabulous radial shrinkage for even a modest longitudinal tension.

Textbooks on materials and mechanical design, even older ones, appear silent on practical consequence of the great difference between *E*/*G* (or *EI*/*GJ*) for metals and for woods or on the high ratios for woods. Silence in an application-driven field suggests minimal importance. But one can at least envision points of relevance. The high values for woods should affect the behavior of unipodal wooden furniture. Thus wooden lecterns and pedestal tables, even if adequately resistant to bending, will be relatively prone to twisting. But one suspects that anticipation of such problems in design comes more from experience than calculation.

4. Cylindrical structures

Tree trunks. Most tree trunks have circular cross sections, so *EI*/*GJ* can be safely equated (recalling equations 3 and 5) with half of *E*/*G*, and geometric issues can be put aside. That makes tree trunks an obvious starting point for asking what our ratio might tell us. First, though, the data just cited for wood, however relevant to its role as construction material, beg a question that cannot be ignored. Do those data say anything about wood as the material of a tree, as opposed to its performance as a sliced and dried commercial material?

Common experience tells us that dead wood and live wood differ mechanically – a dead twig snaps; a live one bends. Unsurprisingly, comparative measurements confirm the observation. Hoffmann *et al* (2003) reported direct comparisons between dried (at 55% RH) and rehydrated sections of the stems of several tropical lianas (woody vines). Dried *Bauhinia* stems had fully twice and dried *Condylocarpon* stems 1.4 times the Young's moduli of rehydrated stems. Ratios of shear moduli for the two were essentially the same, 1.9 and 1.4. The quotients, *E*/*G* or *EI*/*GJ*, thus differed only minimally.

Roughly the same results emerge in a comparison of my values (Vogel 1995) for freshly collected trunks of small trees with those of Bodig and Jayne (1982) for lumber of the same species. For the five species in common, two gymnosperms and five angiosperms, slicing and drying raises *E* by factors of 1.8 and *G* by 1.75 on average, with lots of variation and no obvious interspecific pattern. *EI*/*GJ*, again on average, remains essentially constant – it drops an insignificant 2.4%. In short, while drying changes the moduli considerably, it has curiously little effect on the ratio of the moduli. Put another way, whatever confers the peculiarly high E/G ratios of wood seems not to depend on the water content of the wood – at the least, these are by no means hydrostatically supported systems.

What, then, are the values of our twistiness-to-bendiness ratio? While data for Young's modulus can be found for a large number of species, shear modulus seems only rarely to be determined. I made paired measurements of the two moduli on freshly-cut lengths of a few trunks, and the data suggest some general patterns. For three hardwood trunks, EI/GJ averages 8.7, with little variation (Bodig and Jayne, 1982, give 8.0 for prepared wood); bamboo culms give an average value of 8.6, insignificantly different. Two softwoods, a loblolly pine and a red cedar have lower ratios, 6.1 and 4.4. (Bodig and Jayne give 8.2 and 4.1). The differences between hardwoods and softwoods, whatever the present values, should not be taken as general, judging from extensive data (Bodig and Jayne 1982, again) on prepared wood, where EI/GJ averages 6.91 ($s = 1.12$) for 52 types of hardwood and 7.51 ($s = 1.13$) for 47 types of softwood.

Clearly the EI/GJ ratios for both prepared wood and freshly-cut circular lengths of small trunks are far higher than the 1.50 of isotropic, isovolumetric circular cylinders. Do high ratios hold functional significance; might they represent, perhaps, a direct product of natural selection? Any assertion of direct selective significance requires stronger evidence. One might well be viewing some indirect consequence of the design of xylem as sap conduits. One expects a set of parallel, longitudinally oriented pipes to be naturally anisotropic Put rubber bands around a bundle of thin, cylindrical, dry pasta, and the structure will twist more easily than it bends, at least if the strands are not too strongly squeezed together.

Still, unless especially symmetrical in both shape and exposure, a tree in a wind will experience both a bending and a turning moment. Accommodating that turning moment with some twist might lessen the associated bending moment by permitting drag-reducing reconfiguration. And the softwoods for which I got low values of EI/GJ, loblolly pine and red cedar, are relatively narrow-crowned and symmetrical trees. But we should resist the seductive appeal of facile functional rationalization. Thus I find the trunks of trees with woven rather than straight-grained wood, such as sweetgum, sourwood, and sycamore, especially hard to split lengthwise for firewood. Yet the ratios for such trees do not differ noticeably from ones that split easily such as oaks and tulip poplar. That should also remind us that the present issue is the utility of a twistiness-to-bendiness ratio as one structural variable, not as the only relevant one. It does not appear to covary with failure point, whether the latter is expressed as strength, extensibility or work of fracture. Nor is there reason to believe that it will correlate with energy absorption before failure.

Woody vines. Of particular interest from the present viewpoint are recent measurements on woody vines, or lianas, common and diverse in tropical forests. The woods of mature lianas mainly bear simple tensile loads rather than the more complex compressive, flexural and torsional mixes of self-supported plant axes. Concomitantly, climbing members of most lineages (and lianas have evolved many times) have lower dry densities and much wider vessels than do the self-supporting trunks considered so far. I looked at attached, climbing stems of two woody vines, a native grape, *Vitis rotundifolia*, and an introduced and escaped ornamental, *Wisteria sinensis*; the first yielded EI/GJ = 2.66, the second an average of 4.48 with especially wide specimen-to-specimen variation (Vogel 1995). These values lie well below the ratios for almost all tree trunks.

But most such plants support themselves at an early stage, and the apical regions continue to do so as they reach out for external assistance. Early stages and the apical portions of climbers have flexural stiffnesses typical of woody trunks; the stiffnesses of later stages are much lower, up to an order of magnitude so. Monocots, lacking secondary (that is, radial) growth, provide most of the exceptions (Rowe *et al* 2006). Again, data on torsional stiffness are fewer, but the latter does not seem to drop in the same manner. For climbing specimens of *Croton nuntians*, EI/GJ drops during ontogeny from about 9 to about 0.8 (Gallenmüller *et al* 2004). For *Bauhinia guianensis* it drops from 9.5 to 5.9, for *Condylocarpon guianense* from its exceptionally low 2.6 to 1.8 (Hoffmann *et al* 2003). In short, all climbers so far tested develop especially low EI/GJ ratios after giving up self-support.

As noted by Gallenmüller *et al* 2004, the ontogenetic change makes functional sense. Not only must freestanding plants support themselves (high E), but climbing plants may depend on flexibility in bending (low E) beyond merely achieving some material economy. Flexibility should limit loading caused by movements of their supportive hosts. The argument for ontogenetic change, though, bears much less on shear modulus (G), since torsional flexibility should have no particular disadvantage at any stage.

[Having given data for low EI/GJ ratios for mature lianas from several studies as well as the case for their functional significance, I do not know what to make of a study by Putz and Holbrook (1991), which reports E and G values for 12 tropical lianas from Puerto Rico. Assuming circular sections (no data are given), they correspond to ratios ranging from 2.0 to 52.8, with an average of 10.0. Omission of the high outlier reduces that average to 6.09. For five trees, by comparison, they give a range of 3.1 to 33.5 and an average of 13.2; again omitting the high outlier drops the latter to 8.1. They provide no information on stage or habit of the specimens.]

Roots. In light of the obvious mechanical importance of roots, data are surprisingly scarce. One might guess that the most serious loads are tensile. One might guess further that a very high Young's modulus could be disadvantageous, quite unlike the situation for an upright trunk – a little "give" will lower the peak stresses of impulsive loads and will facilitate load-sharing among an array of nearly parallel roots. And one might anticipate a lesser degree of functional relevance for shear modulus. Therefore the *EI*/*GJ* ratio may serve mainly to illustrate the possible variation in mechanical behavior of woody structures built with lengthwise conduits and thus anisotropic "grain."

To provide such a contrast, I dug up and tested a few roots (Vogel 1995). Young's moduli are indeed low. For instance, for loblolly pines (*Pinus taeda*), $E = 6.01$ GPa for trunks but only 1.13 GPa for roots, a sufficient difference to be obvious when pieces are handled. Shear modulus differs less, with *EI*/*GJ* dropping from 6.12 to 2.34. So roots appear to be much less anisotropic than are trunks, parallel conduits notwithstanding.

Bones. The nearest analog among animals to tree trunks must be the long leg bones of large terrestrial mammals – elongate, vertical, gravitationally-loaded, and cylindrical or nearly so. (Most but not all long bones are close to circular in section. Cubo and Casinos 1998 give extensive data for those of birds and mammals, noting as an extreme the tarsometatarsi of a parrot with a cross-sectional aspect ratio of 2.0.) Long leg bones must face slightly more diverse loads than do tree trunks, such as those from a variety of muscles and postures and, in particular, the impact loads of running. Still, they will not ordinarily experience major torsional loading – except when we attach long, transverse levers, skis, without provision for load-sensitive release. And they incorporate no stiff-walled, lengthwise vascular elements analogous to xylem.

The Young's modulus of long bones are about three times greater than those of tree trunks, about 20 GPa, exceeding even bamboo culms, at about 15 GPa. But their anisotropy is far less, as indicated either by measurements along transverse axes or in terms of our ratio of *EI*/*GJ*. Human femurs average 2.59. Bovine femur, somewhat more resistant to both flexion and torsion, has a similar ratio, 3.14 (Reilly and Burstein 1975).

What about bones of more complex shape and which might bear more diverse loads? Dental interests, in both the extant and the extinct, have stimulated measurements on primate mandibles. Those of both rhesus macaques and humans have somewhat higher Young's moduli than those of corresponding femurs, but they have notably lower twistiness-to-bendiness ratios – 1.48 for macaques (Dechow and Hylander 2000) and 1.62 for humans (Schwartz-Dabney and Dechow 2003) – applying *I*- and *J*-values for cylinders for the sake of comparison. These ratios do not quite correspond to those of isotropic, isovolumetric materials, though. Poisson's ratios for similarly stressed bone run between about 0.2 and 0.4, significantly non-isovolumetric. Equation (1) and a value of $\nu = 0.3$ (and assuming a circular cross section) gives a still lower ratio, 1.3. In short, ratios for bones are at or below ratios for woods. And the twistiness-to-bendiness ratio appears to give a convenient, if rough, indicator of anisotropy.

Additional items with circular cross sections will appear further along, deferred to facilitate specific comparisons with non-circular ones.

5. Non-circular structures

Shape has yet to play much role in the present discussion, and except for the few data for mandibles, attention has been restricted to elongate circular cylinders. My colleague, Stephen Wainwright (1988) makes a case that such cylinders form the morphological baseline for much of the diversity of macroscopic biological form. As noted earlier, deviations from circular sections increase flexibility to lengthwise twisting loads relative to flexibility in the face of bending loads. And, as we have just seen, analogous increase in *EI*/*GJ* can be produced by incorporation of anisotropic materials. How might nature combine these two routes to the same end?

Petioles and herbaceous stems. My original impetus for invoking the ratio came from an investigation of what the leaves of a variety of broad-leafed plants did in high winds (Vogel 1989). Most reconfigured into cones and (for pinnately compound ones) cylinders, thereby reducing flutter, which might shred them as it does flimsy flags in winds, and drag, which might uproot or break the parent trees. With leaves exposed individually, their petioles (leaf stems) should feel only tensile loading. But when, as should be more common in nature, groups of leaves were exposed to winds, they typically reoriented into stable clumps. Clumping requires that petioles twist lengthwise. Thus structures that resist bending, acting as cantilever beams that hold leaf blades extended from branches, should at the same time accommodate twisting. That argues for elevated ratios of *EI*/*GJ*, whether achieved by material or geometric specialization.

