28. *Rugged-Landscape Models*

Amazing the things you find when you bother to search for them.

—attributed to Sacagawea

In this chapter, we study the rugged-landscape model. Like spatial and hedonic models, the rugged-landscape model defines an entity as a collection of attributes. Each set of attributes maps to a value. The goal is to modify attributes to construct an entity of highest value. This model originated in ecology to study evolution. It is now also used to study problem solving, competition among firms, and innovation. That will be our focus here. We apply the model to reveal how interdependence in the effects of attributes makes innovation difficult and leads to path dependence in the solutions found and also leads to a greater variety of solutions. We also see how more difficult problems benefit from a greater diversity of problem solving approaches.

The chapter consists of three parts followed by a discussion of how to extend the model to capture competition. In the first part, we describe the ecological model of a fitness landscape and show how we can reinterpret it as a model of problem solving and innovation. In the second part, we discuss the implications of ruggedness within a one-dimensional model. In the third part, we present the NK model of rugged landscapes, which extends the one-dimensional model to an arbitrary number of binary dimensions.

The Fitness Landscape

The *fitness landscape model* assumes species have features or traits that contribute to their fitness, loosely defined as their reproductive potential, and that individual members of a population differ in how much they have of a particular trait. If we plot the amount of the trait on the horizontal axis and the fitness of species on the vertical axis, we produce a graph known as a fitness landscape in which points of high elevation correspond to high fitness.

To draw a landscape in which the trait corresponds to the length of a coyote's tail, we hold all other attributes of a coyote the same, vary the length of the tail, and measure the effect on fitness. To draw the graph, we need to know how the tail contributes to fitness. Suppose that a coyote's tail helps to balance the coyote when it jumps, and that it signals happiness, fear, or aggression. We begin at the left of the horizontal axis with a tail of length zero. It cannot carry out either function, so it has a fitness of zero. As the tail becomes longer both balance and signaling improve. Thus, fitness increases with tail length. At some point, say eighteen inches, the tail may be an ideal length for contributing to balance. If the tail becomes longer, the coyote will be less agile. Longer tails may continue to improve signaling value, so perhaps a tail of length twenty inches produces the most overall fitness. Once the tail becomes longer than twenty inches, fitness falls. The resulting graph, shown in figure 28.1, has a single peak.

Figure 28.1: A Mount Fuji Landscape

This landscape is known as a *Mount Fuji*. Such landscapes often occur in the real world. Mount Fuji problems are considered easy. We expect that evolution or learning will always find the peak when encountering one. Imagine a population of coyotes with different lengths of tails. Selective pressure would result in coyotes with tails of approximately twenty inches. Coyotes with tails of that length optimally blend balance and signaling. They have the highest fitness and produce the most offspring, resulting in more coyotes with twenty-inch tails. If we think of this as an optimization problem, we see that any hill-climbing algorithm would locate the peak.

We can apply one-to-many thinking and reinterpret this as a problem of product design—specifically, the problem of designing a coal shovel. Suppose that we have already decided on the length of the handle and the shape of the pan. The remaining design decision is how large to make the pan. Pan area will correspond to the trait on the horizontal axis. On the vertical axis, we plot how much coal a worker could shovel in an hour given that pan size.

As before, we start at the far left, which corresponds to a shovel pan with zero surface area. The technical term for a zero-surface-area shovel is "stick." A stick cannot shovel coal and has value zero. As we increase the pan—say, to the size of a teaspoon, then a tablespoon, then a toy shovel—we make the shovel more and more effective. The graph of shovel fitness slopes up. At some point, the pan area becomes too large. Lifting the shovel becomes a chore, and the amount of coal that a person can shovel in an hour decreases with further increases in pan size. When the pan area becomes sufficiently large, no one would be able to lift the shovel and fitness will be zero. Once again, we have a Mount Fuji landscape. And once again, we should expect to be able to find the peak, the ideal pan size for our shovel.

The idea of plotting the efficiency of shovels as a function of pan size to determine the optimal shovel was developed by Frederick Taylor. In the 1890s, Taylor and others ushered in an era of scientific management in which manufacturing decisions—how fast to move the assembly line, how strong to make the weld, how many breaks to give workers—were modeled as rugged-landscape problems. Many of the great industrialists of the twentieth century including Henry Ford, John D. Rockefeller, and Andrew Carnegie contributed to this movement toward efficiency, or what now is commonly called Taylorism.

The move away from artisans making individual and distinct products to large-scale manufacturing, in which processes were broken into parts and each part was optimized and then routinized, led to increases in efficiency but also, in the eyes of many, the dehumanization of labor. Herein lies a welcome reminder about the need for multiple models. Any single model simplifies the world and highlights only some dimensions. Scientific management models focused on process efficiency. This led to criticism. Making decisions based on efficiency of output caused other objectives, such as the happiness and well-being of workers, to fall by the wayside.

