Chapter 14 Multi-objective optimality in biology

14.1 Introduction

So far we considered evolution towards a single objective, such as maximizing the growth rate of bacteria. A single objective is appropriate for carefully controlled experiments. In nature, however, biological systems usually need to carry out multiple tasks. Bacteria, for example, need to grow quickly and also to survive stresses. Multiple tasks lead to a fundamental **tradeoff**: no design can be optimal at all tasks at once. There is no animal that can fly like an eagle, swim like a dolphin and run like a cheetah.

In this chapter, we will ask how evolution can optimize in the presence of multiple tasks. This is the art of optimal tradeoffs. We will see that multiple tasks lead to simple geometrical patterns in biological data. These patterns can help us to understand the evolutionary tradeoffs at play.

14.2 The fitness landscape picture for a single task

Let's start with the classical framework for evolutionary theory, the fitness landscape picture. Consider the evolution of a bird's beak. The **genotype**, DNA, leads to the **phenotype**, the shape of the beak, which leads to fitness, by eating seeds. The better the beak is at eating seeds, the more the bird will have viable chicks and grand-chicks on average, and the higher its fitness. It will pass its genes to the next generation. Thus

$$genotype \rightarrow phenotype \rightarrow fitness$$

Now suppose we take a ruler and measure the beak length, width, depth, curvature and so on. These are called beak **traits**, T_i. Each beak phenotype can be represented as a point in a space whose axes are the traits, called **trait space**. We can (in principle) plot the fitness of each beak in trait space, $F(\vec{T})$, where $\vec{T}$ is the vector of traits, resulting in a **fitness landscape.** The fitness landscape is a multi-dimensional version of the fitness function from the previous chapter. The fitness landscape can have hills and valleys. Figure 14.1 shows the contours of a fitness landscape as a function of two traits. Natural selection will tend to converge to the summit of the fitness landscape, to the phenotype that maximizes fitness. Phenotypes will perhaps form a cloud around the peak due to randomizing forces.

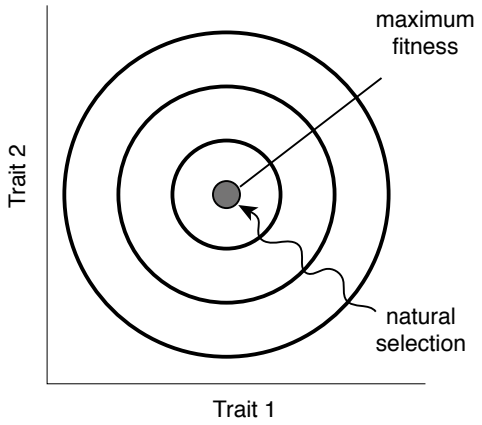

Figure 14.1

14.3 Multiple tasks are characterized by performance functions

But what if the beak needs to do two different tasks that both contribute to fitness: to crack seeds and to pick pollen from flowers? You can't be optimal at two tasks with one beak. Cracking seeds require a beak shaped like a pliers, whereas picking pollen requires a beak shaped like a pincers (Fig 14.2). In this case we need to modify the $genotype \rightarrow phenotype \rightarrow fitness$ picture, and add in the notion of **performances** at the two tasks (Arnold, 1983). The genotype determines the traits of the phenotype, $\vec{T}$, which determine performance at task 1 (cracking seeds) $P_1(\vec{T})$ and performance at task 2 (picking pollen) $P_2(\vec{T})$. Fitness is a function of these two performance functions: $F = F(P_1(\vec{T}), P_2(\vec{T}))$

Figure 14.2

$$genotype \rightarrow phenotype \begin{matrix} \nearrow \\ \searrow \end{matrix} \begin{matrix} performance\ at\ task\ 1 \\ performance\ at\ task\ 2 \end{matrix} \begin{matrix} \searrow \\ \nearrow \end{matrix} fitness$$

Notably, fitness is an **increasing** function of the two performances ($\frac{dF}{dP_i} > 0$): make a beak better at both tasks and fitness is sure to increase.

The precise form of the fitness function depends on the **niche**. In some niches one task is more important than the other, so that $F(P_1, P_2)$ gives more weight to that tasks. In other niches the other task is more important. The precise shape of the fitness function F in each niche is usually not known. But, as we will see below, conclusions can be reached that do not depend on knowing the form of F.

14.4 Pareto optimality in performance space

So which beak shapes will evolve under the tradeoff between these two tasks? Engineers routinely need to solve this type of problem. They use an approach called **Pareto optimality**. Suppose that you want to design a car. The design specifications require performance at two tasks, say acceleration (time from 0 to 100km/h) and fuel economy (say Km/liter). You take all possible designs, and plot them according to their performance at the two tasks. In this plot, whose axes are the two performances, each design is a point in **performance space**.

Now consider a design B. If there exists a design A that is better than B at both tasks (has higher performance at both tasks), we erase design B (Fig 14.3). We say that B is dominated by A. Erasing in this way all points which are dominated by another point, we remain with the **Pareto front** (Fig 14.4)**.** It is the set of designs that cannot be simultaneously improved at both tasks. This front is what engineers care about.

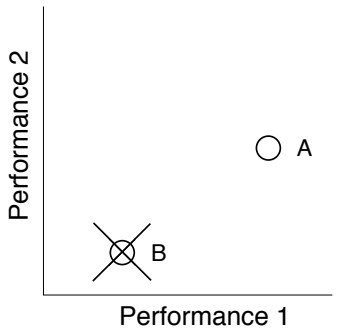

Figure 14.3

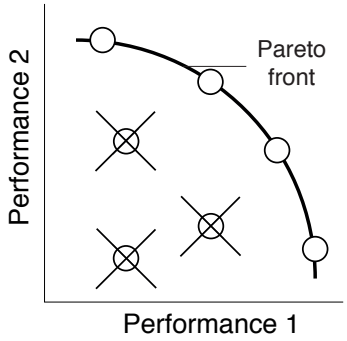

Figure 14.4

Which design from the front you choose is based on the market niche of the car: a family car requires better economy at expense of acceleration, and a sports car requires better acceleration at the expense of economy (Fig 14.5).

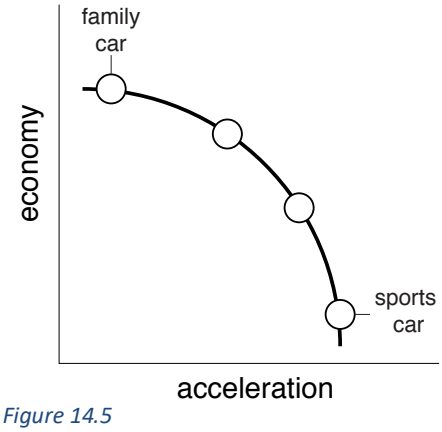

Figure 14.5

The Pareto optimality idea can also be used, in a lighthearted way, to help us understand how to choose scientific problems ((Alon, 2009)). The axes are how feasible and interesting the problem is.

14.5 Pareto optimality in trait space leads to simple patterns

In biology, unlike engineering, *we usually don't know what the tasks are in advance*. We can make an educated guess, but we can't be sure. Thus, we can't directly use performance space to do Pareto optimality, because we don't know what tasks to compare, and, even if we did, we usually cannot evaluate the performance of each phenotype at each task.

Remarkably, we can still make progress. We simply plot the data in *trait space*, using all the traits that we can measure directly. The axes are the traits, and each phenotype is a point in this space. For example, each beak is a point in a space of traits such as beak width, depth, curvature and so on. We will now see that evolution under several tasks makes the data show particular geometric shapes. These shapes can help us discern the number of tasks, and even what the tasks might be. For example, when there are two tasks at play, the data will fall on a line segment (or sometimes on a slightly curved segment as discussed below). The two ends of the segment give us clues about what the tasks are.

To see where this line-segment geometry comes from, let's imagine that each of the two tasks has a performance function, $P_1(\vec{T})$ and $P_2(\vec{T})$. The contours of these performance functions are plotted in trait space in Figure 14.6. The peak of each performance function is a special phenotype, called the **archetype**- it is the phenotype (trait combination) that is best at the task. If there was only that one task, evolution would converge to the archetype. Archetype 1 is the best beak for

seeds, and archetype 2 is the best beak for pollen. Performance drops with distance from the archetype.

We want to find the beak shape that maximizes fitness, $F(P_1, P_2)$, where F can be any increasing function. The surprise is that, no matter what F is, one can prove that, under certain assumptions, the optimal solution must fall on the *line segment that connects the two archetypes*. The assumption of the theorem is that the performance functions drop with a metric distance from the archetype (for example, with circular contours in Fig 14.6). But for some reason, as we will see when we look at data, the theorem seems to work well even in cases where it has no right to.