Petioles did have values of *EI*/*GJ* well above the isotropic, isovolumetric baseline, whether their cross sections were circular or non-circular. Non-circular ones, though, had higher ratios than did circular ones. Thus a typical circular petiole, red maple (*Acer rubrum*) had a ratio of 2.8; ones with some lengthwise grooving (as in figure 4a) averaged about 5.0. If twisting to cluster is important, one might guess that grooving should be more common among shorter than among longer petioles – short ones would need more twist per unit length. Mami Taniuchi (unpublished) examined the

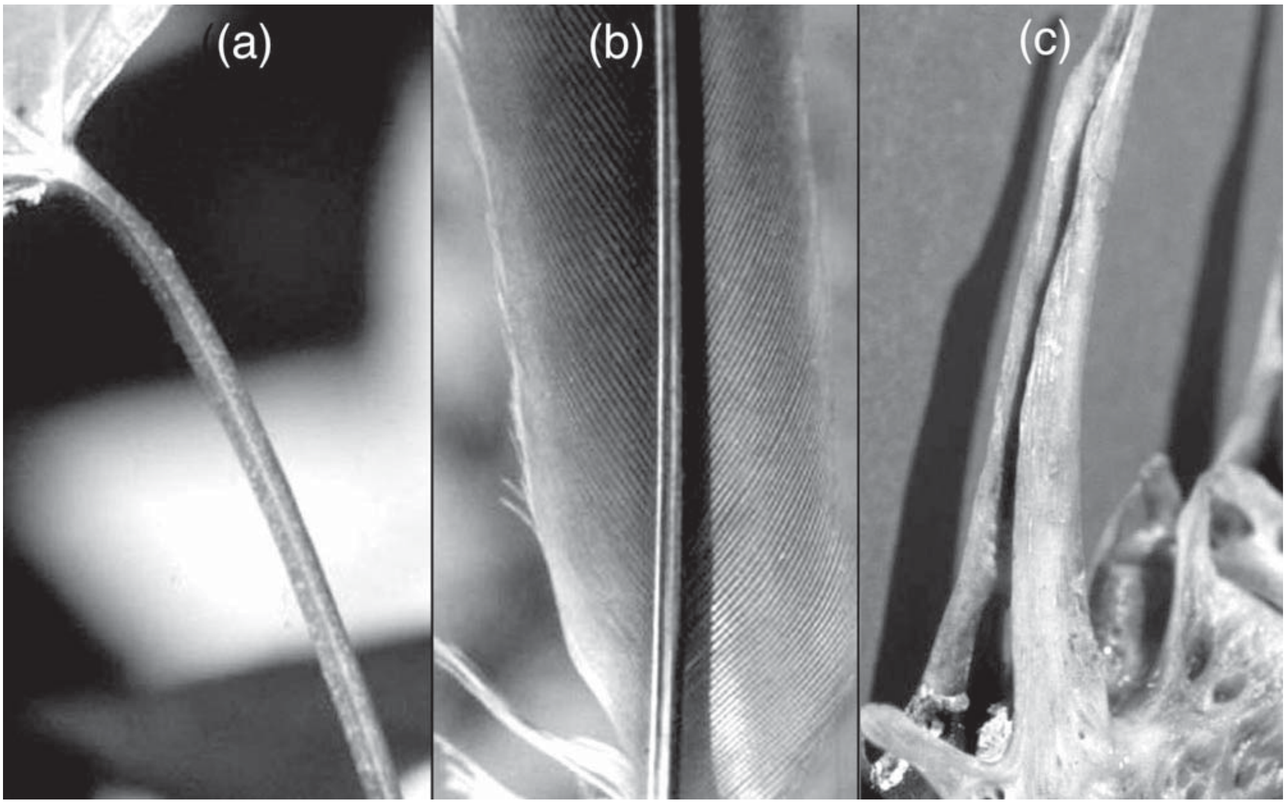

Figure 4. Some biological beams with longitudinal grooves. **(a)** Petiole of sweetgum (*Liquidambar styraciflua*); **(b)** feather shaft of blue jay (*Cyanocitta cristata*); **(c)** neural spine of an unidentified bony fish.

literature on petiole lengths and cross sections; she found some indication of such a correlation, but at the margin of statistical significance. While we cannot safely assert specific adaptation – that structural variation has been driven by functional imperatives – convergences such as this one would provide evidence of the operation of natural selection on a specific feature (Vogel 1996).

The value for bilaterally flattened petioles of white poplar (*Populus alba*), with lateral bending (as in pictures of wind-driven clustering) was 7.7. That bilateral flattening characterizes the genus, which includes trees such as cottonwood (*P. deltoides*) and quaking aspen (*P. tremuloides*). Niklas (1991) gathered more extensive data for the last, finding that petioles increased in *EI*/*GJ* from 2.11 to 9.62 as they developed. All of these species have a strong propensity to oscillate in modest winds, using a mechanism described by Bschorr (1991), and much speculation has focused on the function of that visually and aurally attractive habit. I subjected individual leaves of *P. alba* to the strongest wind I had available, about 30 m s^{-1}.. While that speed shredded most other leaves, these suffered no obvious damage. From that and casual observations through binoculars of a variety of leaves in thunderstorms, I suggest that instability, especially torsional, at modest speeds goes along with good reconfigurational ability at high speeds. *Populus* leaves, with their flattened petioles, are simply the extremes in both regards. In short, the low-speed shimmering simply represents an otherwise functionless concomitant of good high-speed performance. And the trees of that genus are especially common at high altitudes, in open plains, and along coasts—places where strong winds are common.

Several cultivated herbaceous stems gave analogous results. For tomato, with a circular section, $EI/GJ = 3.9$; for cucumber, with a cruciform section, $EI/GJ = 5.4$. While the former grows upright and free-standing, the latter is either recumbent or climbing – but I hesitate to draw any functional inference on the basis of this limited comparison.

Yet another comparative study also found relatively low values for circular sections, but ones that still remain above the expectation for structures made of isotropic materials with reasonable Poisson's ratios. Niklas (1997) looked at the hollow internodes of six herbaceous species, all of these with circular sections between their cross-wise septae dividing the internodes. Despite wide phylogenetic diversity and over four-fold ranges of *E*'s and *G*'s, the ratios of *E*/2*G*, or *EI*/*GJ*, varied very little from 2.5.

However far beyond our baseline, even the highest figures so far noted should not be regarded as extreme. The flower stems of daffodils (*Narcissus pseudonarcissus*) are hollow and lenticular rather than quite circular in section. They bear single apical flowers well to the sides of their long axes, flowers that "dance" in the intermittent gusts common near ground level – as alluded to by several British and American poets. As put by William Paley (1802), posthumously famous as the darling of the anti-Darwinians, "All the blossoms turn their backs to the wind, whenever the gale blows hard enough to endanger their delicate parts." As confirmed by Etnier and Vogel (2000) wind on the off-axis flowers produces torsional loading of their stems, causing them to swing around and "face" downstream. That reduces the drag of the flowers, in effect the flexural load of the wind. We reported an average *EI*/*GJ* value of flower stems of 13.3, with the remarkably low standard deviation of 1.0. By comparison, tulip flower stems, with circular stems that bear axially symmetric flowers, had *EI*/*GJ* values of 8.3 ± 3.2

Two other structures have far higher values, the flower stems of a sedge (Ennos 1993) and the petioles of a banana leaf (Ennos *et al* 2000). Sedge (*Carex*) flower stems stand erect but curving to one side. Thus winds will load them both flexurally and torsionally. They are triangular in cross section; for an isotropic, isovolumetric material that might raise *EI*/*GJ* from 1.50 to 2.49, as noted earlier. In fact their ratios proved much higher. While both flexural and torsional stiffness dropped by more than an order of magnitude with height above the ground, *EI*/*GJ* changed much less and peaked about half-way up. Its values ranged from 22 to 51. What appears responsible for such radically high values are mechanically isolated, lengthwise, peripheral bands of lignified material in the stems.

Similarly, banana petioles, exceptionally large for herbaceous structures, extend both upward and outward. In contract with sedge flower stems, they have U-shaped cross sections. But like sedge stems, they have peripherally concentrated, longitudinal, isolated lignified elements. These play a major role in permitting twistiness-to-bendiness ratios from 40 to 100, the highest of any natural structures yet measured. These high *EI*/*GJ* values ensure that, rather than bending, banana petioles will twist away from the direction of the wind.

Other non circular structures. Most present data on elevation of the twistiness-to-bendiness ratio through use of non-circular sections come from stems and petioles. But that predominance should not be taken as indicative of an unusual reliance on the device by such structures. It more likely reflects the predilections of investigators and a rare case in which more has been done with plants than with animals. Our animal data at this point mainly provide fingers pointing to systems ripe for more extensive scrutiny.

Gorgonian corals, common in shallow tropical seas, are relatively non-rigid but still erect. Jeyasuria and Lewis (1987) reported *E* and *G* values for 13 species from the West Indies, although with no information on cross-sectional shape except a note that some were circular, others elliptical. Assuming circularity produces *EI*/*GJ* values from 1.6 to 6.5 with an average of 3.9. Of interest in the present context is a comment that species in which polyps surround the circumference of the branches twist more easily, while those with single rows of polyps on the sides of the branches twist less easily. They interpreted the difference as reflecting a greater need to maintain torsional orientation where polyps are aligned in rows.

The joints of arthropod legs rarely if ever incorporate analogs of our hips, ankles, shoulders and wrists, all capable of considerable rotation. When dismembering a decapod crustacean, for instance, one quickly learns that legs disarticulate when twisted. Perhaps absorbing torsional loads through shaft twisting could reduce the demands on such torsionally vulnerable joints. Most leg segments of arthropods appear circular, but quite often a strip of especially thin cuticle extends lengthwise. This thin region may be most familiar in the walking legs (pleopods) of crabs and lobsters, whether fresh, frozen or cooked. It might provide the functional equivalent of a groove or other *I*-lowering device, increasing the relative flexibility of the segmental shafts to torsional loads.

Despite a wealth of descriptive information and illustrations, I found no measurements of the behavior of torsionally-loaded leg segments. So in parallel to the measurements on petioles, I tested a few hind tibias of freshly caught locusts, (probably *Dissosteira carolina*), these being the least tapered segments of their most powerful jumping legs. *EI*/*GJ* averaged 6.4, providing some support for the argument. Specimens were quite vulnerable to local buckling, so only very slight torsional strains could be imposed, and three-point bending tests had to be done with loops of thread rather than the usual point contacts.

The vane-bearing shafts (rachises) of the long, outer feathers of birds, especially those of tails and wings, are more obviously non-circular in section. They may have lengthwise grooves on their lower surfaces (as in figure 4b), or they may be nearly square in section with thinner lower than upper sides. I made a few measurements on pieces of shaft from the primary wing feathers of song sparrows (*Melospiza melodia*), obtaining *EI*/*GJ*-values of about 4.8 – again indicative of a structure that preferentially accepts torsion.

One can make a similar functional argument, here based on flight aerodynamics. Wing feathers, like propeller blades, should not bend excessively. But animal wings, incapable of full rotation, must alternately move up and down. So angles of incidence of wings and primary (wing tip) feathers for producing lift and thrust must shift between half-strokes. For primary feathers that means reversing their lengthwise twist. Proper torsional flexibility can enlist aerodynamic forces to cause that switch in twist, eliminating dependence on muscles and nerves. That the grooving or thinning is ventral rather than, as in petioles, dorsal (= abaxial), comes as a surprise only until one realizes that, while petioles hang from branches, birds hang from feathers – feathers in flight bend upward rather than downward.

Another set of non-circular structures that might bear investigation are the neural (dorsal) spines sticking up and rearward from the centra of the vertebrae of large, bony fishes. These (as in figure 4c) often have U-shaped sections. A resulting increase in torsional flexibility might be important in permitting a fish to bend its trunk. If the spines were fully vertical, bending would not impose any torsional load, but their rearward tilt requires that they twist as the overall fish bends.

6. Planar systems

Up to this point attention has centered on cylindrical or nearly cylindrical structures. Hollow structures were fully circular, with no lengthwise openings into their lumens. Planar and near-planar structures experience and respond to torsional stressees in ways of equal biological cogency; among nature's designs, flat surfaces simply happen to be somewhat less common – or at least less diverse. Not that they are truly rare – examples include the leaves of higher plants, many macroalgal fronds, the vanes of feathers and the wings of insects.

Most insects use indirectly-acting muscles to power the strokes of their flapping wings, muscles that attach at neither end to the wing articulations but instead act by reshaping the cuticle surrounding their thoracic chambers. The small direct muscles that insert on the bases of the wings supposedly rotate and camber the wings, making the changes necessary between each alternating half-stroke. In addition, they have been held responsible for reversing the lengthwise twist between half-strokes – as mentioned for the wing feathers of birds – to maintains a near-uniform angle of attack along a wing's length. (The propellers of airplanes, ships and turbines face no such problems since they, unlike flapping wings, normally rotate in a single direction.) The precise phasing of these direct muscles drew little attention despite the severe demands that would place on a neuromuscular system dealing with wings that beat hundreds of times each second.

Ennos (1988) drew attention to a more realistic mechanism, one in which their intrinsic and locally tuned torsional flexibility enabled insect wings to use aerodynamic forces to effect these rapid changes in wing contour. That paper forced reevaluation both of the role of the direct muscles and of the role and arrangements of wing veins. A second paper (Ennos 1995) provided a general analysis of the torsional behavior of cambered plates, as found in leaves (especially grasses), feather vanes (as noted earlier), and cuticular plates in arthropods.

Coincidentally, the first successful human-built aircraft took advantage of torsional flexibility in just the same manner – if without the rapid reversals demanded by flapping. In the 1903 Wright Flyer, the pilot controlled turns by shifting a slide that, through cables, warped opposite wings so they would produce different amounts of lift. Vertical struts connected upper and lower wings with deliberate omission of the cross bracing that would have provided torsional stiffness. What look like cross bracing (and are inaccurately shown as such in many drawings) were, in fact, those wing-warping cables. Hinged ailerons, as still used, soon replaced wing warping, initially as Henri Farman's way, in 1908, of circumventing the Wright's patent on their system of control (Anderson 1997).

Rolling a flat surface into a cylinder without sealing the joint to prevent shear produces an apparently cylindrical beam or column that resists bending but has exceptional torsional flexibility. One can produce a model by doing nothing more elaborate than rolling up a sheet of paper. Such an arrangement occurs in many xeromorphic (drought-adapted) grass blades. According to the usual explanation, the device reduces water loss; noting that stomata are on the inside of the cylinder, it views the roll as a functional addition to the sunken stomata of more planar xeromorphs. The roll might also have a mechanical role, although no one seems to have looked into the possibility. The tops of tall grasses ordinarily bend to one side, so wind will load blades torsionally. As suggested earlier, torsional flexibility can reorient such a structure so more surface area is parallel to flow and downwind from other surface. That would, of course, decrease drag and reduce the chance of flexural buckling – what agricultural scientists call "lodging."

The parent phenomenon, aeroelasticity, has been of considerable interest to aircraft designers, but they typically focus on trouble rather than on adaptive mechanism. Tilting a wing to change its lift usually moves the center of pressure fore or aft, changing the torque on the wing and on its attachment to the fuselage. In at least one aircraft used during World War One, a Fokker D8, the wings sometimes detached as a disastrous result (Gordon 1978). Aeroelasticity has at least occasionally be put to positive use in aircraft – for example, a recent small, high performance military jet (a US F-18) has been fitted with torsionally aeroelastic wings. Aeroelasticity deserves more attention from biologists. We touched on it when considering the reconfigurations of leaves in winds; similarly, it takes on importance for tree trunks and similar structures in rapid flows; and it might be used to induce oscillations that detach seeds and spores into dispersing currents of air or water.