The landscape model may seem to be a relatively obvious idea: plot the fitness, efficiency, or value of a characteristic as a function of a trait or attribute and then climb up the hill to find the optimal amount of the trait. Thinking of solving a problem as climbing up a hill may also seem little

more than a metaphor. The validity of these critiques is not in question. However, by constructing a formal landscape model we will produce nontrivial insights.

Rugged Landscapes

When we allow for multiple attributes and for the contribution of one attribute to interact with those of others, we produce a rugged landscape— that is, a landscape with multiple peaks. Consider designing a couch in which we must choose the thickness of the cushions and the width of the arms. Let the value of a design equal its expected sales in the market, which correlates with aesthetic quality. If the couch has thick cushions, then wide arms may create a more appealing aesthetic. If the cushions are thin, then the ideal couch may have thin arms. A two-dimensional plot of expected sales as a function of arm length and cushion thickness will have two peaks. One peak corresponds to a couch with thin arms and thin cushions. The other will have thick arms and thick cushions.

Interdependent effects between variables create ruggedness on the landscape. Ruggedness has several implications. First, different approaches to finding the highest point on a rugged landscape may locate different peaks. So too may different starting points. Thus, ruggedness creates sensitivity to initial conditions and the possibility of path dependence. Each of these implies that landscape ruggedness contributes to outcome diversity. Ruggedness also implies the possibility of suboptimal outcomes. These are represented as *local peaks* on the landscape.

[Figure 28.2](#) shows a rugged landscape with five peaks. Four of these peaks are local peaks, points whose neighboring points all have lower values, and one is the *global peak,* the point with the highest value. To see how search could land on a local peak that depends on the initial point, imagine beginning from a point and then climbing uphill. This is known as a *gradient heuristic* or a *hill-climbing algorithm.* On a rugged landscape, gradient heuristics get stuck on a local peak.

If the starting point is at the far left, the gradient heuristic would locate local peak 1, which is not optimal. If the gradient heuristic starts in the

region denoted by Basin 2 in figure 28.2, then it locates local peak 2. Each of the other peaks, including the global peak, has a region such that if the gradient heuristic begins in that region, it will locate that local peak. These regions are called *basins of attraction* and are identified in figure 28.2. The global peak has the smallest basin of attraction. If we were to choose a random starting point and apply the gradient heuristic, the global peak is the least likely to be found.

The basins of attraction depend on the heuristic. A different heuristic may produce different basins. For example, consider the heuristic *go to the right* which moves to the right until finding a local peak. This heuristic produces identical local peaks as the gradient heuristic, but those peaks have different basins of attraction, as can be seen by comparing figure 28.3 with figure 28.2.

Figure 28.2: A Rugged Landscape with Five Peaks

To find an optimal or near-optimal peak on a rugged landscape requires either diversity or sophistication. The value of diversity should be self-evident. If distinct heuristics locate different peaks, then applying multiple, diverse heuristics to a problem will produce multiple, diverse local peaks, and one can choose the best from among these.[1] The same result will occur if one applies the same heuristic from different starting points: distinct local optima will be found and the best among them can be chosen.

Note also that the ruggedness of the landscape, as measured by the number of peaks, correlates with problem difficulty. However, a problem can be difficult to solve yet not have a rugged landscape. The problem of finding a gold coin in a cornfield would be represented by a flat landscape with a single peak at the coin's location. The landscape would not be rugged, but the coin would be very difficult to find.

The NK Model

We now describe the *NK model*, which allows us to formalize the connection between interactions and ruggedness.[2] The model represents objects, or what we might call alternative solutions, as binary strings of length N. The value of an object equals the sum of the contributions of each bit on the string. The K term in the model refers to the number of other bits that interact with each bit to determine its value. If K equals zero, the value function is linear. If K equals $N - 1$, then every bit interacts with every other and the value of each string is random. Thus, we can think of increasing K as tuning the ruggedness of the landscape to somewhere between Mount Fuji and random.

Figure 28.3: Basins of Attraction Produced by the "Go to the Right" Heuristic

NK Model

An **object** consists of N bits, $s \in \{0, 1\}^N$.

The **value** of an object is $V(s) = V_{k1}(s_1, \{s_{1k}\}) + V_{k2}(s_2, \{s_2 k\}) + \cdots + V_{k2}(s_2, \{s_{2k}\})$ where $\{s_{ik}\}$ equals a randomly selected set of k bits other than i, and $V_{k1}(s_1, \{s_{1k}\})$ is a random number drawn from the interval $[0, 1]$.