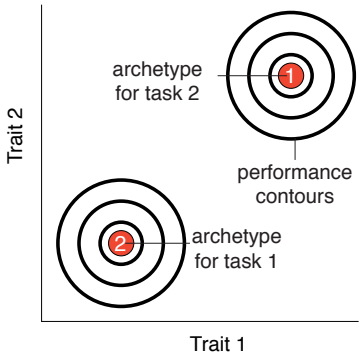

Figure 14.6

To understand why phenotypes fall on the line segment between the two archetypes, consider a phenotype B that is *not* on the line segment (Fig 14.7). The performance of B in each task is determined by its distances to the archetypes.

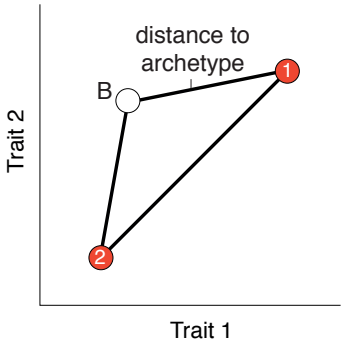

Figure 14.7

There is a phenotype A on the line segment connecting the archetypes which is closer to both archetypes (by the triangle inequality) as shown in Fig 14.8. Phenotype A therefore has better performance at both tasks, and therefore higher fitness than B. In an evolutionary race, A would win, and we can therefore erase B.

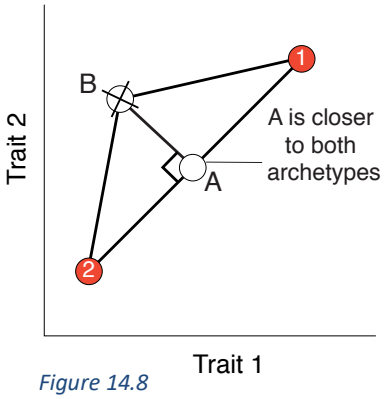

Figure 14.8

Now, there was nothing special about point B, so we can erase all of the points and remain with the line segment between the two archetypes (Fig 14.9). This is the set of phenotypes that cannot be improved at both tasks at once- the **Pareto front** (plotted in trait space, not in performance space).

14.6 Two tasks lead to a line segment, three tasks to a triangle, four to a tetrahedron

Thus, a tradeoff between two tasks predicts phenotypes on a line segment in trait space. Suppose we measure many beak traits, say 100 traits, making a 100-dimensional trait space. The beaks will still fall on the line segment between the two archetypes in this 100-dimensional trait space (Fig 14.10a). Measuring any two of these traits will still show a line, because the projection of the line on any plane is a line (gray line in Fig 14.10a). Thus, it is not too important which traits you measure, as long as they have to do with the same tasks.

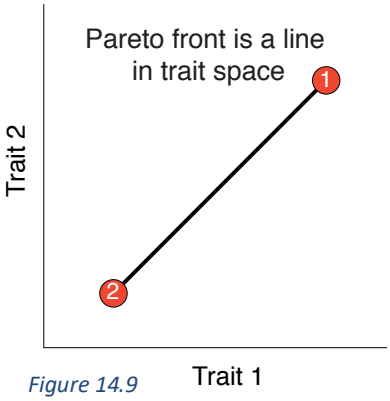

Figure 14.9

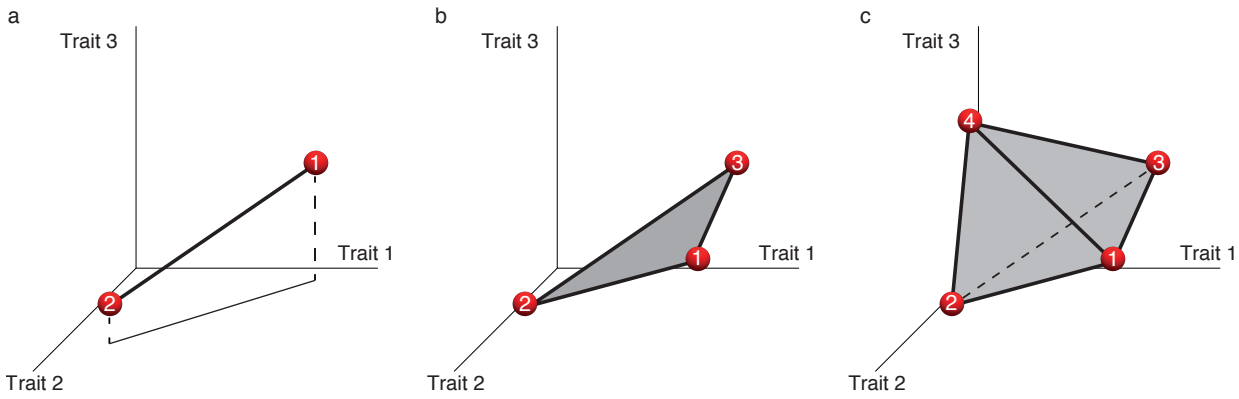

Figure 14.10

If there are three tasks, we expect the optimal phenotypes to fall inside a triangle, whose three vertices are the three archetypes (Fig 14.10b). If there are four tasks, the phenotypes will fall inside a tetrahedron (Fig 14.10c). In the case of very many traits, we can use dimensionality reduction methods such as principal component analysis (PCA) to visualize these shapes.

In general, a tradeoff between k tasks will result in a Pareto front shaped as a **polytope** with k vertices (a polytope is the generalization of a polygon or polyhedron to any dimension). Each vertex is an archetype for one of the tasks. A proof is given in solved exercise 14.1.

The key idea is that fitness is not just any function of traits $F(\vec{T})$, it is an increasing function of k performance functions of the traits $F(P_1(\vec{T}), P_2(\vec{T}), \ldots, P_k(\vec{T}))$. The maxima of these performance functions define k points in trait space, which a polytope. The maximum of F needs to be close to these k points, and hence inside the polytope.

If you make a nonlinear transformation of the traits, (eg measure T^2 instead of T), the polytopes will be deformed, (Fig 14.11). Deformed shapes can result from other situations, such as a non-metric decline of performance functions (Exercise 14.6). Even if the shapes are deformed, they still have sharp vertices at the archetypes.

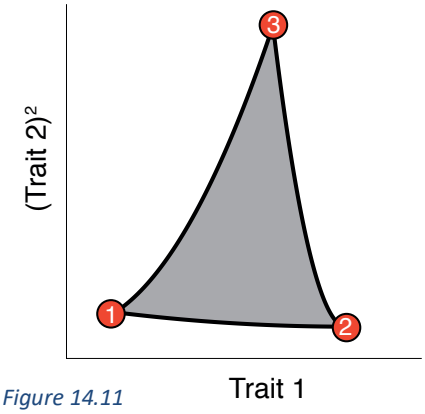

Figure 14.11

The neat use of this approach is to *discover what the tasks are directly from biological data*. The polytopes can help us to infer the tasks: The phenotypes closest to a vertex should be **specialists** at something, and that something gives clues to what the task might be (Fig 14.12). Phenotypes near the

center of the polytope should be **generalists**. This approach of inferring the tasks from the geometric shape of the data in trait space is called **Pareto task inference** (**ParTI**) ((Hart *et al.*, 2015)).

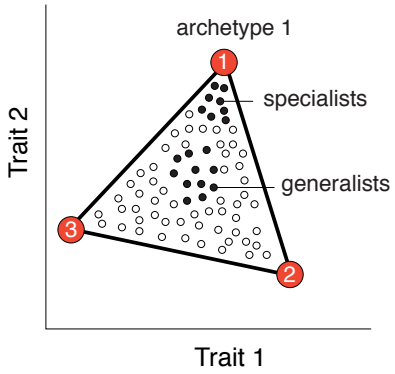

Figure 14.12

14.7 Tradeoffs in morphology

Now let's look at some data. We'll start with animal morphology, and then move to proteins and gene expression. Morphology is a field that measures the shapes of organisms, and morphology books are full of lines called **allometric relationships**. For example, the molar teeth of rodents (the three big teeth at the back of the mouth called M1,M2 and M3) vary in shape between rodent species. One can plot each species in a trait space of the relative tooth areas, the ratios M2/M1 and M3/M1[1]. These are dimensionless traits that normalize out the total size. In this trait space, the rodent species fall on a line (Fig 14.13) ((Kavanagh, Evans and Jernvall, 2007)). Most tooth configurations are not found, and thus most of trait space is empty.