7. Jointed systems

When considering arthropod legs, mention was made of joints and their behavior in torsion. Joints ordinarily bring muscles into the picture, along with an ability to make short-term adjustments of mechanical behavior. One can, for instance, deliberately stabilize a wrist to resist either twisting or bending as a task demands. When we jump, we spontaneously adjust the compressive stiffness of our legs, mainly at our ankles, so the stiffness of legs plus surface remains nearly constant over a wide range of surface compliances (Ferris and Farley 1997). We also adjust ankle torsional stiffness, if slightly less so (Farley *et al* 1998). Whether our actions are voluntary or involuntary and for both wrists and ankles, the two stiffnesses are separately controllable.

Several biological systems make use of beams of alternating joints and stiff portions. Etnier (2001) looked at such multi-jointed systems in horsetails (*Equisetum*), crinoid (echinoderm) arms and crustacean antennae, obtaining EI/GJ values not greatly different (1.8 to 6.6) from those of most simple biological beams and columns. Here, though, the possible functional significance of values above that baseline of 1.5 for isotropic, isovolumetric cylinders remains unclear – a pattern, if there is one, in the behavior of such jointed systems has yet to emerge. What does emerge is further evidence that natural structures can achieve a range of values with joints and muscles as well as by material and geometric characteristics of elongate, passive, solid elements themselves.

8. Dissecting the variables

For simple circular cylinders, again, E/G would serve just as well as would EI/GJ in providing a handy index for the degree of anisotropy. Admitting a role for shape, as defined by the two moments of inertia, permits us to dissect anisotropy into an instructive trio of components. Stretching the usual meaning of anisotropy we can regard the ratio of the elastic moduli, E/G, or as used here, $E/2G$, as "material anisotropy." That of the second moments, $2I/J$, then becomes "geometrical anisotropy," with a convenient base line of 1.0 for cylindrical sections. And their product, EI/GJ, constitutes "structural anisometry," in the form of a twistiness-to-bendiness ratio.

From either two ratios, of course, one can get the third. The easiest route will usually consist of measuring EI/GJ, obtaining (with the caution previously noted) $2I/J$ from cross-sectional shape, and then calculating $E/2G$. For only a few cases can EI/GJ yet be teased apart in this way. For daffodil stems, for instance, $E/2G = 10.0$ since $EI/GJ = 13$ and $2I/J = 1.3$ (Etnier and Vogel 2000). For sedges (Ennos 1993) a $2I/J$ of 1.25 implies an $E/2G$ of roughly 30. Such values, together with the elevated EI/GJ ratios of circular beams such as bones and tree trunks, suggest the following general characteristic of biological designs. Structural anisometry comes far more from an unusually high E relative to G, that is, from high material anisometry, than from high I relative to J, that is, from geometric anisometry. Put another way, deviation from circularity commonly indicates a high EI/GJ, but ordinarily it represents the lesser element of causation. For this reason EI/GJ provides almost as good an indicator of underlying material anisometry as does $E/2G$.

That forms a striking contrast with the structures of our human technology. We take a piece (or melt) of metal or plastic and then treat it as a homogeneous material in fabricating some desired shape. In effect, we accept a value of $E/2G$ of about 1.4 as a given. Particularly when making large structures, most shifts in the structural ratio, intentional or incidental, come from adjustment of geometric anisometry. Older materials such as stone, cement, and brick get similar treatment. Structures such as I-beams and metal fence posts get high EI/GJ ratios entirely as consequences of their cross-sectional shapes. And in our mechanized, mass-productive society, we factor out of our factories and leave to the province of craft workers the paradigmatic high E/2G material, wood.

9. Looking still further afield

Hydroskeletons. Everything so far implies, first, that material and structural specialization can raise the twistiness-to-bendiness ratio above a baseline value and, second, that functional advantage may be gained by its elevation. By extending the ratio downward we can display quite a different set of biological designs on the same spectrum. Lengthwise anti-grooves will not lower EI/GJ, but an widespread arrangement should do so quite effectively. It consists of an incompressible but highly non-rigid core surrounded by an outer flexible skin within which fibers run helically in both directions. These so-called hydroskeletons provide support, among others, for limp annelid and stiff nematode worms, for the tube feet of echinoderms, for the mantles of squid, and for the bodies of sharks.

Twisting such a fiber-wound, pressurized cylinder one way puts tensile stresses on one set of fibers; twisting it the other way stresses the other set. In all such systems the fibers in the outer membrane are relatively inextensible, so they strongly resist such torsion. At the same time, little except stretch of the membrane itself limits bending. The net effect has to be a low twistiness-to-bendiness ratio. That torsional stiffness is rarely noted. Clark and Cowey (1958) provide the classic description of hydroskeletons, Alexander (1988) puts their operation into the context of motility as well as support, and Niklas (1992) extends the discussion to plants. (The only known hydrostatic supportive systems with lengthwise and circular reinforcement rather than such crossed-helical windings are those of non-bony mammalian penises, as discovered by Kelly (2002). Their function demands resistance to lengthwise compression, provided by their combination of circumferential fibers and a constant volume interior. Lack of torsional stiffness probably matters little.)

Unfortunately, hydrostatic supportive cylinders do not lend themselves to the kind of mechanical testing that might measure *EI*/*GJ* ratios – imagine an apparatus that twists a worm. I made a few measurements (Vogel 1992) on a partially hydrostatic system, young shoots of sunflowers (*Helianthus*) about 10 to 15 cm high – the average ratio was 1.4, slightly below our base line of 1.5. But the scatter was wide, with some specimens yielding values around 1.0. Fully hydrostatically structures should have considerably lower values.

Just as a high *EI*/*GJ* ratio should provide functional advantage to many terrestrial, gravitationally-loaded systems, a low ratio might be maladaptive under such circumstances. Perhaps the scarcity of classic hydroskeletons in mature terrestrial systems compared to aquatic ones comes from that need to hold erect structures of a far higher density than that of the medium.

A *tale of a tall building*. Finally, an example of unexpected relevance of this twistiness-to-bendiness ratio—and torsion generally. This tale of a faulty tower comes from Levy and Salvadori (2002). When completed, in 1974, the John Hancock Tower, a tall (234 m), slender building, in Boston, MA, was applauded for its elegance. But soon thereafter a variety of problems appeared. While, among the effects of wind, falling exterior glass panels gained the most notoriety, a more interesting failing was the way the building underwent quite unpleasant twisting motions. Between the high aspect ratio of its cross section, about 3: 1, its rhomboidal shape, and the much-admired lengthwise grooves running up the smaller two sides, the structure turned out to be unexpectedly lacking in torsional stiffness – too little *GJ* for its *EI*.

The cure, not an inexpensive one, consisted of two parts. About 1500 t of diagonal bracing were added. And, occupying the far ends of the top floor and resting on a thin layer of oil, a pair of 275-t masses were connected to the structure through springs and shock absorbers. These passive dampers, tuned to oscillate in opposite phase with the building, compensated for most of the torsional motion.

Nature's designs may be less rigid than those things we build, but perhaps they can provide guidance if one applies an adjustment of scale. Traditional rigidity takes a disproportionate amount of material in very large structures. So our tall buildings sway and our large aircraft flex their wings. Thus natural design at the more modest scales of animals and plants may hold relevance for our efforts when we work at much larger scales.

The present discussion holds a final lesson, one not new to this or the earlier essays. In reading and reviewing material in biomechanics, I have all too often encountered an inappropriate level of confidence in the applicability of formulas obtained from the engineering literature. For that matter, I have been guilty of the practice myself. They have to be taken with a larger discount than the ones we biologists encountered in our physics course. Both their intrinsic accuracy as descriptors of reality and the conditions for them to apply may be severely restricted – in ways that may matter more to us than to their originators. In general, the more complexly multidimensional the underlying physical situation, the further along on a spectrum from precise predictors through rules of thumb to completely inapplicable are the equations one encounters.

Acknowledgements

Melina Hale suggested looking at feathers, Andy Rapoff steered me into the literature on bone, Hugh Crenshaw, more agile than I, caught the locusts, Mami Taniuchi, as noted, analyzed the literature on petioles, Matthew Healy supplied the quotation from Paley, Charles Pell brought up fish backbones, and Steve Nowicki provided fresh sparrow feathers. Kalman Schulgasser and Shelley Etnier provided valuable general discussion.

References

Alexander R M 1988 *Elastic mechanisms in animal movement* (Cambridge: Cambridge University Press)

Anderson J D Jr 1997 *A history of aerodynamics* (Cambridge: Cambridge University Press)

Bodig J and Jayne B A 1982 *Mechanics of wood and wood composites* (New York: Van Nostrand Reinhold)

Bschorr O 1991 Winderregte blattschwingungen; *Naturwissenschaften* **78** 402–407

Clark R B and Cowey J B 1958 Factors controlling the change of shape of certain nemertean and turbellarian worms; *J. Exp. Biol.* **35** 731–748

Cubo J and A Casinos 1998 The variation of the cross-sectional shape in the long bones of birds and mammals; *Ann. Sci. Naturelles* **1** 51–62

Dechow P C and Hylander W L 2000 Elastic properties and masticatory bone stress in the macaque mandible; *Am. J. Phys. Anthropol.* **112** 553–574

Ennos A R 1988 The importance of torsion in the design of insect wings; *J. Exp. Biol.* **140** 137–160

Ennos A R 1993 The mechanics of the flower stem of the sedge *Carex acutiformis*; *Ann. Bot.* **72** 123–127

Ennos A R 1995 Mechanical behaviour in torsion of insect wings, blades of grass and other cambered structures; *Proc. R. Soc. London* **B259** 15–18

Ennos A R, Spatz H-Ch and Speck T 2000 The functional morphology of the petioles of the banana, *Musa textilis*; *J. Exp. Bot.* **51** 2085–2093

Etnier S A 2001 Flexural and torsional stiffness in multi-jointed biological beams; *Biol. Bull.* **200** 1–8

Etnier S A 2003 Twisting and bending of biological beams: distribution of biological beams in a stiffness mechanospace; *Biol. Bull.* **205** 36–46

Etnier S A and Vogel S 2000 Reorientation of daffodil (*Narcissus*) flowers in wind: drag reduction and torsional flexibility; *Am. J. Bot.* **87** 29–32

Farley C T, Houdijk H H P, Van Strien C and Louie M 1998 Mechanism of leg stiffness adjustment for hopping on surfaces of different stiffnesses; *J. Appl. Physiol.* **85** 1044–1055

Ferris D P and Farley C T 1997 Interaction of leg stiffness and surface stiffness during human hopping; *J. Appl. Physiol*. **82** 15–22

Gallenmüller F, Rowe N and Speck T 2004 Development and growth form of the neotropical liana *Croton nuntians*: the effect of light and mode of attachment on the biomechanics of the stem; *J. Plant Growth Regul.* **23** 83–87

Gordon J E 1978 *Structures, or why things don't fall down* (London: Penguin Books)

Hoffmann B, Chabbert B, Monties B and Speck T 2003 Mechanical, chemical and X-ray analysis of wood in the two tropical lianas *Bauhinia guianensis* and *Condylocarpon guianense*: variations during ontogeny; *Planta* **217** 32–40

Jeyasuria P and Lewis J C 1987 Mechanical properties of the axial skeleton in gorgonians; *Coral Reefs* **5** 213–219

Kelly D A 2002 The functional morphology of penile erection: tissue designs for increasing and maintaining stiffness; *Integr. Comp. Biol.* **42** 216–221

Levy M and Salvadori M 2002 *Why buildings fall down, updated and expanded* (New York: W W Norton)

Niklas K J 1991 The elastic moduli and mechanics of *Populus tremuloides* (Salicaceae) petioles in bending and torsion; *Am. J. Bot*. **78** 989–996

Niklas K J 1992 *Plant biomechanics: an engineering approach to plant form and function* (Chicago: University of Chicago Press)

Niklas K J 1997 Relative resistance of hollow, septate internodes to twisting and bending; *Ann. Bot.* **80** 275–287

Paley W 1802 *Natural theology* (London: Charles Knight 1856)

Putz F E and Holbrook N M 1991 Biomechanical studies of vines; in *The biology of vines* (ed.) F E Putz and H A Mooney (Cambridge: Cambridge University Press) pp 73–97

Reilly D T and Burstein A H 1975 The elastic and ultimate properties of compact bone tissue; *J. Biomech*. **8** 393–405

Rowe N P, Isnard S, Gallenmüller F and Speck T 2006 Diversity of mechanical architectures in climbing plants: an ecological perspective; in *Ecology and biomechanics* (eds.) A Herrel, T Speck and N P Rowe (Boca Raton FL, CRC Press) pp 35–59

Schwartz-Dabney C L and Dechow P C 2003 Variations in cortical material properties throughout the human dentate mandible; *Am. J. Phys. Anthropol.* **120** 252–277

Vogel S 1989 Drag and reconfiguration of broad leaves in high winds; *J. Exp. Bot*. **40** 941–48

Vogel S 1992 Twist-to-bend ratios and cross-sectional shapes of petioles and stems; *J. Exp. Bot.* **43** 1527–32

Vogel S 1995 Twist-to-bend ratios of woody structures; *J. Exp. Bot.* **46** 981–985

Vogel S 1996 Diversity and convergence in the study of organismal function; *Israel J. Zool.* **42** 297–305

Wainwright S A 1988 *Axis and circumference*. (Cambridge, MA: Harvard University Press)

*e*Publication: 25 April 2007

Series

Living in a physical world
XII.
Keeping up upward and down downward

STEVEN VOGEL
Department of Biology, Duke University, Durham, NC 27708-0338, USA
(Fax, 919-660-7293; Email, svogel@duke.edu)

1. Introduction

In defining an organism's immediate physical situation, one begins with position and orientation, just as when distinguishing between translational and rotational motion. Position always matters, even if no exact specification need be given for the pelagic or the aerial. Orientation can be ignored in only a few instances – non-motile spherical unicells in a continuous medium, perhaps a colonial *Volvox* in a pond, under some circumstances maybe bird eggs and round nuts.