$K = 0$**:** Results in a linear function of the bits.

$K = N - 1$**:** Any bit change produces a new random contribution from each bit.

The NK model framework provides a wonderful space to explore ideas and ask questions. The first question we ask is how the number of local optima depends on the number of interaction terms. We then ask how the height of the global optimum depends on the number of interaction terms. At the moment, both of those questions are ill-posed because we have yet to define how we are searching the space of possibilities, that is what heuristic

we are using. Recall that what the set of peaks depend on our choice of heuristic.

In what follows, we rely on the *single-flip algorithm*. This algorithm chooses each attribute in sequence and switches the attribute's state. If changing this attribute results in a higher value, the switch is adopted. Otherwise the attribute is returned to its original state. The choice of this algorithm can be motivated in two ways. It can be interpreted as a crude model of genetic mutation, where good variants take over in the population and bad ones die. It is also the most natural way to represent a hill-climbing algorithm in this space.

We first evaluate the NK model with $N = 20$ and $K = 0$. When $K = 0$, each attribute's contribution to the total value is independent of the other attributes. The single-flip algorithm can identify the better state for each attribute and the global optimum. Thus $K = 0$, no interactions, corresponds to a Mount Fuji landscape. Each state's value is uniformly distributed in the interval [0, 1]. It can be shown that the higher of two random draws from a uniform distribution has an expected value of image. If we average the contributions across twenty attributes, the global optimum will also have an expected value of image.

At the other extreme, $(N = K - 1)$, each attribute is connected to every other attribute. When the state of an attribute is switched, the contribution of every attribute will change. It will be a new random number drawn uniformly from the interval [0, 1]. The value of the object will be the sum of these twenty new random numbers (one for each attribute), meaning that each flip of an attribute results in a value for the entire object that is uncorrelated with the earlier value. Thus, the landscape will be incredibly rugged—just as likely to go up at any point as it is to go down.

By applying that insight, we can derive the expected number of local peaks. If we start from any alternative, the single-flip algorithm compares that alternative to each of N alternatives. For example, starting from the alternative with all bits taking value zero, the algorithm will evaluate the N alternatives in which exactly one bit takes value one.

image

A local peak must have a higher value than each of these N alternatives. The probability that the original alternative has the highest value equals image. Therefore, the number of local peaks approximately equals the number of possible alternatives, 2^N, divided by N. For $N = 20$, that calculation yields fifty thousand local peaks. With so many local optima, the single-flip algorithm rarely locates the global peak.

The relevant question is not the number of these local optima, but their values. It remains only to compare the expected average value of these local optima with the expected value of the global optimum. That comparison will determine how well the single-flip algorithm performs. To calculate theses values, we can use the central limit theorem. It is not difficult to show that the expected value of a local optimum equals approximately 0.6 while the expected value of the global optimum equals a little more than 0.75.[3] Comparing these to the global optimum for the case $K = 0$, which equals image, reveals that the local peaks on the rugged landscape have lower values than on Mount Fuji, but the global peak has a higher value.

This begs the question of what happens in between these two extremes, as we increase the number of attribute interactions, K, from zero to $N - 1$. The answer is that we see both effects. The increased number of interactions produces a higher global peak, but more, and therefore lower-value, local peaks. Assuming that the search uses the single-flip algorithm, then computational investigations of this model show that for small K, the benefit of the interactions—a higher global peak—outweighs the increase in local peaks. So initially, the expected value of a local peak increases in K. The growing number of local peaks means that the average value will decrease. So if you were stuck using the single-flip algorithm, you would prefer a relatively small K value, around, say, 3 or 4. But why should we be constrained to use this simple heuristic of switching a single attribute? Evolution by mutation may be constrained to this heuristic, but we are not. We could switch the state of two attributes or even three. A more sophisticated algorithm will reduce the number of local optima.

Ruggedness and Dancing Landscapes

The NK model implies that we want a moderate degree of interdependence as that creates higher peaks. Many-model thinking demands that we step out from the particular assumptions of the model and consider the logic that drives that result. We find that the logic consists of two parts. The first rests on combinatorics: the number of pairs of combinations increases with the square of the number of pairs and the cube of the number of triples. Thus, interdependent effects create more possibilities of beneficial interactions. The second part rests on the fact that we need only keep the better combinations. Imagine grabbing any four food items to make a snack. Four items implies six possible combinations of two items. Suppose that we grab the following set of four: {*pickles, bananas, chicken, caramel*}. Of the resulting six pairs—bananas and pickles, pickles and chicken, caramel and pickles, bananas and chicken, caramel and bananas, and caramel and chicken—only one sounds remotely appealing. We only need choose that option. We enjoy the caramel bananas. We ignore the rest.[4]

A similar logic applies in evolutionary systems. Phenotypic combinations that produce positive interactions—a hard shell and short sturdy legs—remain in the population, while survival of the fittest works against combinations that produce negative interactions. We do not encounter many slow-footed, tasty animals with vibrant colors. If they ever existed, they have been caught and eaten.