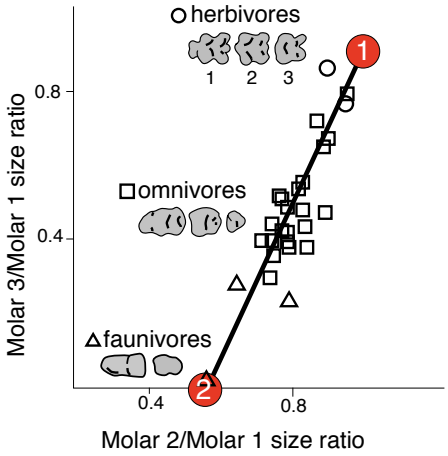

Figure 14.13

Each rodent species is represented by a point on the line. The position on the line depends on what the rodent eats. Plant eaters (herbivores) are found at one end, meat eaters (faunivores) at the other end, and omnivore generalists in the middle. This suggests a plant-eating archetype with equal sized molars (flat molars with area ratios of 1:1:1), and a meat-eating archetype with spiky molars with area ratios of 2:1:0. The line provides a rule in which the area of the middle molar is the average of its two neighbors. This rule that applies also to dinosaur teeth, allowing fossil hunters to infer how much meat versus plants the dinosaur ate.

Kavanagh et al ((Kavanagh, Evans and Jernvall, 2007)) also perturbed the development of rodent teeth, by adding morphogens or by blocking morphogen diffusion. Most perturbations led to tooth phenotypes that are still close to the line. This finding is related to the robustness of the developmental pathways, and to their ability to generate useful shapes even under perturbations, a feature called canalization (Chapter 12). Some perturbations, however, led to phenotypes far from the line, showing that the empty trait space is not impossible, and can be reached.

[1] Tooth areas, the traits favored by morphologists, give straight lines. If we plotted tooth length or volume instead of area, the line would be curved. Thus, nonlinear transformation of traits matter.

Morphological data also shows triangles. An example is found in the classic study of Darwin's finches by Peter and Rosemary Grant ((Grant, 1986)). The Grants lived on a tiny island in the Galapagos and observed finch evolution over decades. They measured five traits for each finch - mass, bone size, beak shape. This 5D data falls on a plane (the first two principal components explain over 90% of the variation). On this plane, the finches fall within a triangle (Fig 14.14). Their diet reveals three tasks: near the three vertices are specialist species at eating large seeds, small seeds and pollen/insects from cactus plants. Species in the middle of the triangle do a combination of these tasks.

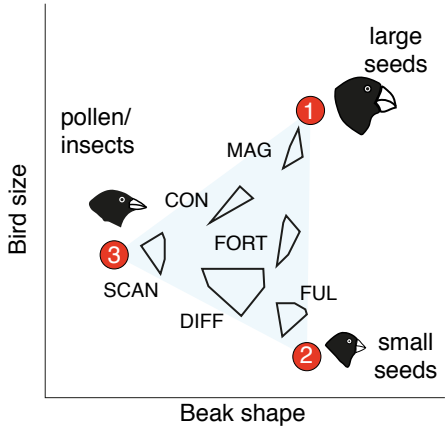

Figure 14.14

A triangle is seen also when each data point is an ant from the same nest (Fig 14.15). E. O Wilson measured the size of leaf-cutter ants versus the relative size of the gland which makes the pheromone for the ant trail ((Wilson, 1980)). He also recorded the behavior of each ant. There are three tasks: staying in the nest and nursing, soldiering, and foraging. Ants fill a continuum inside the triangle defined by these three archetypes.

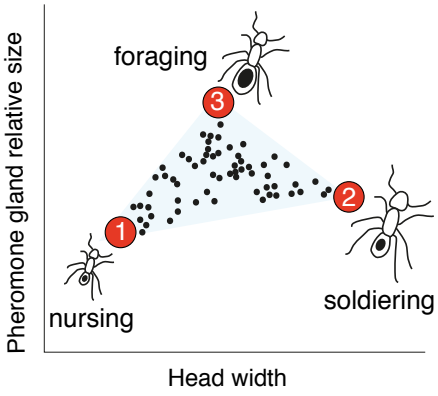

Figure 14.15

You might ask what is the functional role of the ants in the middle of the triangle? Why not make three clusters of specialists- optimal nursers, foragers and soldiers, without the generalist ants in the middle which are suboptimal at all tasks. Ant researchers believe that one reason for generalists is dynamic flexibility. Suppose the nest is attacked- there is no time to make more soldiers. Instead, generalist ants can be recruited to supply the needed tasks quickly. We will use this as a metaphor soon for division of labor between cells in an organ.

13.8 Archetypes can last over geological timescales

We can also ask whether the archetype positions in trait space move over long evolutionary timescales. A model system for this question is ammonites, marine creatures with detailed morphological data from 300 million years of evolution. The detailed data was collected in part because ammonite fossils are used to date rocks.

Ammonite shells can be described in a trait space with two parameters, as proposed by paleontologist David Raup ((Raup, 1967)) (Fig 14.16). In this trait space, the outer shell is a logarithmic spiral, whose radius grows with each whorl by a factor W, the whorl expansion rate. The inner shell is also a logarithmic spiral, with a constant ratio between the inner and outer shell radii, denoted D.

Figure 14.16

In this W-D trait space, ammonite shapes fill out a triangle (Fig 14.17). There is empty trait space, without ammonites, at large D and W. This empty trait space includes shells shaped like French horns, which are found in other clades, but not ammonites. The three archetypes at the corners of the triangle match the shell shapes that are optimal for three tasks: economy (maximal internal volume per shell material), swimming (lowest drag), and predator avoidance (rapid growth of shell diameter) ((Tendler, Mayo and Alon, 2015)).

There were three mass extinctions in which ammonites were wiped out except for one or two surviving genera. For example, the red dots in figure 14.17 mark the two surviving genera after the Permian/Triassic extinction 252 million years ago. Remarkably, in about 10 million years after each extinction, ammonites diversified to refill essentially the same triangle. This suggests that tasks and archetypes did not move much in this case.

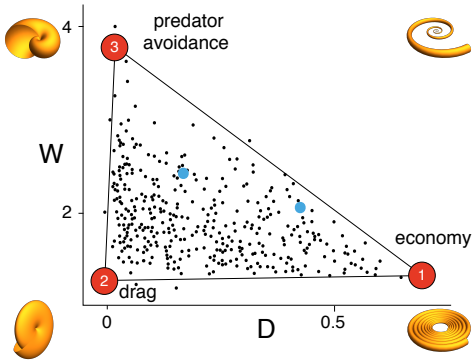

Figure 14.17

Only with the last extinction that wiped out the dinosaurs, 65 million years ago, this triangle-filling trick didn't work, perhaps due to competition with mammals. There is only one surviving genus in the ammonite lineage, called *Nautilus*.

If archetypes remain relatively fixed, there remains the question of how radically new tasks appear. How did novelties like vision and flight evolve, given that such tasks require complex organs such as eyes and wings. Organisms must somehow move out of an existing polytope towards a new archetype (e.g. flight performance).

Current thinking is that adaptation to a novel task arises by reuse of parts that have already evolved for a different task. An example is the evolution of wings from body appendages that served as thermal regulation devices. These appendages had selection pressure to grow in order to better radiate heat. When the appendages were large enough, they allowed the organism to glide, sparking selection pressure for aerodynamic gliding properties. Finally, the gliding appendages allowed rudimentary flight, and selection pressure worked to improve their

performance as wings. This picture is called **stepping-stone evolution**, because each new task is a stepping stone to the next.

14.9 Tradeoffs for proteins

Let's turn now from animals to proteins. A protein can also have multiple tasks. For example, Rubisco, one of the most abundant proteins in plants, is tasked with capturing CO_2 from the air and adding it to a sugar molecule that can be used to build biomass. All of the carbon in our bodies comes from Rubisco that made the plant biomass that is the basis for our food.