We ordinarily treat orientation mechanisms as matters of coordination, putting them in treatments of neurobiology. Detectors, such as proprioceptors, provide information with which nervous systems direct appropriate muscular activity. We less often look at the underlying physical situations, at potential perturbing forces, at the options available for reorientation, and at devices for maintaining orientation. (But I must immediately applaud a symposium held a few years ago, reported as Fish and Full 2002.) One essay certainly cannot do the subject justice – especially since we know quite a lot about the subject in a somewhat scattershot way – but perhaps the main bases can be touched.

Elementary textbooks of physics recognize three mutually exclusive situations, ones in which bodies in their positions subject to a gravitational (or analogous) force are either stable, unstable, or neutral. As in figure 1, the distinctions hinge on differences in either their own mass distributions relative to their footprints or on the contours of the supporting surfaces. (The issue of substratum contour will be put aside as a second-order one, left for the time being to people concerned with behavior.)

A neutrally stable object simply has no preferred orientation. At most, perturbation adds rotational momentum, which then keeps the thing rotating. The commonest neutrally stable objects are rigid cylinders (one neutral axis) and spheres (two such axes) lying on rigid surfaces, and objects in continuous media whose centers of gravity coincide with that of the fluid they displace. I will say no more about these latter cases, relatively uncommon among macroscopic systems, simply noting that neutral stability opens a possibility for ground-level wind dispersal – one exploited by, at least, Russian thistles (tumbleweeds) in the drier parts of North America.

Almost all our passive possessions are stable around at least two axes – tip one a bit north or south, east or west, and it returns by itself to its previous orientation. Most often this comes about because the work of perturbation raises the object's center of gravity, and the gravitational restoring force then drives its subsequent reorientation. Put another way, the perturbation generates a restoring force, and the system enjoys the resulting negative feedback. For instance, we make pencils with hexagonal cross sections to stabilize their rotational axes. Petroski (1990) describes the origin of the practice, one especially handy for cultures with sloping desks. By contrast, while an unstable object may be stationary, any perturbation will upset the balance and produce runaway reorientation – a process with positive rather than negative feedback. As we will see, seemingly impractical, even dangerous, precariousness turns out to be widespread among living organisms.

"Stable" and "unstable" positions require quantitative qualifiers if one considers all but minimal perturbations. Turn a stable object far enough and it ordinarily becomes unstable, so if need be one could plot, say, turning moment against angle, with the shift from stable to unstable marked by the angle at which the moment dropped through zero. This kind of static stability requires another note as well,

Keywords. Falling; flying; locomotion; maneuverability; stability; swimming

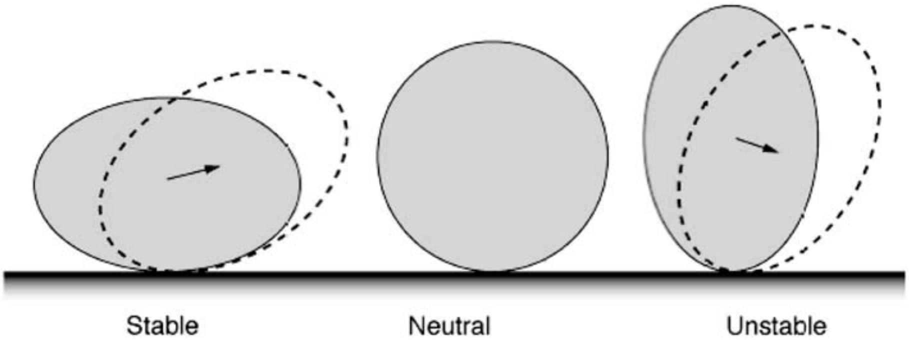

Figure 1. The stability of three gravitationally-loaded rigid bodies on horizontal, rigid substrata – assuming, of course, uniform density and no other forces.

something to bear in mind. Most often the downward force of gravity enforces resistance to lateral perturbing forces, but any other unidirectional and externally applied force would do. Whatever the origin of the force, crucial to the shift of orientation are lines of action and their moment arms.

For many biological situations, this static view proves inadequate. Often we need to consider dynamic stability as well. Acceleration in effect tips the direction of the restoring force, and it may shift its line of action as well. The speed of application of a perturbation commonly bears on its consequences and the effectiveness of any active response. As does the duration of a period of instability; for instance, that of any airborne phase of a terrestrial gait. The activities of an organism itself can either produce or offset instability – one thinks immediately of the location of control surfaces on swimmers and fliers relative to their centers of mass, buoyancy, and pressure. Responses to perturbation can be sluggishly overdamped or sufficiently underdamped to permit transient, sustained, or even increasing oscillations. Problems of dynamic stability have bedeviled vehicular design at least since wheelbarrows came into use, in China, over a thousand years ago.

2. Static stability – sessile systems

For sessile organisms well attached to substrata, the issue of stability in the present sense remains moot. Only attachment strength, along with the magnitude and line of action of any lifting force and vulnerability to peel failure, hold consequences for a limpet, snail, mussel, or waterpenny beetle on a rock. In moving water, only a few sessile organisms, such as flounder (plaice), manage to hold position without some secure attachment, and these are neither erect nor exposed to especially rapid flows. Among sessile terrestrial ones, including some of the largest, though, remaining in place may rest on gravitational stability.

Many, probably most, terrestrial plants manage to attach themselves to the ground with sufficient strength to resist the turning moment of wind-induced drag. But reliance on well-ramified, tension-resisting roots becomes ever more problematic with increasing size. Greater height increases both the speed of local winds and the turning moments they cause. And a greater area of foliage raises the drag caused by a given wind. However one assumes that attachment effectiveness scales, it will most likely increase with a lower exponent than that moment. Some quite large plants do appear to manage mainly by ground-grabbing, most notably bamboos and tropical trees that can take advantage of an ample general tangle of roots in the soil. Some, where I live most notably large specimens of the loblolly pine (*Pinus taeda*), limit turning with a stiff, deep, central taproot, essentially a downward extension of the trunk.

Essay 7 (Vogel 2006a) considered a tree that resists uprooting, not by attachment to the ground, but by being gravitationally stable. With ample weight, a low or deflection-resistant center of gravity, and a wide, stiff, partially buried base, the "up-" in "uprooting" takes on an especially literal significance. When such a tree does uproot, the lower portion of the trunk commonly rests a meter or two above the ground. Figure 2a shows such a tree, one that grew in a fairly open and unsheltered location. An instructive variation of the arrangement has repeatedly appeared in trees that live in the shallow water of swamps. Weight near the base, where it will not move laterally when wind-loaded, increases stability most effectively. But what matters is effective weight, that is, weight less buoyant force. The densities of almost all fresh woods lie below that of water, and even the few denser ones are not much denser. So trunk volume below the water line has little stabilizing value – and these trees produce conspicuous trunk enlargements just above the water, as in figure 2b.

3. How to stand on legs

A tripod can be stable if its center of gravity falls in the area defined by three straight lines with the legs at their ends.

Figure 2. **(a)** The lower part of a willow oak (*Quercus phellos*) grown in a fairly exposed location. **(b)** The lower part of a baldcypress (*Taxodium distichum*) in the shallow water of a coastal swamp in North Carolina, USA.

Additional supporting legs merely increase the number of lines needed to establish that area, assuming flexibility of leg length or joints or else substrate compliance allows the extra legs to make useful ground contact. Naively, then, we might assume that unipeds (as are some standing birds) and bipeds (ourselves) are unstable when standing, while quadrupeds, pentapeds (kangaroos, at times), and yet more leggy creatures can stand stably. That assumes amply stiff joints, a boundary condition effectively defining the other end of biological reality. In fact unipeds and bipeds need not be intrinsically unstable since feet can provide sufficient contact area to circumscribe the line of action of gravitational force. But for them stability does demand broad, stiffly articulated feet. The instructive exception, not hard to experience, consists of standing on stilts – virtually impossible without fairly frequent changes in contact points.

Since no animal in nature engages in bipedal stilt-standing, no standing animal need be intrinsically unstable. Yet however easily achieved, few if any standing animals take full advantage of stable postures. Even we humans, while we appear to stand directly over our feet with our weight borne by compression of our leg bones, deliberately court instability. We tilt slightly forward and then offset that shift of center of gravity by muscular action, principally through tension in the large muscles of our calves and the backs of our thighs. (Hasan 2005 describes the remarkably complex system involved.) We – and most (perhaps all) erect quadrupeds – continuously sense position and adjust the output of our tonic muscles. Sensing is critical, carried out, as Sir Charles Sherrington (1906) pointed out a century ago, by an elaborate proprioceptive system that signals forces and lengths of muscles. We pay scant attention to its operation unless doing something mildly unnatural such as standing for a prolonged period on one leg. Toy horses, cows, and humans stand only precariously; real ones are not such easy pushovers.

Standing posture varies systematically with body size in a way that makes sense for a slightly unstable system which does not profligately produce force and work for continuous corrective movements. A large mammal stands almost perfectly erect, clearly bearing nearly all its weight on the lengthwise bones of its legs. A small mammal stands in a crouch, with leg joints flexed to one degree or another. A crude rationale goes as follows. Muscle makes up about the same fraction of body mass of all mammals, and the contractile force a muscle produces varies with its cross-sectional area. So, all else equal, the small mammal can exert more forceful corrections and maintain a less bone-

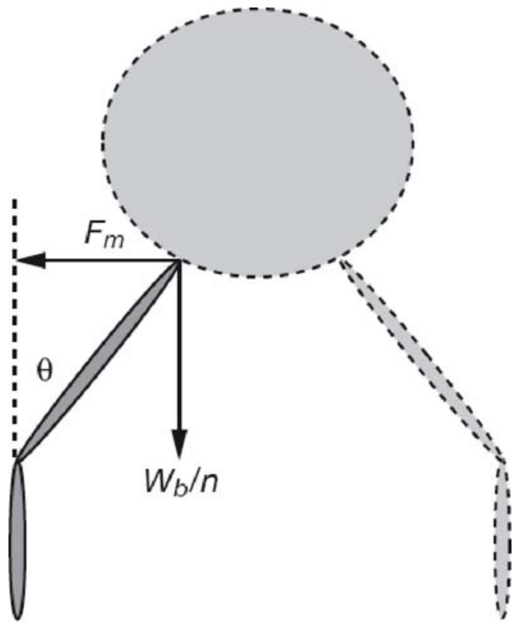

Figure 3. Adjoining long bones of a standing animal with the flexion angle, θ, between them. W_b is body weight; F_m is the force the relevant muscles must exert to offset the flexion; n is the number of (identically loaded) legs.

supported and more unstable posture relative to its body mass. If, as in figure 3, a joint is flexed by an angle, θ, body weight is W_b, and muscle force is F_m,

$$\frac{F_m}{W_b} \propto \tan\theta \cdot \tag{1}$$

F_m scales with the square of body length, W_b scales with length cubed, so their ratio must scale inversely with length and, assuming creatures of similar shape and density, force relative to body weight should vary inversely with length or $W^{-0.33}$. Thus (with constant gravitational acceleration), flexure angle should vary as

$$\tan\theta \propto M_b^{-0.33}. \tag{2}$$

While I have not seen a direct test of the prediction, Biewener (2003) predicts and supports a similar one, taking a somewhat different approach. He cites data that give $M^{+0.26}$ for "effective limb mechanical advantage," close to the inverse of the flexure angle used here – the difference between 0.26 and 0.33 is unlikely to be significant. He notes (and I completely agree) that, by lowering the center of gravity, a flexed stance confers advantages when an animal accelerates, both for linear and angular acceleration. In short (one might say), it imposes a cost in stability – more forceful corrective motions – but enhances maneuverability.

One predicts, therefore, that the height when standing of a mammal's center of gravity with respect to body mass will vary with an exponent less than the 0.33 of isometry – a rationale for our common observation that small mammals are on average long and stand low, big ones are short and stand high. Anticipating just a bit, the flexed stance and concomitant change in muscle location also lower muscular efficiency in locomotion. Muscle does best if shortened slowly, as measured in muscle lengths per unit time – "intrinsic rate of contraction". But the advantage in this respect of smaller size normally gets offset by the higher stride frequency of smaller animals. Add the size-dependent variation in location, and the small animal needs higher intrinsic rates and has to pay a higher price in cost of transport, energy relative to mass moved times distance.

Cows, as expected, stand on almost unflexed legs. A practice termed "cow-tipping" enjoys a widespread body of folklore, at least in the United States. Supposedly one or a few people who sneak up at night on a sleeping, standing cow in a field can push it over, with distinctly detrimental effects on the animal – taking advantage of the narrow window of stability consequential to a high center of gravity and closely spaced legs. An analysis by Lillie and Boechler, at the University of British Columbia, (2005) concludes that a standing cow has sufficient stability to require an impractically large force for such a prank, about 3000 N (equivalent to 300 kg). Thus if single human can push about 300 N at the requisite height, ten synchronized pushers would be needed. (Lillie and Boechler assume what I think is an overly generous estimate of the push a typical human can exert; Cotterell and Kamminga 1990 cite a datum for maximum pull of 280 N, which ought to be about right for pushing as well and which I have just rounded off).