We encountered a similar intuition in our model of search. When we have an abundance of possibilities, we prefer variation. The same logic applies here: combining pairs (and triples) produces abundant possibilities. And we would have preferred that these many possibilities had high variation in value. We then have a greater likelihood that one of them has a very high value. Given that interdependent effects increase variation, on the whole they are advantageous, but only up to a point. As we have just seen, too many make the landscape random. Ideally, then, we have a moderate number of interactions. Some argue that if the number and size of interactions can evolve or adapt, then systems should naturally evolve to rugged landscapes with high peaks. This would suggest that systems tend toward complexity and not equilibrium or randomness.[5] When and whether that is true is exactly the sort of question that is fun to explore with models.

One final point: We have taken the landscape as fixed. In ecological and social systems, the landscape that a species or firm confronts depends on the actions and attributes of others. An adaptation by a competing species, or a change in strategy by another firm, will shift and rearrange the fitness landscapes of competitors. We can now reinterpret our earlier models of spatial and hedonic competition as movements on dancing landscapes. Those movements could lead to an equilibrium, where each player stands on a local or global peak, or, competition on dancing landscapes may lead to complex patterns of actions and outcomes. Even a cursory glance at ecosystems, politics, and economics suggests that we see more of the latter. One reason that we see so much complexity may well be that much of our world consists of adaptive and purposive actors maneuvering on dancing landscapes. To make sense of that complexity, we need many models.

Do We Patent Knowledge?

Our well-being rests atop a centuries-long accumulation of knowledge that includes the laws of physics, the combustion engine, double-entry accounting, the germ theory of disease, X-rays, and HTML. Knowledge is often a public good. Knowledge is always non-rival. It may or may not be excludable. Exclusion requires verification, which is easier when a physical artifact embeds the knowledge. Verifying that someone used an algorithm or technique for solving a problem may be impossible. Verifying that someone embedded that algorithm in a software program is not.

When knowledge can be excluded, we confront a choice. We could treat knowledge like roads and national defense, and tax people to fund its production. Governments pay people to think through grants and indirectly by supporting universities. Governments also allow people to patent knowledge. A patent creates an incentive to produce knowledge by creating a period of exclusive rights to use the knowledge and to charge others for its use. In the United States and Europe, patents last twenty years from the date of filing.[6] Advocates of patents argue that private entities would have little incentive to spend years developing a better mousetrap, computer algorithm, or sound system if their discoveries could be used by anyone for

free. They argue that patents overcome the incentive problems inherent in knowledge production.

Boldrin and Levine construct an argument that borrows ideas from several models to argue against patents.[7] In models in which ideas can be combined, patents can hinder innovation by limiting recombinations. One firm's patent on touch-screen technology may reduce the number of other firms that design new products that incorporate that technology. Absent patent protection, the technology could be incorporated into more products. Innovation would increase.

Proponents of patents push back by noting that while slowing innovation may be bad, without patents the reduction in investment would be much larger. To counter that claim, Boldrin and Levine use a logic partly based on our diffusion model. A useful product based on new knowledge will spread quickly through the population of buyers. That was true of the radio, television, Google's search engine, and Facebook. This creates a *first-mover advantage*. The innovator can still benefit, but only by producing something with the idea. With a patent, an inventor can wait for others to implement the idea and profit.

Boldrin and Levine also question how much credit the inventor deserves anyway. If breakthroughs were the result of a solitary genius, and most ideas would never have been produced without incentives, then the case for patents is stronger. The rugged landscape model suggests that difficult problems may have multiple workable solutions. New inventions, particularly those that combine existing ideas and technology such as the car, the telephone, and online auctions, may be natural occurrences not acts of genius. Any number of people might have made these innovations given the ideas swirling around in the community of thinkers. The simultaneity of major discoveries—calculus (Isaac Newton and Gottfried Leibniz), the telephone (Alexander Graham Bell and Elisha Gray), and the natural selection theory of evolution (Charles Darwin and Alfred Russel Wallace) —supports that inference. In sum, many-model thinking shows advantages and disadvantages to patents. The deeper, more nuanced understanding the models provide argues for a more flexible patent policy. Perhaps some ideas—those that many people might have discovered and those that could

recombine with other ideas—should have different lengths or types of patents, or even not be patentable at all.