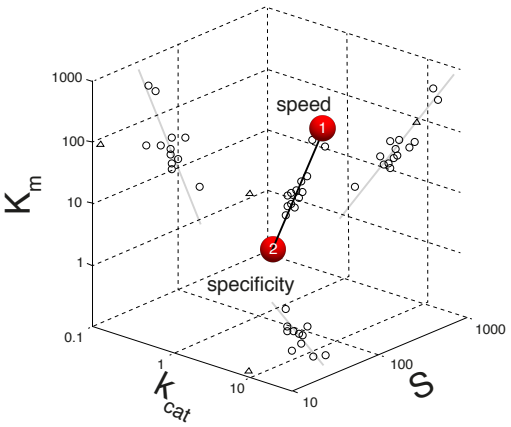

Figure 14.18

Rubisco can be characterized by a trait space with four kinetic parameters. Two of these parameters are the catalytic speed k_{cat} and affinity K_m to CO_2. The other two are the catalytic speed and affinity, k'_{cat} and K'_m, for the main competitor of CO_2, oxygen O_2. Capturing O_2 instead of CO_2 is a mistake that requires energy to correct.

To study tradeoffs in Rubisco, Yonatan Savir et.al. compiled these four kinetic traits from 30 photosynthetic organisms ((Savir *et al.*, 2010)). They found that the 30 Rubiscos fall approximately on a line in the 4D trait space. Figure 14.18 shows the data in the space of three traits, k_{cat}, K_m and the specificity $S = k_{cat}K'_m/k'_{cat}\ K'_m$, together with the projections of the data on the three planes. At one end of the line segment are the fastest Rubiscos, which occur in organisms like corn, known as C4 plant, that can concentrate CO_2. Since these plants reach a high CO_2 concentration inside their leaves, they do not need to worry about oxygen. At the other end are the slowest Rubiscos, but which bind CO_2 most strongly. These occur in organisms which do not concentrate CO_2 and face competition from O_2. Thus this protein seems to evolve under a speed-specificity tradeoff.

14.10 Tradeoffs in gene expression

This ParTI approach can also be applied to gene expression. At first glance, gene expression in a cell might seem very different from beaks or proteins. Cells can rapidly change gene expression according to their needs, whereas if you are born with a beak of a certain shape, you are stuck with it. Still, gene expression also faces tradeoffs.

Consider a brief time period, say a second, in which the cell can make say 1000 proteins. You can't make proteins to optimize rapid growth and at the same time make proteins to optimize stress resistance. Growth and stress require very different sets of proteins. The cell needs to choose

which protein portfolio to express based on the expectation of the future. Thus, the cell faces tradeoffs between tasks and thus it makes sense to look for polytopes in gene expression data.

Indeed, gene expression of the top 200 promoters in *E. coli*, which make up 90% of the total promoter activity, falls on a line segment (Fig 14.9). Here trait space is a space of gene expression, in which each axis is the fraction of the total promoter activity in the cell devoted to promoter i, with $i = 1 \ldots 200$. At one end of the line is the growth archetype, in which gene expression is focused on making ribosomes and machinery for biomass production. At the other end is the survival archetype in which cells express stress-response genes (and a small amount of ribosomes too in order to restart growth when things improve).

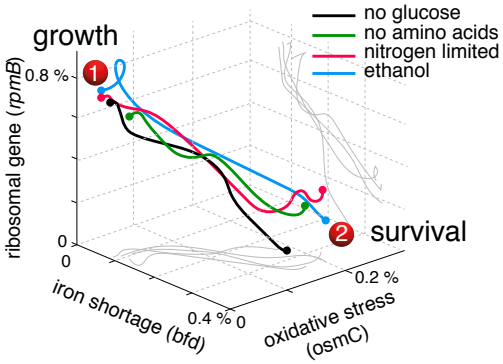

Figure 14.19

When placed in a test tube with nutrient, *E. coli* starts out close to the growth archetype, and grows exponentially until it begins to deplete the nutrient and pollute its environment. It gradually slides down the line to the survival archetype until conditions are so bad that growth stops. *E. coli* follows approximately the same line for different nutrients and conditions.

Thus, all that *E. coli* needs to do in a new condition is decide about its position on the line. This means that it needs to choose a number θ between zero and one, with the growth archetype at $\theta = 0$ and stress archetype at $\theta = 1$. To choose this number, *E. coli* uses a simple mechanism to put its gene expression on a line. This line-making mechanism is based on competition between two **sigma factors**, proteins that bind RNA polymerase (RNAp) and allow it to bind sites in promoters (Fig 14.20). One factor, σ_{70}, binds sites in the promoters of growth genes and the other, σ_S, binds promoters for stress-response genes.

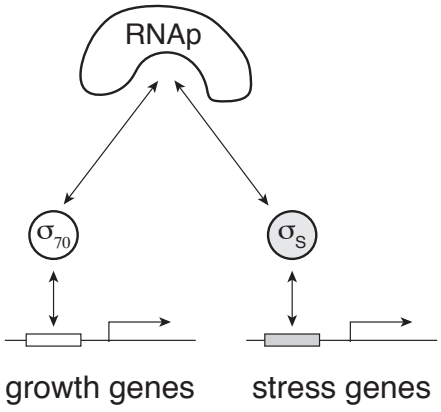

Figure 14.20

Thus, the position on the line is given by the fraction of RNAp bound to σ_S, namely $\theta = \sigma_S/(\sigma_S + \sigma_{70})$. *E. coli* has signaling systems that read the environment and accordingly produce and degrade the two sigma factors, in order to determine where the cell lies between the tasks of growth and survival. The coordinates of the archetypes are encoded in the strength of the sites for the two sigma factors in each promoter (many promoters have binding sites for both sigma factors). A polytope with k vertices can be achieved by a similar design with k competing factors.

14.11 Division of labor in the cells that make up an organ

We now turn from bacteria to gene expression in human cells. Human tissues are made of different types of specialized cells: brains are made of neurons and livers are made of hepatocytes. Having different cell types for each tissue allows a useful division of labor, assigning metabolic tasks to the liver and thinking tasks to the brain.

What about division of labor within cells of a given type, say the hepatocyte cells in the liver? Recall the ants, which divide labor towards a collective goal of colony survival and reproduction. Are there specialists and generalists also within a cell type?

Analysis of gene expression from individual cells all from the same organ indicates that division of labor is widespread. Gene expression of cells of a given cell type typically falls in a continuum bounded inside shapes with pointy vertices ((Korem *et al.*, 2015), Adler et al, 2019). The tasks of the cell type can thus be inferred.

For example, liver hepatocytes are famous for doing multiple functions. They synthesize blood proteins and other essential compounds, they detoxify the blood, get rid of ammonia by turning it into urea, and regulate glucose levels by storing it into glycogen or making it from amino acids when needed (gluconeogenesis). Individual liver cells fill out a tetrahedron in gene expression space, where each axis is the expression of gene i, with i=1…20,000. This tetrahedron is plotted in Fig 14.21, where each point is a cell, and the axes are the first three principal components of gene expression.

At the vertices of the tetrahedron are cells that specialize at four key tasks: synthesis of blood proteins (such as albumin), gluconeogenesis, detoxification and, surprisingly, lipid metabolism/iron homeostasis. Each archetype has additional secondary tasks, so that each specialist carries out a 'syndrome of tasks': for example, the gluconeogenesis archetype also produces the antioxidant glutathione.

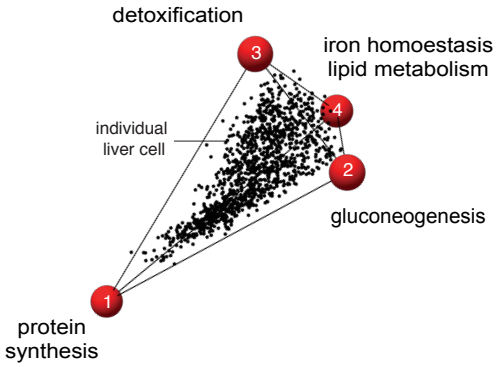

Figure 14.21

The specialist cells have a particular arrangement in space across the liver ((Halpern *et al.*, 2017)). The liver is made of repeating hexagonal columns called liver nodules, about 15 cells across. The cells that specialize in synthesis (albumin, glutathione) tasks that require much oxygen, are found in the oxygen-rich boundary of the hexagons, near the portal veins. Cells that specialize in detoxification, which requires less oxygen, are found at the oxygen-poor center of the hexagon where the central vein drains the lobule. Lipid/iron specialist cells are found in the middle. This placement of specialists at positions best suited to their task helps maximize the organ performance at all tasks (Adler,2019).

These individual liver-cell experiments are an example of the technologies that provide the ability, undreamed of when I was a postdoc twenty years ago, to measure thousands of numbers from

each individual cells. They produce massive amounts of data. How we can analyze such information-rich experiments in biology? The challenge is that human beings cannot visualize high-dimensional data. We can deal with a 2D pictures or sometimes a 3D volume, but 4D, not to speak of 20,000D, is alien to us. That is why we need approaches to reduce dimensionality into something we can comprehend.