Pushing force, though, may not be the key constraint that renders the stories apocryphal. More importantly, cows do not sleep standing up, and when standing, they have the usual dynamic instability and ever-vigilant reflexes that one experiences if one tries to tip over a dog or cat. If the cow can respond by modestly widening its stance, even without an overall shift of its center of gravity, about 4000 N or 14 pushers would be needed – quite a challenge to deploy without excessively disturbing the cow. As Young-Hui Chang has found (personal communication), even a flamingo that stands quietly on one leg above a splayed foot makes continuous minor muscular adjustments as directed by its proprioceptive system.

4. And how to walk on legs

Moving about on legs adds other destabilizing factors. Indeed, for most forms of legged locomotion requiring continuous stability would greatly limit motions and gaits; and in practice most such systems are at least slightly unstable most of the time. Two unavoidable factors loom largest. First, progress demands pushing rearward on the substratum, ensuring that the line of action of the propulsive force lies below that of the resistive forces of inertia (resisting acceleration) and drag (resisting speed). Second, moving demands that at least one leg be lifted from the substratum, so an animal must be supported by one less than its normal complement.

That mismatch between an animal's rearward push and the resistance to that push produces a moment that turns the animal head-backwards. In practice, a forward shift of the center of gravity provides the compensatory head-forward moment. Acceleration, most notably at the start of a walk or run, takes more of a shift than does moving steadily; with only minor air resistance for most forms of terrestrial locomotion getting mass into motion greatly exceeds drag as an impediment. One leans forward when walking with the greatest tilt when one starts, and walking when half submerged in water (more resistance, less gravitational force) takes a greater tilt. We start walking by deliberately falling forward, and we cannot stand in quite the same posture we use while walking at anything over the most modest pace. Adding resistance by asking that a person push or pull makes matters worse, eliciting the same response, an increased forward tilt.

Increasing the number of legs from two to four does not solve the problem, although it may make it easier to manage by not demanding that the center of gravity shift to a statically unstable location. So asking that a quadruped push or (more commonly) pull still exacerbates instability. At one time horses were often harnessed to the carriages of the ostentatious with checkreins that kept their heads high. That limited their pulling abilities and obviously distressed them. A 19th century children's classic, *Black Beauty* (Sewell 1877) made much of that and other abusive practices and probably contributed to its abandonment.

When walking, bipeds such as ourselves sway slightly from side to side as one shifts support back and forth from one leg to the other, again displacing support from beneath the center of gravity. Again, one cannot stand in most positions one assumes while walking at all but the slowest speed – snapshots of standing postures cannot be arranged into a walking sequence. Thus to fore-and-aft we add lateral instability, although the latter averages out over time. At least lateral shifts put the line of action of the center of gravity closer to the ground-contact point of the leg that will then take the load and provide propulsion. Penguins, relative to their heights the shortest-legged of birds, sway the most; they use the gravitational shift as part of a pendulum-like energy minimization scheme (Griffin and Kram 2000). That sway characterizes a quadrupedal gait called "pacing" or "racking" in which both right and then both left legs move simultaneously. Horses can be taught to pace; long-legged giraffes and camels normally pace, most likely because these long-legged animals can then use greater leg-swinging angles without front-hind interference. Pacing, of course, imposes the same instability as bipedal walking, losing a major advantage of quadrupedalism. Still, both some especially short-legged walkers and some especially long-legged walkers sway – if for different reasons.

A variable called "duty factor," the fraction of the time a given leg provides support, helps us judge whether stable locomotion is even possible. A two-legged walker cannot be fully stable even at a duty factor of 1.0 – standing still – as already noted. Oddly enough, that does not demand that bipedal walkers have a continuously vigilant and actively intervening proprioceptive system. Small wind-up walking toys or similar unpowered downhill walkers do quite well, although they usually have feet of biologically unreasonable areas. Their centers of gravity may never lie directly above a leg, but properly timed foot-falls limit falling to one side or the other. Coleman and Ruina (1998) have devised a more elaborate version of such a self-compensating downhill walker, a physical model that can be put together from widely available toys ("Tinkertoys"), as well as a theoretical treatment of this kind of stability-while-in-motion. Powered versions (Collins *et al* 2005) walk with efficiencies comparable to those of walking humans with only the most minimal control.

A four-legged creature can be stable, but only if it never has more than one leg off the ground, that is, if the duty factor equals or exceeds 0.75. And it can do that only if it shifts its center of gravity slightly by, say, tilting the body, away from whichever leg is held aloft, as in figure 4, keeping the center of gravity over the triangle formed by the contact points of the remaining legs. In effect, a leg has to give the ground a slight body-tilting push just before breaking contact. Quadrupeds do walk stably (postural reflexes aside), but mainly when they go quite slowly, as when stalking prey. We might expect slow walking to demand rather than merely allow static stability, but work on chelonians – turtles and tortoises – tells us quite the opposite. While duty factors generally run well over 0.75, Jayes and Alexander (1980) found that at times only two feet contacted the ground. They provide a persuasive (if counterintuitive) argument that eschewing stability permits slower and more efficient muscle action.

But slower locomotion does impose a lower tolerance for instability. As Alexander (2003) notes, while forces need only balance when averaged over a stride, during any unstable period an animal falls with gravitational acceleration, g. And some relatively constant fraction of leg length, h, must limit the permissible falling distance. As an indicator of the need to preserve stability, he suggests a dimensionless expression based on these variables and on stride frequency, f:

$$\frac{g}{2f^2h}. \tag{3}$$

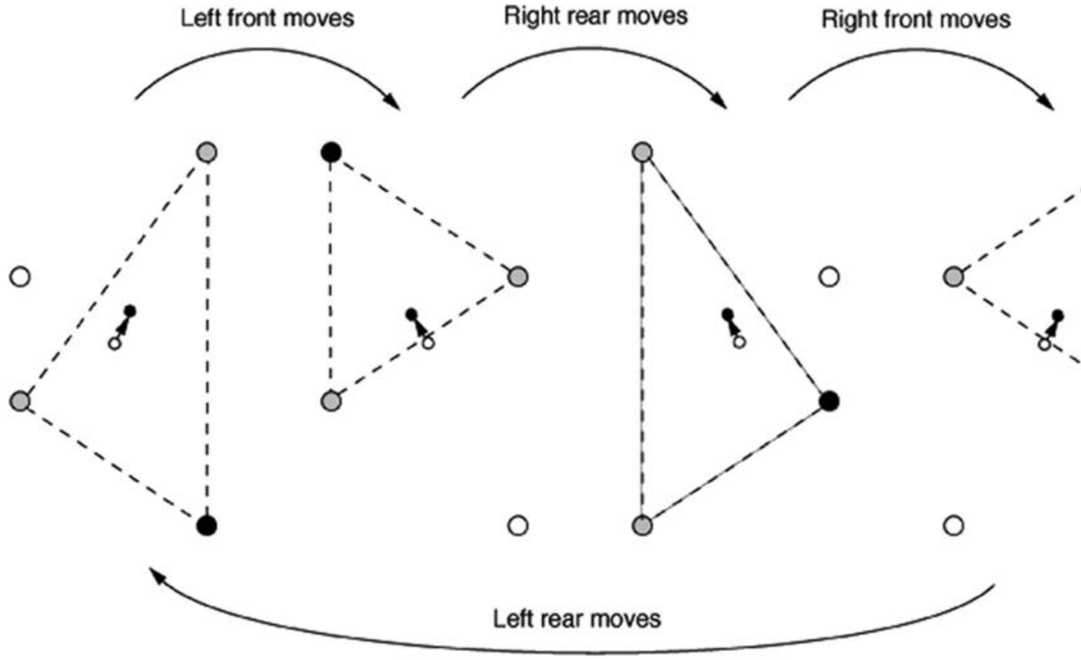

Figure 4. The normal stepping pattern for a quadrupedal walker. The most recent footfall is shown dark, the one just lifted is light. Arrows indicate shifts of the center of gravity needed to move it from the center of a standing stance to equivalent positions with respect to a line between diagonally opposite supporting legs.

Thus instability becomes more tolerable as an animal increases its stride frequency, especially if this can be done without much of a concomitant decrease in leg length. Put in practical terms, at low speeds and thus low stride frequencies, stable gaits work better; as speeds increase, unstable gaits become ever more practical and higher degrees of instability more tolerable. Dogs can tolerate a lot of instability when galloping but less when walking; turtles, low to the ground and making infrequent strides, should be much less tolerant of instability – although, as just described, they can still be unstable.

Alexander's (2003) approach parallels a suggestion I made (Vogel 2003) about the minimum speed for galloping, also based on maximum fraction of leg length that an animal can fall between foot-falls. I invoked the Froude number as a predictive variable; expression (3) amounts to the reciprocal of the Froude number (*Fr*) if animals swing their legs similarly, so speed is proportional to the product of stride frequency and leg length, *fh*:

$$Fr = \frac{v^2}{gh}\cdot \tag{4}$$

Six legs permit unconditional static stability. A hexapedal animal need only support itself on alternative triangles, and a duty factor of 0.5 is ample. Insects, paradigmatically hexapedal, use such a stable gait at low speeds but become increasingly unstable as they move faster (Ting *et al* 1994). (Of course not all insects walk on all six legs – for instance, praying mantises and mosquitoes use only four of their six.) At the highest speeds some, such as cockroaches and ants, have fully aerial phases (Full *et al* 2002) – as we do when running. Further increase in number of legs further reduces the maximum duty cycle consistent with stability, but six legs are the fewest paired legs that enjoy the option of fully stable walking without shifting centers of gravity. That has stimulated considerable interest in the way insects walk and run by designers of walking robots and robotic vehicles.

5. The stability of aircraft, living and non-living

An object standing or moving on the earth's solid surface faces two planes of potential instability, both vertical, resulting from the way its center of gravity lies above the point of contact with the substratum. For bilaterally symmetrical movers these planes are side-to-side and fore-and-aft. Alternatively, we can adopt anatomical practice and designate them transverse (or frontal to the human anatomist) and sagittal.

Devices moving through continuous media encounter in addition a horizontal plane of potential instability, frontal to the anatomist (except ones dealing with upright humans). For names we normally turn to the world of airplane design and performance, perhaps to avoid ambiguity about whether changes within any of the planes refer to changes in position (translation) or, as here, changes in orientation (rotation).

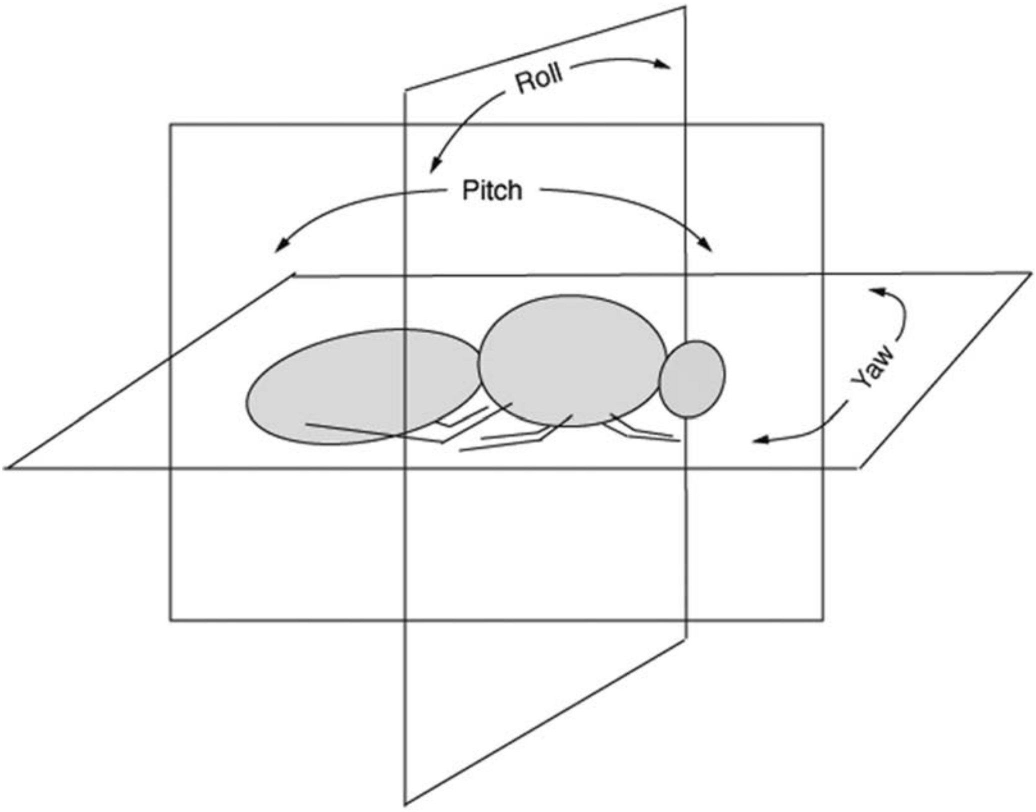

Figure 5. The three planes in which a flier can turn.

As in figure 5, that community defines the planes “roll” for the transverse or sideways-tipping, “pitch” for fore-and-aft rotation, and “yaw” for this last side-to-side swinging in a horizontal plane.

Dealing with three planes complicates both achieving directional control and analyzing how organisms might be managing just as it amplifies the importance of the matter. (In fact, perturbation about one axis commonly affects orientation about a different axis, introducing another, and a quite common, element of instability.) Putting aside degree of stability, a stable aircraft returns to its previous orientation after a perturbation without active adjustment of its controls. An unstable one either does not return or deviates even further – in whatever plane is at issue.