The problem is that each dimensionality reduction method has an implicit model of what the data is like. If the data is unlike that model, the method can be misleading. Consider a commonly used approach, called data **clustering**. Data clustering assumes that data is arranged in distinct, well-separated clouds, and clustering algorithms can easily detect those clouds. However, clustering does poorly when data is arranged in a continuum– there are no natural distinct clusters.

In this case, other algorithms, known as **archetype analysis algorithms** ((Mørup and Hansen, 2012)), can help detect whether data can be approximated as a continuum filling a polytope, and how many vertices the polytope has ((Hart *et al.*, 2015)). These algorithms focus on the outside contours of the data. As in any approach, be wary of artifacts that make data spuriously look like a line or triangle.

14.12 Variation within a species lies on the Pareto front

Let's end by returning to animal morphology, in an example that opens up new questions. This example helped to start the ParTI framework, when Kathy Kavanagh showed me data on bird toes. The fourth toe of the bird has four bone segments called phalanges- similar to the bones in our fingers. Each bird species can be plotted in a trait space whose axes are the areas of three of the phalanges normalized by the area of the fourth. In this 3D trait space, birds fall on a plane and on that plane they fill out a triangle (Fig 14.22).

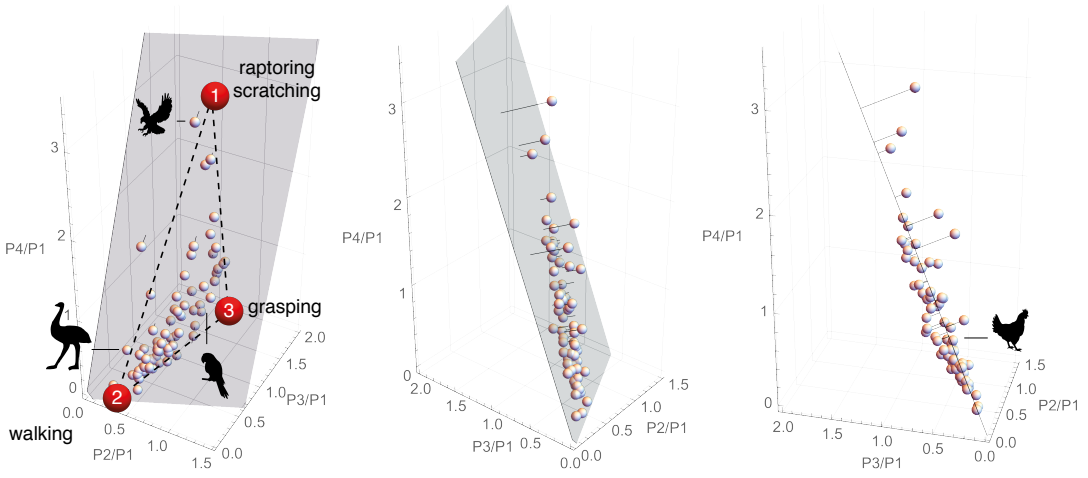

Figure 14.22

If I wrote this chapter clearly, you can see that the triangle indicates that the bird toes have three tasks. We can infer the tasks by looking at the birds closest to each vertex of the triangle. We see parrots and other perching birds near one archetype. The task is grasping, and the toe shows the biomechanical optimum for grasping a branch, namely equal-sized (and curved) phalanges. The second archetype is close to ostriches and other walking birds - it is the walking archetype with phalanges that decrease in area with ratios 4:2:1:0. Again, a biomechanical optimum: If you ever wondered about the difference between your hands and feet, your hands are for grasping and have equal sized parts, and your feet are for walking and have a long footpad with short toe for kicking off. The third archetype is for raptoring/scratching, and has a long fourth phalanx that provides a good-sized lever for the talon.

But let's ask a different question. The point for a given species, say chicken, is the average over all chickens measured. What if we look at *individual* chickens? Each chick is born with slightly different bone ratios (and these ratios are set for life already in the egg). This variation is due to the combination of the parent's genomes that provide each individual with a unique combination of genetic differences called **polymorphisms**. Variation further arises from the randomizing effect of noise during development. So individuals will form a cloud in trait space around the chicken average. Does this cloud go in all directions, or is it flattened like a pancake along the triangle defined by different species?

Kathy Kavangh tested this by hatching 100 chicken eggs and 100 zebra-finch eggs ((Kavanagh *et al.*, 2013)). She found that the cloud of variation is flattened like a pancake along the triangle defined by different species (Fig 14.23). Exaggerate the difference between two individuals and you get a caricature of another bird species.

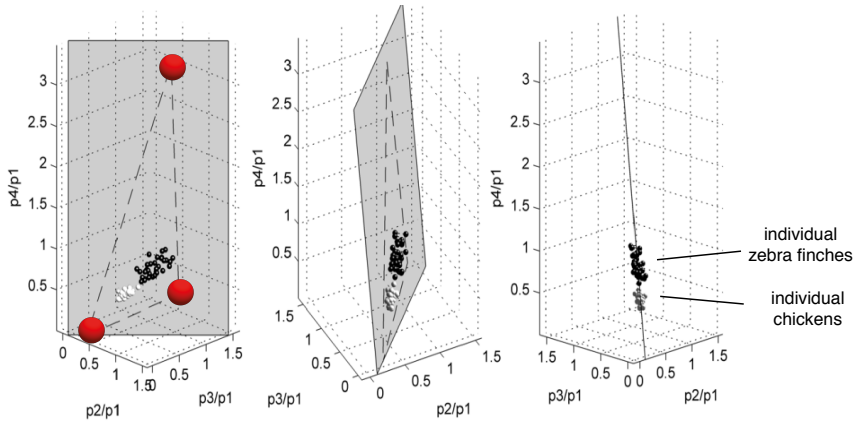

Figure 14.23

How can this be? As mentioned above, variation between individuals comes from a combination of noise and polymorphisms. To maintain the chicks on the front, the prevalent polymorphisms must push the phenotype along the front, but importantly not off of the front. We will call such polymorphisms **aligned polymorphisms** (Fig 14.24). Aligned polymorphisms can be selected because chickens have a range of niches in which walking, grasping and scratching are differentially important ((Sheftel *et al.*, 2018)). Each individual gets a mix of polymorphisms that create a cloud of variation aligned with the front. Aligned polymorphisms help keep the phenotype close to the optimum in at least one possible niche; polymorphisms that move the phenotype off of the front face the risk of a competitor on the front with higher performance at all tasks.

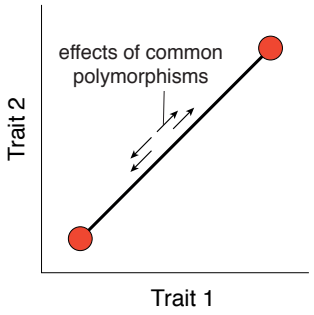

Figure 14.24

Furthermore, in order to produce the observed variation, developmental patterning mechanisms need to somehow focus the effects of noise along useful directions aligned with the front, a kind of 'aligned canalization'.

The importance of aligned genetic variation and canalization can be appreciated if we recall that every individual is a never-tried-before combination of mom and dad, a pastiche of their millions of genetic differences. It's like designing a new jet plane and flying it out of the hangar without ever testing it. Aligned polymorphisms and canalization mean that every offspring has a good chance to lie somewhere on the Pareto front, and therefore can be competitive in one of the niches available to the species[2]. What types of developmental mechanisms and genetic population structure is needed for this type of variation is an open question.

To sum up, natural selection is usually a multi-objective optimization problem. Organisms are therefore rarely optimal for a single task, but instead evolve under tradeoffs. Evolutionary tradeoffs lead to patterns in phenotype space in which a continuum of possibilities is bounded within polyhedral-like shapes with pointy vertices. The pointy vertices can be used to infer the tasks at play. The position of each phenotype relative to the vertices (the archetypes) tells us how important each task was in its evolution. This notion applies in principle to any scale - molecules, circuits, cells, organisms – as long as natural selection has had enough time, population size and genetic variation to approach the optimal tradeoffs between tasks.

Further reading

Natural selection as a multi-objective problem

[2] One more requirement is that the genetic variation be additive in the sense that multiple polymorphisms do not generate components perpendicular to the front. Such additivity can itself be seleted ((Sheftel *et al.*, 2013)).

(Arnold, 1983) " Morphology, performance, fitness"

Evolutionary tradeoffs and Pareto task inference

(Shoval *et al.*, 2012) "Evolutionary Trade-Offs, Pareto Optimality, and the Geometry of Phenotype Space"

(Sheftel *et al.*, 2013) "The geometry of the Pareto front in biological phenotype space."

Adler ..Alon (2019) – new liver paper

Animal morphology

(Kavanagh, Evans and Jernvall, 2007) "Predicting evolutionary patterns of mammalian teeth from development"

(Grant, 1986) "The ecology and evolution of darwins finches"

(Kavanagh *et al.*, 2013) "Developmental bias in the evolution of phalanges"

(McGhee, 2006) "The geometry of evolution: adaptive landscapes and theoretical morphospaces. Cambridge"

Algorithms for archetype analysis

(Mørup and Hansen, 2012) "Archetypal analysis for machine learning and data mining."

(Hart *et al.*, 2015) " Inferring biological tasks using Pareto analysis of high-dimensional data. "

Exercises

Exercise 14.1: Mathematical proof for the main theorem in this chapter. Prove that if fitness is an increasing function of k performance functions in trait space, and that each performance function $i = 1 \ldots k$ has a maximum at a point called archetype i, and that performance drops with Euclidean distance from the archetype, then the point of maximum fitness is found inside the polytope defined by the k archetypes (Shoval , 2012).

Solution:

Each phenotype is described by a vector of traits T (for convenience we will drop the vector sign from now on). Fitness F is an increasing function of the performance at the k different tasks, $F(T) = F\left(P_1(T), P_2(T), \ldots, P_k(T)\right)$. Each performance function P_i has a maximum at archetype i, A_i, and performance decreases with Euclidean distance from the archetype $P_i(T) = P_i(||T - A_i||)$. We will show that the optimal phenotype (the phenotype that maximizes F) is a

(i) a weighted average of the archetypes.

(ii) Weights are positive and sum to one.

which means that the optimal phenotype is inside the polytope defined by the k archetypes. Another way to say this is that the optimal phenotypes are convex combinations of the archetypes.

The optimal phenotype maximizes fitness, $dF/dT = 0$. Let's denote the distance from the archetype of task i by $r_i = ||T - A_i||$, so that $dr_i/dT = 2(T - A_i)$. Using the chain rule, $\frac{dF}{dT} = \sum_i \frac{dF}{dP_i}\frac{dP_i}{dr_i} 2(T - A_i) = 0$. Thus the optimal T that solves this equation, T_{opt}, is a weighted average of the archetypes

$$T_{opt} = \sum_i \theta_i A_i$$

with weights

$$\theta_i = \frac{\frac{dF}{dP_i}\frac{dP_i}{dr_i}}{\sum_j \frac{dF}{dP_j}\frac{dP_j}{dr_j}}$$

where all derivatives are at T_{opt}. Note that the the weights sum to one, $\sum \theta_i = 1$. The weights are positive $\theta_i > 0$ because F increases with performances (so that $\frac{dF}{dP_i} > 0$) and performance decreases with distance from the archetype ($\frac{dP_i}{dr_i} < 0$) so that all terms have a negative sign which cancels out. Hence

(i) The optimal phenotype is a weighted average (convex combination) of the archetypes $T_{opt} = \sum \theta_i A_i$

(ii) The weights are positive and sum to one ($\theta_i > 0$, $\sum \theta_i = 1$). Another way to say this is that the k maxima of the k performance functions define a k-1 dimensional shape in trait space (the convex hull of those k points), and optimal phenotypes are trapped within inside this shape. For two tasks, $k = 2$, this shape is a line segment: $T = \theta A_1 + (1 - \theta)A_2, 0 \leq \theta \leq 1$.

14.2 Phenotype position is determined by fitness and performance gradients:

This exercise shows that the position inside the polytope can provide information. Exercise 14.1 shows that the optimal phenotype for a given fitness function F is a weighted average of the archetypes.

(a) Show that the weight for each task can be interpreted as the importance of the task to fitness times the sensitivity of its performance to changes in the traits.

(b) Show that in a niche in which one task has much greater effect on fitness than the other tasks, the phenotype will be close to the corresponding archetype. This is a specialist phenotype.
(c) In the case of ammonites, how can the position in the triangle help us to understand what might have been the selective conditions for each genus?
(d) Discuss the effects of adding a new task to k existing tasks. The new task has a small effect on fitness and can be carried effectively out by a large range of traits.

Solution to (a): the weights from exercise 14.1 are

$$\theta_i = \frac{\frac{dF}{dP_i}\frac{dP_i}{dr_i}}{\sum_j \frac{dF}{dP_j}\frac{dP_j}{dr_j}}$$

The weight for task i is therefore the normalized product of the importance of the task to fitness dF/dP_i times the sensitivity of the performance to changes in the trait dP_i/dr_i.

14.3 Multiple tasks break the symmetry of neutral spaces:

Consider a system with two tasks. Task 1 has a performance function whose maximum is at point A, the archetype for that task, and performance decays with Euclidean distance form the archetype. Task 2, however, has a maximum performance not attained at a single point, but instead *in an entire region N* in trait space (Fig 14.25). Such a region, in which all points have the same performance, is called a **neutral space.** Performance decreases with Euclidean distance from N (distance to the closest point in N).

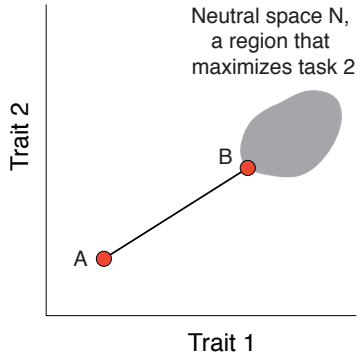

Figure 14.25

(a) Show that only one point in N is on the Pareto front. It's as if task 1 "chooses" one point in N. It thus breaks the symmetry between the points in N.

(b) Show that the Pareto front is a line segment that connects point A with a point on the boundary of N.

(c) Give a possible biological example for a task which is maximized at a region and not a point in trait space.
(d) What would happen if task 2 was the only task affecting fitness?

Solution: (a) The chosen point is the point in N closest to A. To see this, let's call this point B. Choose a point C in N other than B. Its performance in task 1 is lower than B (because B is closer

to A by definition), and its performance at task 1 is the same as B since both are in the neutral space. Hence C is dominated by B and can be removed. We are left with point B.

14.4 The Pareto front is where performance contours are tangent:

(a) Show that the Pareto front is the set of points in which the contours of the performance functions are externally tangent.

(b) Use this to explain why, when performance functions decline with Euclidean distance from their maxima, contours are circular and the set of tangent points is a line (Fig 14.26a).

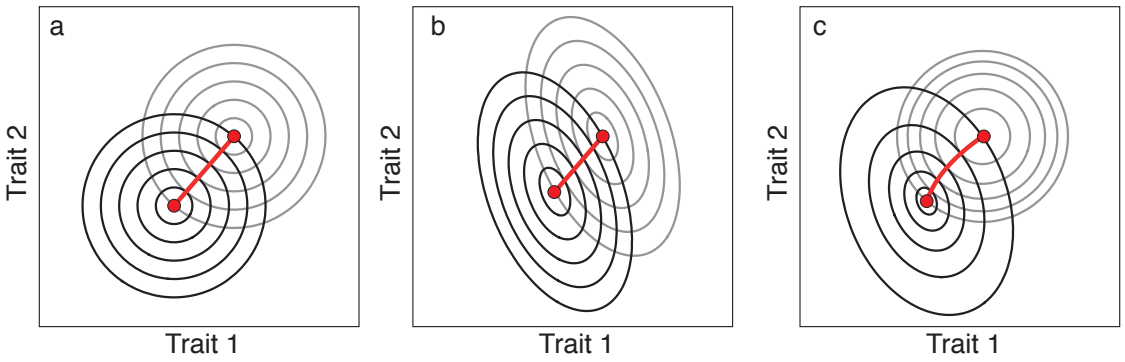

Figure 14.26

14.5 Pareto front is a line if both performances decay with the same inner-product norm:

This exercise shows that when the performance functions decline not with Euclidean distance, but all decline with the same inner-product norm (giving different traits differential impact on performance, with elliptical contours), the theorem of 14.1 still applies.

Consider the case of two tasks, in which performance functions decay with distance given by an inner product norm, $r_i = (T - A_i)^T Q (T - A_i)$where Q is a positive definite matrix. The inner product norm is the same for both tasks. Their contours are therefore concentric parallel ellipses around the archetype, such that both tasks have ellipses of the same orientation and eccentricity (Fig 14.25b).