In a short article in *Evolution*, in 1952, John Maynard Smith drew attention to the considerable orientational instability of most extant flying animals. Maynard Smith, before taking a degree with the great evolutionary biologist J B S Haldane, had spent some time as an aeronautical engineer, so he brought a new perspective on flight to both paleontologists and physiologists. He compared present-day insects and birds and the mature pterosaurs to the earliest then-known flying ancestors of each. (He had inadequate information about early bats.) He suggested that, as active flight developed in each lineage, evolution produced ever less stable designs. In particular, control in pitch appeared, as with aircraft, to present the greatest challenge. Early forms tended to have greater development of horizontal surfaces on their rear portions – lateral plates on insects such as the dragonfly-like Paleodictyoptera, lateral membranes on the pterosaurs, and large tails on the birds – that stabilized flight much as rear feathers stabilize our arrows. He noted that while many extant birds have large tails, most deploy them mainly for take-offs and landings, when tails lower the stalling speed and thus facilitate operation without runways. (The remarkable pictures in Dalton 2001 provide graphic illustrations of just this.)

What permitted the instability was the concurrent evolution of ever more competent flight control, that is, of neuro-muscular systems. What drove the process was the inherent trade-off in flying machines between stability and maneuverability – a much less forgiving but essentially identical situation to that faced by legged locomotion on land. That contemporary birds are quite unstable renders them dangerously flawed models for human aircraft. Many early attempts to build airplanes foundered from inadequate appreciation of the fact. At least one case proved fatal, that of Otto Lilienthal, in 1896, author of the pioneering analysis, *Bird Flight as the Basis of Aviation* (1889), and pilot of what we would call hang-gliders. Others, notably Samuel

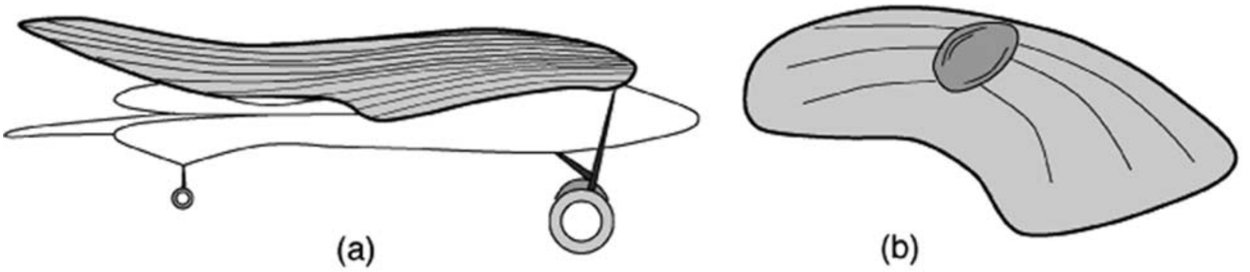

Figure 6. **(a)** The Etrichs' glider of 1906, traced from a photograph of a full-scale model at the Owl's Head Transportation Museum, Rockland, ME; struts and cables have been omitted. **(b)** The seed-leaf of *Alsomitra*.

Langley, took great pains to assure pitching stability – the most troublesome plane and the one that doomed Lilienthal. Langley's best full-scale airplane, which failed (at least) from structural weakness, achieved inherent pitch stability with tandem wings, one pair behind the other.

Unlike Lilienthal and others, the Wrights gave considerable attention both to stability and to control by adjustments of the aircraft rather than the position of the aviator (Culick and Jex 1987). Their later gliders and 1903 Flyer had sufficient stability to be safe and reliable in breezes but not so much as to compromise control. Indeed, the only fundamental change made thereafter consisted of substituting for the canard wing in front the now familiar pitch-controlling horizontal tail. Canard wings persist only in a few high-performance aircraft (usually with fast-acting computer control offsetting intrinsic instability) and possibly as the "hammerheads" on sharks in the genus *Sphyrna*. On the latter, they may (a good study is overdue) facilitate following the contour of the sea bottom or suddenly lunging downward. The "bow-rudder," a front paddle directed obliquely forward by white-water canoists approaching rocks, works the same way except in yaw.

Stable fliers do exist in nature, as one might anticipate, where active control remains out of the question. As with terrestrial stability, plants provide the exemplars. A descending, autogyrating samara of a maple or other tree must be able to recover from the perturbations of wind gusts or branches encountered on route. Non-autogyrating gliders, closer to airplanes or gliding birds, are rare. The most famous is the gliding seed-leaf of the Javanese cucumber *Alsomitra* (formerly *Zanonia*). As with so many phylogenetically odd animal gliders it lives only in the understory of the especially high dipterocarp forests of southeast Asia – perhaps because the nearly still air found there spans an unusually great height range. It provided a model for some early aircraft, those of the Etrichs, who built a series of *Zanonia*-winged craft, beginning with gliders (figure 6), in the first decade of the 20th century. Most likely as a reaction to Lilienthal's death, they wanted assurance of stability. Which they achieved, producing a glider that was almost unmaneuverable (Bishop 1961). Vincenti (1990) provides a fine historical perspective on the gradual appreciation of the issues involved.

They thereby inadvertently sacrificed straight-line performance. Putting that tradeoff in context involves the so-called lift-to-drag ratio of an airfoil, hydrofoil, or propeller blade. In effect, such a device generates lift (L), a force normal to its motion through the medium, at the price of an increase in drag (D), the force tending to retard its motion. The ratio not only represents a kind of efficiency, but it translates directly into the range a glider can go in still air. It does this by setting the "glide angle," θ, the angle with respect to the horizontal, at which a passive craft will descend:

$$\cot\theta = \frac{L}{D} \text{ or } \tan\theta = \frac{D}{L}\cdot \qquad (5)$$

Gliding thus maximizes distance by maximizing L/D, which varies somewhat with speed. It produces the greatest time aloft at a slightly lower speed and thus a slightly lower (but maximal for that speed) L/D.

The *Alsomitra* seed-leaf in nature develops a relatively poor lift-to-drag ratio, 3.7, almost certainly in the interest of stability. In a wind tunnel, its optimum (at the price of some stability) is still a mediocre 4.6, about half of what an insect wing under equivalent circumstances can reach. Its swept-back wings put the seed and center of mass lie ahead of the aerodynamic center (the point of action of the lift-drag vector), so it has the equivalent of the expected tail. As a result, the increased lift of any inadvertent upward pitch of the nose is far enough aft to offset rather than amplify the perturbation. And normal operation at an angle of attack well below the stall point gives extra margin for that stabilizing effect. (Azuma and Okuno 1987; Azuma 2006).

Birds, by contrast, fly as remarkably unstable craft – one cannot easily launch a stuffed bird on a smooth and fairly flat path. As a result, it has been argued (Harris 1989) that their use as models hampered the development of airplanes. Still, the degree of instability varies a lot; and, again, varies more-or-less with maneuverability. The *Alsomitra* seed-leaf accepted a lower-than-ideal lift-to-drag ratio in the interests of stability; birds make analogous departures from best performance in the interest of maneuverability. Other things equal, a longer, narrower wing gives a higher ratio than does a shorter, broader wing – again something glimpsed but not always appreciated by earlier aircraft designers. Why, then, do sea birds have particularly long and narrow wings while those that glide over land are shorter and wider, with splayed primary feathers instead of pointed tips? These terrestrial gliders apparently accept a lower best L/D – in effect a lower efficiency – and thus steeper minimum glide angles in order to be able to turn more sharply and thus fly in tighter circles. That should be important when gliding in an ascending thermal torus or an updraft over irregular terrain. Stabilizing against yawing – more important in the more erratic winds over land – has been suggested as another function of splayed primaries (Pennycuick 1975), but with little hard evidence of either that action or its utility.

Terrestrial birds also gain in ability to fly slowly without stalling – increasing the angle at which wing meets air to boost lift (which varies with the square of speed) at low speeds – and thus the ability to land and take off at near zero airspeed. (Norberg 1990 give a good review of the contrast.) Enhancing that low speed performance, terrestrial soarers have lower wing loading (W), the ratio of weight (mg), to wing area (S):

$$W = \frac{mg}{S} \approx \frac{L}{S} \propto \frac{v^2}{S}\,. \tag{6}$$

Weight to be supported translates into lift – roughly, since lift is conventionally defined as normal to flight direction rather than to the horizon. The lift required to offset body weight depends, again roughly, on the square of flying speed. So lower wing loading implies a lower weight-sustaining minimum flight speed.

Note that L/S in eqn 6 scales with linear dimensions – lift here is lift required to support weight, not lift as proportional to wing surface at constant speed. That scaling makes wing loading higher for a larger but otherwise similar flier. Such scaling underlies not only the non-similarity of fliers of different sizes but also the higher flying speeds of larger craft – with their associated diminution of maneuverability as well as their higher takeoff and landing airspeeds and higher weight-specific power requirements for sustained flapping flight. Thus the 70 kg extinct bird *Argentavis magnificens* had to soar and could live only in a region of steep slopes and high winds (Chatterjee *et al* 2007).

While equipped with nothing analogous to the splayed primaries of terrestrial soaring birds, the same trade-off has been recognized in bats by Aldridge (1986), Norberg and Rayner (1987), and Dietz *et al* 2006). Bats that fly through forests have shorter, broader wings, and they weigh less relative to wing area so they can fly more slowly – paying a price in power (in effect overall performance) for that maneuverability. By contrast, bats that fly in open areas have longer, narrower wings, more weight relative to wing area, and they fly faster and more economically.

Like birds and bats, flying insects are fairly unstable. Extant forms lack aerodynamic stabilizers such as tails or abdominal protrusions; in any case these latter would be of limited service during hovering. Hind legs sticking out into the airstreams of the two wings seem to give some ruddering action in some forms, and mobile abdomens (as in many wasps) provide at least the possibility of adjusting centers of mass and drag. As Dudley (2000) notes, the way the wing stroke centers above the body, with wings almost (or actually) touching at the top but not the bottom of the stroke, provides some degree of stability in both roll and yaw. (The torque induced by a center of lift and thrust above the center of drag should matter little for a flier that can control the direction of the resultant of lift and thrust.) And the elongation of bodies fore and aft – heads and abdomens, the latter sometimes quite long – should give a bit of pitch stabilization, at least for transient perturbations.

The issue of stability may bear on the origin of flight as well as on the subsequent evolution of flying lineages. As pointed out by Ellington (1991) (and summarized by Dudley 2000), a long, circular cylinder, held obliquely, can descend at glide angles as low as 40º, no worse than some non-flying but gliding vertebrates. Instability, though, wrecks the scheme at all but Reynolds numbers (length times speed, in effect) still lower than those of flying insects. The difficulty traces to an aerodynamic center in front of the center of gravity, meaning that any upward pitch will be magnified until, at equilibrium, the cylinder will descend vertically while oriented horizontally, parachuting rather than gliding. Small winglets protruding from the sides – minor cuticular extensions for insects – and located toward the rear fix the problem, making the cylinder stable in pitch. (Moving the center of gravity forward, where it is in extant insects can help also.) Yawing stability can be achieved by adding a caudal filament, and roll stability as well comes if a bilateral pair of diagonally rearward-pointing cerci replace the filament. Of course overdoing the appliances on the rear end produces the equivalent of a fletched arrow, which will both descend and be oriented vertically and thus have the maximal drop speed and gain no horizontal distance at all.

In effect, a flier picks some combination of three variables, stability, maneuverabilility, and performance, each in practice multidimensional. As we have seen, both the exceptional stability of the *Zanonia* seed-leaf and the maneuverability of terrestrial soarers come at a price in level-flight performance. In general, an increase in any one of the three variables extracts a price with respect to one or both of the others. None, though, lend themselves to definitions that combine precision with practicality, and designers face no definitional limit to what a particular combination can do. Birds and bats, as D E Alexander (2002) points out, may be less constrained in that choice than insects and human aircraft since they can vary wing geometry quickly and extensively.

6. The inputs for aircraft

Only with adequate control systems can instability be tolerated, much less capitalized upon. Proprioceptive feedback loops, gravity and acceleration detectors, and associated anticipatory and dynamic devices permit walking, running, climbing, and all manner of terrestrial acrobatics. However complex these tasks, they pale before those of control in continuous media. And the less stable the flier or swimmer, the greater its dependence on control. Moreover, even a stable flier faces at least two problems practically unknown in land-based locomotion. First, air and water cannot be relied upon to remain at rest with respect to the earth, that is, to their underlying solid substrata. Thus sensing what goes on in an animal's immediate vicinity may limit clues to its overall motion. The slower the animal, the lower its speed through the local medium relative to motion of the medium itself. So the problem must bedevil organisms more than it does our boats and aircraft. And second, heading provides only a limited cue about course. Face north and move forward at 1 m s^{-1} in a wind or current from the east at the same modest speed, and you actually progress northwest at 1.4 m s^{-1}.

Small airplanes (excluding high-performance military craft) are, by design, about as stable as they can be without overly compromising their ability to make necessary changes of direction. The new pilot learns, in the words of Molly Bernheim (1959) that when things go awry, "Let go! The airplane can look after itself better, now, than you can do! Turn it loose! Then, and only then, you may guide it gently where you want it to go." Even so, airspeed and ground speed may be quite different, and heading may not equal course. Moreover, our land-based sensory equipment can mislead us. For instance, semicircular canals cannot reliably separate gravitational from angular accelerations. So a banked turn feels no different from straight flight, and the pilot must read the instruments, not the receptors medial to the seat of the pants.