(a) Sketch the contours.
(b) Show that the Pareto front is a line segment (use exc 14.4).
(c) Compute the Pareto front for the case of k tasks. Show that it is exactly the same as in the case of Euclidean norms- a polytope with vertices at the archetypes (hint: rotate and dilate space until the contours are circular).

14.6 Pareto front is curved if performances decay with different norms

Consider the case of two tasks, and performances that each decays by a different inner-product norm (Fig 14.26c). Their contours are therefore concentric parallel ellipses around each archetype, but with different orientation and eccentricity (Fig 14.26b).

(a) Explain why the Pareto front is curved. Show graphically that most curved front occurs when the long axes of the ellipses are at 45 degrees to each other, and the long axes are much longer than the short axes.
(b) Consider the case where the long axes of the ellipses for the two tasks are orthogonal to each other. Sketch the contours. Is the front curved or straight?
(c) What is the biological meaning of the difference in norms and contour shape?

14.7 Bounds for general performance functions:

This exercise shows that even when we know nothing about the shape of the performance functions (no norms or even no monotonic decay from the archetype), one can still bound the Pareto front in a region between the archetypes that lies between certain contours (Sheftel,2013).

Consider the case of two traits and two tasks. Performance functions can have a general form, with global maxima at the archetypes A_1 and A_2.

(a) Show that the Pareto front is bounded inside the region between two special contours, C_{12}, the contour of performance 1 that goes through A_2, and C_{21}, the contour of performance 2 that goes through A_1 (Fig 14.27a).
(b) Suppose that the performance functions have local maxima in addition to the global maximum, such that the contours C_{12} and C_{21} have disconnected pieces that go around the local maxima (Fig 14.27b). Show that generally, the local maxima do not affect the Pareto front. Show that only when the local maxima of the two performance functions happen to be close to each other, the Pareto front can have multiple disconnected pieces (14.27c).

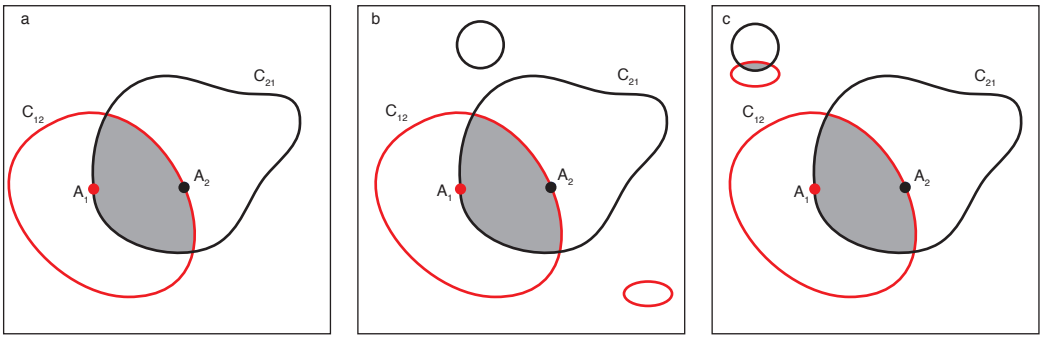

Figure 14.27

14.8 Too few traits measured: Suppose a system has four tasks, but only two traits are measured experimentally. What kind of shapes would describe the optimal phenotypes in the 2D trait space?

Solution: Triangle or kite. The four tasks generally lead to a tetrahedron with 4 archetypes at the vertices. Measuring two traits means projecting the tetrahedron on the plane. Projections of a

tetrahedron are shaped as a quadrangle (kite) or a triangle. The latter case is when one of the archetypes is occluded.

14.9 Empty regions in the polytope: The theorems we discussed so far are silent on the question of where in the polytope the points can lie. This exercise will show that some regions of the polytope can be empty (forbidden) if the performance functions have certain curvatures. We will use a 1D example, with a single trait T and two tasks.

(a) Show that the Pareto front is the line segment between the two archetypes.

(b)Show that a condition for optimality is $\frac{d^2F}{dT^2} < 0$.

(c) Show that this requires a condition on the curvature of the performance functions d^2P_i/dT^2.

(d) Show that when both curvatures are positive, there is an empty region with no phenotypes.

(e) What happens when the performance functions are Gaussians that decay with distance from the archetype?

(f) What other reasons might explain an empty region inside a polytope (hint: consider physical constraints on the phenotype).

14.10 Mass-longevity triangle ((Szekely *et al.*, 2015)**):** Plotting the longevity of mammalian and bird species versus their mass shows a continuum inside a triangle-like shape (Fig 14.26). At the three vertices are shrews (that weigh a few grams and live about 2 years), elephants and whales (tens of tons, ~100 years) and small bats (a few grams, ~50 years). Near the bat archetype are mammals that live in trees and social mammals that live underground (naked mole rat). Flying birds are found near the bat archetype and walking birds near the bottom edge of the triangle. Interpret these findings in terms of tradeoffs and tasks.

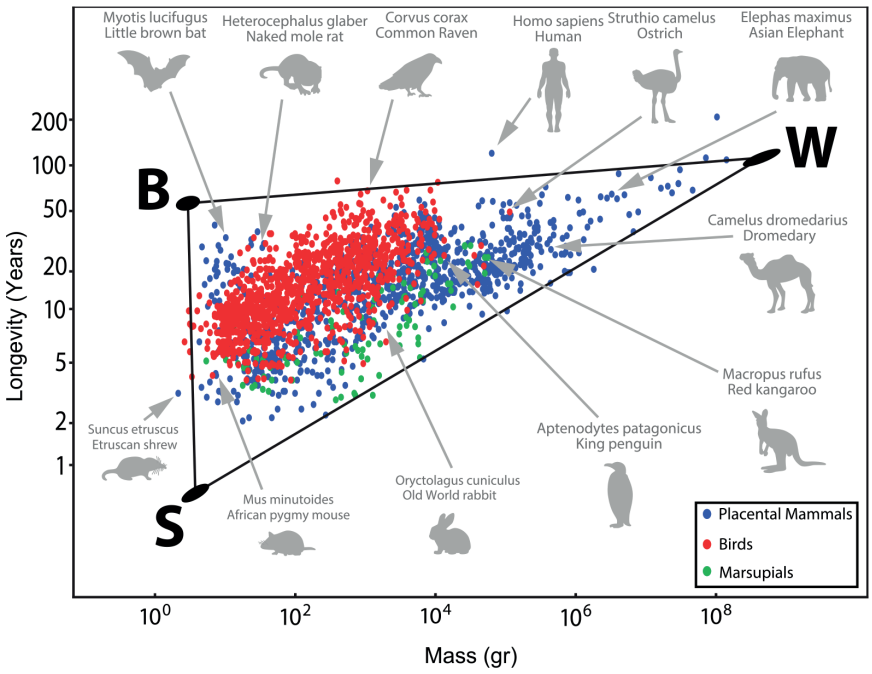

Figure 14.28

14.11 Different modules of tasks: Suppose that an organism has two parts or modules, each with a different set of tasks and traits. For example, a bird has beak traits devoted to tasks of eating and toe traits devoted to the tasks of walking/grasping. Suppose that each module has two traits and two tasks.

(a) What would the Pareto front look like in the 4D trait space?

(b) What would happen in a ParTI analysis if we didn't realize that toes and beaks are separate and mixed the traits together into one big dataset?

(c) Harder problem: Devise an algorithm to detect in a large dataset whether there are multiple separable modules of traits and tasks, each with each own Pareto front (give outline of algorithm, 100 words).

14.12 Molecular mechanism for polytopes in gene expression: In this exercise we consider a mechanism that can generate a polytope with k vertices for gene expression. It is a generalization of the σ-factor mechanism in Fig 14.20. Suppose that k transcription factors X_i , with $i = 1 \ldots k$, regulate genes, but are active only when bound to protein Y, in the complex $[X_i Y]$. The binding is strong with affinity K_i. Each gene promoter has binding sites for one, some or all of the X_i . The expression of gene j when $[X_i Y]$ binds its site is w_{ij}, and the effects of the regulators add up (SUM gate).

(a) Show that the expression of gene j is $T_j = Y_T \Sigma_i w_{ij} X_i \, K_i^{-1} / \Sigma_i X_i K_i^{-1}$ where Y_T is the total level of Y.

(b) Show that gene expression lies in a polytope with k vertices in gene expression space whose axes are T_j.

(c) What are the coordinates of the archetypes in terms of the mechanism parameters?