How then do animals get the sensory input critical for active control? Visual signals provide widely used references for both orientation and location, and both birds and insects typically have large brain areas devoted to processing visual input. The horizontal cells of bird retinas have been know at least since the work of the great neuroanatomist, Ramón y Cajal, a century ago; they purportedly select horizontal lines for attention in a process of horizon-detection that would distinguish level flight from banked turns. Were night-flying birds, insects, and, of course, bats not so common and accomplished, one might declare visual input essential – which it certainly cannot be. Equipping an enclosed volume of air to work as an altimeter should be simple enough; a human acquires a pair that can be painfully effective on airplane flight with plugged Eustachian tubes. (And swim bladders can signal depth changes in fishes.)

The flight motors of insects, on which work has been extensive, have at least three additional sources of sensory input. Bending of antennae and setae equipped with mechanoreceptors at their bases provide information about local airflow, including, most likely, the airflow on each side caused by the wings themselves. Additional mechanoreceptors on the wings and in cuticular structures adjacent to them can not only signal what the wings are doing but, because of the non-rigidity of all the relevant structures, can provide feedback on the loads the wings encounter. Since oscillating wings act as gyrosensitive devices, the receptors should receive dynamic inputs. Several groups of insects, most notably the true flies (Diptera) have developed gyrosensitivity further, converting one pair of wings (the hindwings in Diptera) into stalks with knobs on their ends that still oscillate as if wings. Finally, the flexible connections between thoraces, housing the flight motors, and both heads and abdomens, permit the latter to provide inertial information – as, for instance, when a turn is initiated either volitionally or inadvertently. Dudley (2000) reviews what has been done in an active area that at this point seems to lack a good synthesis.

The principal difficulty for both flying animals and our understanding of how they manage, is the lack of an obvious source of earth-referenced data. Vision cannot form the sole such source, but sensing cuticular deformation can do little to augment it, particularly in a domain in which ambient winds rival or even exceed flight speeds and in which the variations that we lump as local turbulence cannot be easily averaged over time or space. People investigating bird migration – which I want to skip over here – have wrestled with the problem for many years. The same problem for both animal and biologist afflicts fishes that hold position in murky, moving water (Howland and Howland 1962).

7. Swimming

Swimming while fully submerged, as is usual among animals, involves the same general trade-offs and opportunities as does flying. Still, as a more forgiving locomotory mode, it affords a wider range of designs and solutions. Per unit distance covered it costs less than flying or any terrestrial gait. Per unit time – since hovering costs next to nothing – it costs far less even than any form of active flying. Only soaring, essentially gravity-sustained, can come close. While economical, though, the same cost-benefit calculation applies to both predators and prey; and suspension feeders that actively swim (or pump) face the outcome of success by all other and previous suspension feeders. So once again we encounter systems balancing stability, maneuverability, and efficient performance. Again, maneuverability typically comes at a price in reduced locomotor efficiency.

Most swimming vertebrates do have a particular non-locomotory instability with little parallel among fliers. Achieving buoyancy with a non-rigid gas-filled container makes them unstable with respect to depth. Thus increased depth compresses the gas, reduces buoyancy, and impels an animal to go still deeper. The problem and its various solutions formed a large part of essay 8 (Vogel 2006b).

That gas bladder contributes to a more subtle problem. If the center of the water displaced by a submerged object lies below the center of gravity, then the object would prefer to be inverted. With the usual convention, the vertical distance between the centers, the so called "metacentric height," is negative. Although the bladder has moved from its ancestral ventral location (like lungs) to a position above the gut, bony fish with swimbladders usually still have negative metacentric heights, giving them unstable normal postures – a dead fish ordinarily goes belly up. Fish that live in moving streams seem to be the most unstable, perhaps because they have to expend power continuously to hold position anyway. Fish that live in still water and do not swim continuously tend to be less unstable. In addition to making "resting" easier, a stationary fish cannot trim a hydrodynamic surface to adjust position and thus lacks that mode of active stabilization (Webb 2002.)

In general, both fish and cetaceans have considerable instability as a result of their motion through water. The situation resembles what we saw for flying animals in that stability and maneuverability are again to a considerable extent antithetical. That speed trades off against maneuverability appears even clearer for the swimmers than for flyers. And stiff bodied forms tend to be both faster and less maneuverable than flexible ones, whether one compares cetaceans (Fish 2002) or fishes (Webb 2002).

Swimmers most often propel themselves with driving devices at their downstream ends, whether the fins of fishes, the flukes of cetaceans or the jets of cephalopods. (But not always – some fish use pectoral fins or opercular jets while penguins, some other birds, and sea lions, to mention a few, use modified forelimbs.) Pushing from behind rather than pulling from in front generates an additional mode of instability. Any inadvertent yaw means that one component of the output of the propulsor will give rise to a turning moment that will amplify the initial yaw. Still, it does not appear to create any noteworthy difficulty despite the analogous instability of rear-propelled, rear-heavy automobiles and the instability of pushes that (except in jet-propelled forms) alternate from one side of the body to the other.

Another form of instability comes from trimming controls in front of centers of gravity. Hammerhead sharks, as noted earlier, may use the hammerhead as a canard wing, a seriously destabilizing device, to gain maneuverability that helps follow at a fixed distance above a non-level substratum. And skates and rays may be doing the same with their relatively anterior "wings" – many species do swim just above the substratum, and all of these elasmobranchs have subterminal rather than (as in bony fishes) terminal mouths. Less extreme are ordinary pectoral fins and flippers, but these nonetheless commonly lie ahead of the centers of both buoyancy and gravity in both fishes and cetaceans (Fish 2002).

In both fliers and swimmers, large forms that chase small forms for food tend to be somewhat slower, less stable, and more maneuverable than closely related large forms that have other modes of feeding. The large predator retains a speed advantage but must offset its intrinsically worse minimum turning radius and maximum turning rate. That trade-off appears strikingly in interspecific comparisons among toothed whales (Woodward *et al* 2006). Humpback whales, notably acrobatic baleen whales, have unusually large pectoral flippers and are not especially rapid swimmers (Nowak 1991) by large-whale standards. In dragonflies, large aerial predators, the motor drives the wings directly rather than indirectly, as in most other insects, by which they gain an unusual degree of independence of control of their four wings (D E Alexander 1986).

Swimming at the surface – partly submerged – encounters a particularly nasty form of the problem of negative metacentric heights. Floating high in the water virtually guarantees a highly negative value unless the craft contains a lot of ballast located deep in the hull. But the problem can be ameliorated in a way unavailable to the submerged swimmer. Most ships have V-shaped or U-shaped hulls. This geometry requires that the center of gravity of the craft be lifted if it rolls either way from exactly upright, supplying a restoring torque. Counterintuitively, perhaps, a broad, flat bottom with sides that then slope inward yields a craft lacking that region of stability. The few animals that swim at the surface of bodies of water usually do have hulls that taper downward – it is especially conspicuous in young sea turtles (Wyneken, personal communication).

8. The "flights" of non-flyers

All fliers, in the end, achieve stability with aerodynamic devices – adjustable wings, deployable tails, and so forth. Once equipped with propulsive or gliding appendages, the output side of control needs only modest augmentation. But what can a non-flier do if it finds itself in midair and has some incentive to land in a specific orientation? While hydrodynamic control can be effected by fairly small structures, only the smallest non-fliers can press ordinary appendages into effective aerodynamic service. Thus some ants that live high in tropical forests make stable flying jumps at respectable glide angles (moving abdomen first, incidentally) by orienting their legs to provide aerodynamic stabilization (Yanoviak *et al* 2005). Larger ones must play with angular acceleration and velocity in a world that awkwardly insists that angular momentum be conserved.

Just as the product of mass and linear velocity gives ordinary momentum, the product of moment of inertia, I, and angular velocity, ω, equals angular momentum, H:

$$H = I\omega. \tag{7}$$

And moment of inertia, essentially the second moment of mass, is the summed elements of mass, m, times, for each element, the square of its distance, r, from the axis of rotation:

$$I = \Sigma mr^2 = \int r^2 dm. \tag{8}$$

Constancy of angular momentum means that angular velocity can be changed only with a concomitant inverse change in moment of inertia. In a world that also believes that mass must be conserved, changing moment of inertia depends on changing the effective overall r, the "radius of gyration."

Increasing angular velocity by decreasing radius of gyration – a figure skater or ballet dancer does that by drawing arms and legs closer to the torso and thus to the axis of rotation. Alternatively, an external contact can impose a moment that imparts angular velocity. A springboard diver can do that in at least two ways (Frolich 1979). A run out along the board gives the whole body a translatory velocity, but the jump at the end, besides imparting an upward component to velocity, slows the lower part of the body (figure 7a). So the diver takes off with some angular velocity and angular momentum. Tucking in legs and arms in midair then increases angular velocity; with (as conservation requires) no change in angular momentum, a somersault ensues. Additionally (or with no run-up), the diver may jump with the body tilted forward so the resulting torque of the vertical push imparts an initial angular velocity (figure 7b). Either way, that angular velocity must be minimized again before entering the water by extending arms and legs.

It has sometimes been contended that conservation of angular momentum requires that any mid-air rotation must trace back to an initial, visually subtle, angular velocity. If by "rotation" one implies time-averaged net angular momentum, then the contention (absent aerodynamic or Coriolis effects) must be correct. Less self-evidently, changes in body orientation – rotation about any axis – can be effected without such net change in angular momentum of the body as a whole. Such orientational changes have been unequivocally demonstrated in springboard divers, acrobats, trampoline jumpers, space walkers, and falling cats (as well as some other mammals). Frolich (1979) and Edwards (1986) provide good descriptions and analyses of how it can be done; Brancazio (1984) gives a quick summary, and Stewart (2004) shows a nice set of color photographs of a falling cat.

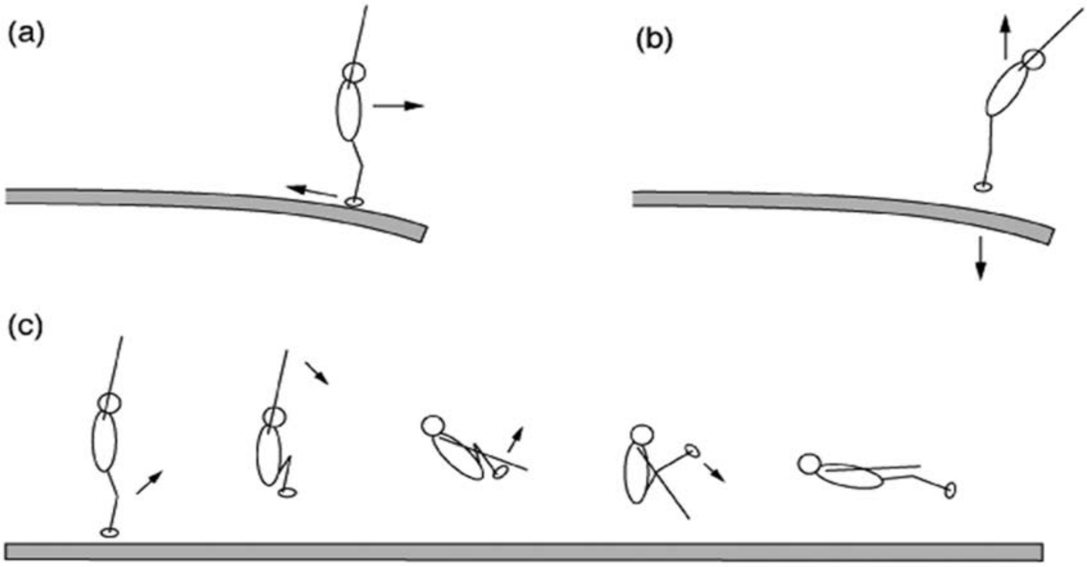

Figure 7. **(a)** Turning during a springboard dive by making the feet lag the torso. **(b)** Turning by taking off with head and torso forward of the final contact point. **(c)** The sequence in which appendages are moved to perform a zero-angular-momentum back quarter somersault, as described by Frolich (1979).

Figure 7c, adapted from Frolich (1979), shows stages of a reorientation about the body's long axis (in pitch) by a diver or gymnast. Initially the body extends full length. Tucking legs up close to the torso reduces the moment of inertia – not in itself inducing much orientational change, but amplifying the subsequent change. Swinging the arms forward and down against the body and thus giving them angular velocity and momentum induces an equal and opposite angular momentum and (of course lesser) velocity of the rest of the body. When the arm motion stops, the motion of the rest of the body has to stop. But the body has shifted orientation substantially – up to about 80º. The arms may then be reextended by moving them along the body axis or swinging them outward to restore the original extended posture. That moment of inertia is a second, rather than a first, moment – incorporating r^2 rather than r – makes even fairly light appendages quite useful, as long as they are reasonably long.

Note a characteristic – and diagnostic – difference between turns that take advantage of initial angular momentum and those that manage despite zero angular momentum. In the first, angular momentum and thus angular velocity remain at the end, so the body keeps rotating unless stopped by some external agency. In the second, the body, having no overall angular momentum, stops as soon as movement of some parts relative to others terminates. One can experience both in a swivel chair. Initial angular momentum just takes push against floor or desk, what one ordinarily does in such a chair. But one can turn, say, counterclockwise, by extending the arms, swinging both clockwise, drawing them back against the torso while moving them counterclockwise, extending them again, and repeating – each time progressing a few degrees. Holding weights in the hands increases the effectiveness of the maneuver. (As an exercise, the reader might now attempt to explain how a child can put a swing in motion.)

Domestic cats, famously able to land on their feet, do just such zero angular momentum turns as they fall. Reportedly, cats can turn 180º around their long axes during a 1-m fall, which takes less than half a second. Peak head acceleration (where the turning begins) has been reported to exceed 120,000º s^{-2} (O'Leary and Ravasio 1994). Tailless cats tail tailed ones in tests. According to Kane and Scher (1969) and Edwards (1986), and as in figure 8, the supine cat begins by arching the back so the whole animal is concave upward. It then twists the body about the vertebral column, beginning with the head, while maintaining that downward concavity, until the whole torso faces downward, and then straightens the back again, halting rotation.

Dogs, less limber, are less adept at righting; in one informal test, a dachshund failed completely, taking umbrage at the imposition. Rats and many other small mammals, though, right themselves quite competently. The behavior not only ensures landing on properly shock-mounted appendages to lower deceleration, but it also must increase drag during long descents, reducing both terminal velocity and the rate of approach to terminal velocity.

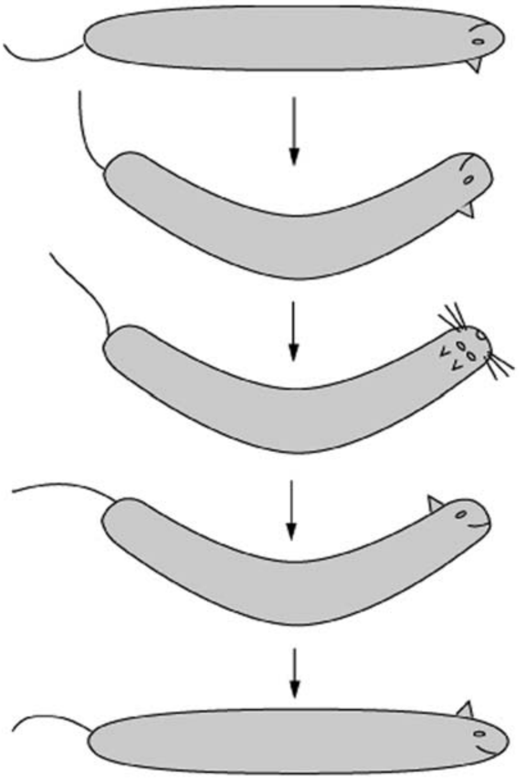

Figure 8. Righting of a mammal's torso with a zero-angular-momentum twist, as explained by Edwards (1986). In practice the twist begins at the head, the tail counterrotates, and movement of the legs (in a manner analogous to that shown in figure 7c) plays at least a supporting role.

Spinner dolphins (*Stenella longirostris*) make spectacular upward leaps, rotating while airborne as many as seven times about the long axes of their bodies. But, by contrast with cats, they make no significant use of zero angular momentum turning; instead, they drive their aerial turning by asymmetrical motion of their flukes just prior to emersion. The behaviour appears to cause dislodgment of remoras as they reenter the water (Fish *et al* 2006). One might expect that the high drag an animal experiences in water renders inertial turning both ineffective and unnecessary. But we should not dismiss the possibility out of hand. Photographs of the so-called pinwheeling maneuver of bottlenose dolphins (*Tursiops truncatus*) (Maresh *et al* 2004) look to my eyes strikingly similar to photographs and diagrams of righting cats.

We know less about non-mammalian cases of righting with zero angular momentum turns. An unfledged bird falling from a nests probably has sufficient plumage to keep it from reach a hazardous terminal velocity – the speed at which drag, speed-dependent, reaches weight. And most of the non-flying, non-mammalian gliders such as flying frogs and lizards, can exert aerodynamic control. An exception might be flying snakes (genus *Chrysopelea*), which while gliding downward do quite a lot of mid-air writhing and maneuvering (Socha *et al* 2005).

The relative utility of inertial and aerodynamic mid-air turning depends on both airspeed and body size. Faster motion favors reliance on aerodynamics, with both lift and drag increasing with something close to the square of speed. Larger size favors inertial turning due to the concomitant reduction in surface-to-volume ratio. Humans can, as we have seen, do quite well at inertial turning. But aerodynamic effects should not be casually dismissed since large animals fall somewhat more rapidly than small ones. Moreover, significant use of inertial turning has recently been demonstrated (along with aerodynamic turning) in flying birds (Hedrich *et al* 2007). Most birds perform downstrokes with extended wings and upstrokes with somewhat flexed wings. Flying straight generates no overall difference in moment of inertia since the two wings cancel each other's asymmetry. But when turning, the outer wing increases amplitude, which will roll that side upward, aiding the extra aerodynamic lift but without extracting a price in drag.

One odd convergence in small mammals suggests routine use of aerodynamic turning. In at least four lineages, long-hind-legged jumping animals have long tails (longer than head + body) with tufts of hair on their ends. Three of these are rodents – kangaroo rats (*Dipodomys*; Heteromyidae), some gerbils and jirds (*Gerbillurus*, *Meriones*; Muridae); and jerboas (*Dipus*, *Jaculus*, etc; Dipodidae). One is a marsupial – the kultarr (*Antechinomys*). Comparably long-tailed animals without long hind legs typically lack such terminal tufts, judging from the photographs in Nowak (1991). Movies of kangaroo rats engaged in intraspecific interactions show extended tails flung vigorously in all directions, with the tufts clearly visible. (*See*, for instance Disney 1953.) Still photographs taken under comparable circumstances often show erected tail hairs (Schmidt-Nielsen, personal observation). Perhaps the tuft, especially when erected, increases the drag of a tail moved laterally, and drag so far from the body provides torque that aids turning of the body in the other direction. Thus aerodynamic and inertial devices might combine in aiding mid-air maneuvering – but I am not aware that the phenomenon has been investigated.

Acknowledgments

Young-Hui Chang, Alphonse Masi, John Mercer, Knut Schmidt-Nielsen, Jake Socha, and Jeanette Wyneken provided sources worth consulting and other useful information. Mary Henderson kindly supplied figure 2b.

References

Aldridge H 1986 Manoeuvrability and ecology in British bats; in *Proc. 7th Int. Bat Res. Conf. Myotis Mitteilungsbl. Fledermauskundler* (eds.) P A Racey and H Roer, pp 157–160

Alexander D E 1989 Wind tunnel studies of turns by flying dragonflies; *J. Exp. Biol.* **122** 81-98

Alexander D E 2002 *Nature's flyers: birds, insects, and the biomechanics of flight* (Baltimore: Johns Hopkins University Press)

Alexander R McN 2003 *Principles of animal locomotion* (Princeton, NJ: Princeton University Press)

Azuma A 2006 *The biokinetics of flying and swimmin,* 2nd edition (Reston, VA: American Institute of Aeronautics and Astronautics)

Azuma A and Okuno Y 1987 Flight of a samara, *Alsomitra macrocarpa*; *J. Theor. Biol.* **129** 263–274

Bernheim, M 1959 *A sky of my own* (New York: Rinehart and Co.)

Biewener A A 2003 *Animal locomotion* (Oxford: Oxford University Press)

Bishop W 1961 The development of tailless aircraft and flying wings; *J. Roy. Aero. Soc.* **65** 799–806

Brancazio P J 1984 *Sport science: physical laws and optimum performance* (New York: Simon and Schuster)

Chatterjee S, Templin R J and Campbell K E Jr 2007 The aerodynamics of *Argentavis*, the world's largest flying bird from the Miocene of Argentina; *Proc. Nat. Acad. Sci USA* **104** 12398-12403

Coleman M J and Ruina A 1998 An uncontrolled walking toy that cannot stand still; *Phys. Rev. Lett.* **80** 3658-3661

Collins S, Ruina A, Tedrake R and Wisse M 2005 Efficient bipedal robots based on passive-dynamic walkers; *Science* **307** 1082–1085

Cotterell B and Kamminga J 1990 *Mechanics of pre-industrial technology* (Cambridge: Cambridge University Press)

Culick F E C and Jex H R 1987 Aerodynamics, stability, and control of the 1903 Wright Flyer; in *The Wright Flyer: an engineering perspective*; (ed.) H S Wolko (Washington: Smithsonian Institution) pp 19–43

Dalton S 2001 *The miracle of flight* (London: Merrell Publishers Ltd)

Dietz C, Dietz I and Siemers B M 2006 Wing measurement variation in the five European horseshoe bat species (Chiroptera: Rhinolophidae); *J. Mammal.* **87** 1241–1251

Disney W 1953 *The living desert* (Vol. 2 in *Walt Disney's legacy collection*, DVD 2006)

Dudley R 2000 *The biomechanics of insect flight* (Princeton, NJ: Princeton University Press)

Edwards M H 1986 Zero angular momentum turns; *Am. J. Phys.* **54** 846–847

Ellington C P 1991 Aerodynamics and the origin of insect flight; *Adv. Insect Physiol.* **23** 171–210

Fish F E 2002 Balancing requirements for stability and maneuverability in cetaceans; *Integ. Comp. Biol.* **42** 85–93

Fish F E and Full R J 2002 Stability and maneuverability; *Integ. Comp. Biol.* **42** 85–181

Fish F E, Nicastro A J and Weihs D 2006 Dynamics of the aerial maneuvers of spinner dolphins; *J. Exp. Biol.* **209** 590–598

Frolich C 1979 Do springboard divers violate angular momentum conservation?; *Am. J. Phys*. **47** 583–592

Full R J, Kubow T, Schmitt J, Holmes P and Koditschek D 2002 Quantifying dynamic stability and maneuverability of terrestrial vertebrates; *Integ. Comp. Biol*. **42** 149–157

Griffin T M and Kram R 2000 Penguin waddling is not wasteful; *Nature (London)* **408** 929

Harris J S 1989 An airplane is not a bird; *Am. Heritage nvention Technol.* **5** 19–22

Hasan Z 2005 The human motor control system's response to mechanical perturbation: should it, can it, and does it ensure stability? *J. Motor Behav.* **37** 484–493

Hedrich T L, Usherwood J R and Biewener A A 2007 Low speed maneuvering flight of the rose-breasted cockatoo (*Eolophus roseicapillus*). II. Inertial and aerodynamic reorientation; *J. Exp. Biol.* **210** 1912–1934

Howland H C and Howland B 1962 The reaction of blinded goldfish to rotation in a centrifuge; *J. Exp. Biol*. **39** 491–502

Jayes A S and Alexander R McN 1980 The gaits of chelonians: walking techniques for very low speeds; *J. Zool. London* **191** 353–378

Kane T R and Scher M P 1969 A dynamical explanation of the falling cat phenomenon; *Int. J. Solids Struct.* **5** 663–670

Lilienthal O 1889 *Der Vogelflug als Grundlage der Fliegekunst*. Trans. 1910 as *Bird flight as the basis for aviation* (London: Longmans, Green)

Lillie M and Boechler T 2005 (formally unpublished and untitled; see *http://web.archive.org/web/20050320145500/http://www.zoology.ubc.ca/~biol438/Reports/CowTip.PDF*)

Maresh J L, Fish F E, Nowacek D P, Nowacek S M, and Wells R S 2004 High performance turning capabilities during foraging by bottlenose dolphins (Tursiops truncatus); *Mar. Mammal Sci*. **203** 498–509

Maynard Smith J 1952 The importance of the nervous system in the evolution of animal flight; *Evolution* **6** 127–129

Norberg U M 1990 *Vertebrate flight* (Berlin: Springer-Verlag)

Norberg U M and Rayner J M V 1987 Ecological morphology and flight in bats (Mammalia: Chiroptera): wing adaptations, flight performance, foraging strategy and echolocation; *Phil. Trans. R. Soc. London* **B316** 335–427

Nowak R M 1991 *Walker's mammals of the world,* 5th edition (Baltimore, MD: Johns Hopkins University Press)

*O'Leary D P and Ravasio M J 1994 Simulation of vestibular semicircular canal responses during righting maneuvers of a freely falling cat; *Biol. Cybernetics* **50** 1–7

Pennycuick C J 1975 Mechanics of flight; in *Avian biology,* (eds) D S Farner and J R King (New York: Academic Press) vol 5, pp 1–75

Petroski H 1990 *The pencil: a history of design and circumstance* (New York: Knopf)

Sewell A 1877 *Black Beauty* (London: Jarrold and Sons)

Sherrington C 1906 *The integrative action of the nervous system* (New Haven CT: Yale University Press; 2nd edition 1947)

Socha J J, O'Dempsey T and LaBarbera, M 2005. A three-dimensional kinematic analysis of gliding in a flying snake, *Chrysopelea paradisi*; *J. Exp. Biol.* **208** 1817–1833

Stewart I (2004) Quantizing the classical cat; *Nature (London)* **430** 731–732

Ting L H, Blickhan R and Full R J 1994 Dynamic and static stability in hexapedal runners; *J. Exp. Biol*. **197** 251–269

Vincenti W G 1990 *What engineers know and how they know it: analytical studies from aeronautical history* (Baltimore, MD: Johns Hopkins University Press)

Vogel S 2003 *Comparative biomechanics: life's physical world* (Princeton, NJ: Princeton University Press)

Vogel S 2006a Life in a physical world. VII. Gravity and life on the ground; *J. Biosci.* **31** 202–214

Vogel S 2006b Life in a physical world. VIII. Gravity and life in water; *J. Biosci*. **31** 309–322

Webb P W 2002 Control of posture, depth, and swimming trajectories of fishes; *Integ. Comp. Biol.* **42** 94–101

Woodward B L, Winn J P and Fish F E 2006 Morphological specializations of baleen whales associated with hydrodynamic performance and ecological niche; *J. Morphol*. **267** 1284–1294

Yanoviak S P, Dudley R and Kaspari M 2005 Directed aerial descent in canopy ants; *Nature (London)* **433** 624–626

*e*Publication: 18 August 2007