(d) In each condition, the cell regulates the concentrations of X_i. What is the concentration combination that leads to gene expression near vertex 1? Near the middle of the polytope? Outside of the polytope?

(e) What happens to the polytope if we delete one of the regulators?

(f) What happens if we change the level of Y, Y_T?

14.13 Aligned mutations in the polytope mechanism: Consider the mechanism of exercise 14.12. Suppose a mutation can change one regulator concentration X_i, one binding site strength w_{ij} or Y_T.

(a) Which mutations move the phenotype along (inside) the front?

(b) Which mutations move the phenotype off of the front?

(c) Which mutations change the archetype coordinates?

(d) Which mutations would you expect to see at high prevalence (that is, which common polymorphisms) in a population of organisms facing niches that share the same k tasks, but where each niche gives different weighting to each task?

14.14 Pareto optimality in engineering

Consider the performance space of car designs, with the performances of acceleration and economy (Fig 14.X).

(a) Where are the archetypes in performance space?
(b) What would be examples of relevant traits in a trait space of cars?
(c) How is ParTI different from Pareto analysis in performance space?

14.15 Tradeoffs in a network motif. Consider the negative autoregulation network motif of chapter 2, with dynamics $\frac{dX}{dt} = \frac{\beta}{1+\left(\frac{X}{K}\right)^n} - \alpha X$ where X is a stable protein. Suppose the tasks are speed (fast response time) and economy (minimal protein production integrated over one cell generation, $\frac{\log(2)}{\alpha}$).

(a) Plot performance space and trait space, with the traits β, K, n .

(b) What is the Pareto front?

(c) When is simple regulation selected?

14.16 Performance space of fold-change-detection (FCD) designs:

Read (Adler *et al.*, 2017). Explain how the Pareto front concept is used to define a handful of FCD circuit designs.

14.17 Aligned canalization: Analyze the French flag model of morphogen pattern formation (Chapter 12). In this model, morphogen X is produced at position $x = 0$ and diffuses with diffusion coefficient D and is degraded at rate α. Cell fate decisions are in thee regions defiend by the points in space where morphogen crosses the thresholds T_1 and T_2.

(a) What is the effect of varying thresholds, D, and α on the patterns?

(b) Define a trait space given by the ratios of the lengths of cell-fate regions. What suite of variations occurs upon changes in thresholds, D, and α?

(c) Could such a design provide aligned canalization for a given set of tasks?

14.18 Optimal arrangement in space (Adler et al, 2019):

Consider a tissue with a spatial coordinate x, with gradients of oxygen and nutrients across x. Cells in the tissue have two tasks, and performance depend on space and on gene expression, $P_i(x) = \phi_i(x)P\left(\left|\left|T(x) - A_i\right|\right|\right)$ where $T(x)$ is gene expression of the cells at position x and A_i are the archetypes. The collective performance at task *i*, summed over all cells , is S_i. The overall function of the tissue is an increasing function of the collective performances $F = f(S_1, S_2)$.

(a) Suppose that performance functions P have negative curvature. Solve for the gene expression profiles as a function of position $T(x)$.
(b) Are there specialist cells and generalist cells?
(c) Task 1 is performed best near *x=0*, and task 2 is insensitive to space ($\phi_1 = 1 - x, \phi_2 = 1$). What is the spatial expression profile?
(d) Relate this problem to the case of hepatocytes in the main text.
(e) What Pareto front shapes might be expected in a tissue with 3D spatial gradients?

14.19 Triangles can result from other reasons: A coin is tossed N times, and the number of heads H is recorded. This is repeated 100 times for N=1,2,..100.

(a) Sketch the data in a trait space whose axes are the number of tosses N versus the number of heads H.

(b) Explain why the data resembles a triangle.

(c) Is there anything special about the data points near the vertices?

14.20 Why only 2-4 archetypes, and not more, in most dataset?

In the systems analyzed in this chapter, we saw evidence for 2,3 or4 tasks, namely lines, triangles or tetrahedra. Why don't we see many more tasks, say 10 or 100?

Solution: To see k archetypes requires k tasks to have the same, large, influence on fitness. More precisely, the traits on the Pareto front are convex combinations of the archetypes $\Sigma_i \theta_i A_i$ with the weight for archetype i proportional to the impact of task i on fitness $A_i = dF/dP_i$ times the sensitivity of the performance at task i to the traits: $B_i = dP_i/dr_i$ namely $\theta_i = A_i B_i / \Sigma_i A_i B_i$ (Exc 14.2). It is reasonable to assume that tasks have widely distributed impacts and sensitivities, so that it is unlikely that many tasks will have large weights that are on the same order of magnitude. Biological systems usually have many additional tasks of small impact and/or small sensitivity (that is, can be performed well by many phenotypes), and these tasks exert a weak 'gravitational pull' on the data. They are undetectable without very precise data. Can you think of such low-impact tasks, for example, for bird beaks?

Adler, M. *et al.* (2017) 'Optimal Regulatory Circuit Topologies for Fold-Change Detection', *Cell Systems*. doi: 10.1016/j.cels.2016.12.009.

Alon, U. (2009) 'How To Choose a Good Scientific Problem', *Molecular Cell*. doi: 10.1016/j.molcel.2009.09.013.

Arnold, S. J. (1983) 'Morphology, performance and fitness', *Integrative and Comparative Biology*. doi: 10.1093/icb/23.2.347.

Grant, P. R. (1986) 'Ecology and evolution of Darwin's finches', *Grant, P. R. Ecology and Evolution of Darwin's Finches. Xiv+458p. Princeton University Press: Princeton, N.J., USA. Illus. Maps*. doi: 10.1167/iovs.10-6486.

Halpern, K. B. *et al.* (2017) 'Single-cell spatial reconstruction reveals global division of labour in the mammalian liver', *Nature*. doi: 10.1038/nature21065.

Hart, Y. *et al.* (2015) 'Inferring biological tasks using Pareto analysis of high-dimensional data', *Nature Methods*. doi: 10.1038/nmeth.3254.

Kavanagh, K. D. *et al.* (2013) 'Developmental bias in the evolution of phalanges', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1315213110.

Kavanagh, K. D., Evans, A. R. and Jernvall, J. (2007) 'Predicting evolutionary patterns of mammalian teeth from development', *Nature*. Nature Publishing Group, 449(7161), pp. 427–432. doi: 10.1038/nature06153.

Korem, Y. *et al.* (2015) 'Geometry of the Gene Expression Space of Individual Cells', *PLoS Computational Biology*, 11(7). doi: 10.1371/journal.pcbi.1004224.

McGhee, G. R. (2006) *The geometry of evolution: Adaptive landscapes and theoretical morphospaces*, *The Geometry of Evolution: Adaptive Landscapes and Theoretical Morphospaces*. doi: 10.1017/CBO9780511618369.

Mørup, M. and Hansen, L. K. (2012) 'Archetypal analysis for machine learning and data mining', *Neurocomputing*. doi: 10.1016/j.neucom.2011.06.033.

Raup, D. M. (1967) 'Geometric Analysis of Shell Coiling: Coiling in Ammonoids', *JPaleo*. doi: 10.2307/1301903.

Savir, Y. *et al.* (2010) 'Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.0911663107.

Sheftel, H. *et al.* (2013) 'The geometry of the Pareto front in biological phenotype space', *Ecology and Evolution*, 3(6). doi: 10.1002/ece3.528.

Sheftel, H. *et al.* (2018) 'Evolutionary trade-offs and the structure of polymorphisms', *Philosophical Transactions of the Royal Society B: Biological Sciences*. doi: 10.1098/rstb.2017.0105.

Shoval, O. *et al.* (2012) 'Evolutionary trade-offs, pareto optimality, and the geometry of phenotype space', *Science*, 336(6085). doi: 10.1126/science.1217405.

Szekely, P. *et al.* (2015) 'The Mass-Longevity Triangle: Pareto Optimality and the Geometry of Life-History Trait Space', *PLoS Computational Biology*, 11(10). doi: 10.1371/journal.pcbi.1004524.

Tendler, A., Mayo, A. and Alon, U. (2015) 'Evolutionary tradeoffs, Pareto optimality and the morphology of ammonite shells', *BMC Systems Biology*, 9(1). doi: 10.1186/s12918-015-0149-z.

Wilson, E. O. (1980) 'Caste and division of labor in leaf-cutter ants (Hymenoptera: Formicidae: Atta)', *Behavioral Ecology and Sociobiology*. Springer-Verlag, 7(2), pp. 157–165. doi: 10.1007/BF00299521.