

Reflections on Researching the Rugged Fitness Landscape

T. Grandon Gill

University of South Florida, Tampa, Florida, USA

ggill@coba.usf.edu

Abstract

The success of an informing system depends upon achieving a fit between multiple entities: the sender, the client, the delivery system, and the task to be performed. Conceptually, the effectiveness of such informing can be modeled as a fitness landscape—a function that maps the characteristics of the system into a scalar fitness value. The assumed shape of that landscape plays a major role in determining how we go about researching a given domain. In the social sciences, we often make the implicit assumption that the processes that we are investigating are largely decomposable, meaning that the impact of a given characteristic on fitness is independent of the values of other characteristics. Many of the statistical tools employed in analyzing social science data, such as regression and structural equation modeling, implicitly depend on such decomposability. In other disciplines, such as evolutionary biology and medicine, the assumption of decomposability is much less prevalent. Instead, the prevalence of interactions between characteristics affecting fitness is taken to be a fact of life, leading to what is called a rugged fitness landscape. This paper explores the nature of such landscapes, the likelihood that they will be encountered in the context of informing systems research, and the implications of ruggedness on how we approach research design.

Keywords: research methods, rigor, relevance, rugged fitness landscapes, adaptation, generalizability, informing sciences, complexity, chaos, decomposability.

Introduction

At the core of the informing sciences is the need to better understand the interplay between the components of an informing system: the sender, the client, the delivery system, and the task to be performed. Implicit in this understanding is a notion of *fit*. For example, if the client is in a rural area of a developing country served only by telephone lines that are subject to intermittent failure and the sender wishes to convey large amounts of information in video format, then the use of an Internet-based delivery system is unlikely to be a good fit with the informing need. Similarly, if the task to be performed is highly interactive and requires ongoing exchange of information with

the client, then a system providing two-way communication is likely to be a better fit than a system that only allows one-way broadcast of information. If we attach a numerical or ordinal value to the level of fit, we can refer to that value as a *fitness* value. If we were to consider a whole variety of different possible informing systems for achieving the same purpose and were to attach a fitness value to each one, we have the beginnings

Material published as part of this publication, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact Publisher@InformingScience.org to request redistribution permission.

of a *fitness landscape*. To develop a complete fitness landscape, we would need to develop a function that can take any combination of informing system characteristics and map them to an associated fitness value. The optimal possible system would then be that collection of characteristics that maps to the highest fitness value. It is probably not an exaggeration to assert that most research in the informing sciences is motivated by the desire to contribute, directly or indirectly, to our understanding of the fitness landscape associated with informing.

In our goal to better understand the fitness landscape for informing systems, it is useful to learn from other disciplines—many of which have their own versions of the fitness function. In economics, for example, consumers strive to maximize utility—a function that maps the bundle of goods and services they consume to a satisfaction-related value. Producers are frequently modeled as attempting to maximize shareholder value, another measure of fitness. In computer science, the concept of fitness is routinely employed in evaluator functions. A chess program, for example, will normally choose its move based upon assessing the fitness of the alternative board positions that may result. Genetic algorithms use reproductive and mutation rules originally observed in natural systems in an effort to seek solutions of maximal fitness. The concept of fitness is particularly prevalent in biological sciences. Evolutionary biologists, for example, view fitness as a survivability function representing the likelihood of reproductive success that may be applied at many different levels—from the gene to an entire species. If fitness is insufficient, the gene or species ultimately disappears if it fails to evolve or adapt.

The present paper is intended to as a non-mathematical introduction to the nature of fitness landscapes, with particular attention being paid to the potential implications for research in the informing sciences. It begins by presenting a continuum used to characterize fitness landscapes that was first introduced in evolutionary biology (e.g., Kauffman, 1993). In this model, landscapes range from decomposable to rugged to chaotic. The paper then demonstrates how the continuum strongly resembles another continuum: that of science and art. The remainder of the paper focuses specifically on rugged landscapes, emphasizing two main themes. First, it argues that the conditions that are likely to lead to a rugged fitness landscape are nearly always going to be present in informing systems. Second, it considers the many ways in which rigorous research conducted on a rugged fitness landscape can—or, more precisely, should—differ from research that assumes underlying decomposability.

Fitness Functions and Landscapes

A fitness function serves to map a set of attributes into a single value that is indicative of the desirability of the particular combination. Conceptually, this function can be represented as:

$$F = f(x_1, x_2, \dots, x_N)$$

Where F is the fitness associated with a particular combination of specific values for the attributes x_1 through x_N . The term fitness landscape is used to refer to the behavior of the fitness function across the set of all possible values of its attributes. Conceptually, this corresponds to the “shape” of the function.

The desirability aspect of a fitness function typically manifests itself in one or both of two ways:

1. *It may signify the survivability of a particular attribute combination.* In biology and in genetic algorithms, for example, entities with higher fitness values are more likely to survive from one generation to the next than those with much lower values.
2. *It may serve to guide choice.* In economics, for example, an underlying axiom of individual behavior involves choosing that basket of goods and services which maximizes utility, which is to say the fitness of the combination.

The differences between the two forms of fitness may be less significant than might first appear. Evolutionary economists, for example, argue that our utility preferences are, in fact, simply the evolved manifestation of characteristics that have—at least in the past—contributed to individual survival (Gandolfi, Gandolfi, & Barash, 2002). Similarly, the survivability of a particular product is likely to depend heavily on the utility it inspires in its prospective customers.

The simplicity of our conceptual representation of a fitness function should not be taken as suggesting that such functions are simple. To the contrary, beyond the matter of how the function behaves—which, as we shall find, can be quite complex—finding a suitable representation for the attributes being considered is, by no means, a trivial matter. Suppose, for example, we wanted to construct the fitness function for a particular recipe that characterized—based upon the arguments supplied—how tasty the resulting dish would be. Among the elements we would need to represent are included:

- The nature and quantity of the ingredients
- The timing of insertion of the ingredients
- The specific actions that we would need to perform upon those ingredients
- The timing of those actions, and
- The tools and equipment required.

Furthermore, among those attributes which are quantitative in character, such as ingredient amounts in our example, the relationship between fitness (taste) and quantity is unlikely to be linear. For example, as shown in Figure 1, it is likely that some optimal amount of an ingredient, such as a sugar, will be present. Either more or less than that amount will lead to lower fitness—meaning the resulting dish will be less tasty.

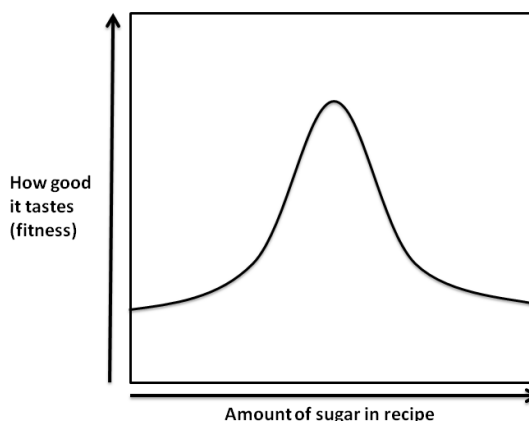


Figure 1: Mapping between taste and amount of sugar in a hypothetical recipe

The shape of the curve presented in Figure 1 is our first example of a fitness peak. If we are trying to maximize the taste value of a 1 kilogram cake (as assessed by some specified individual) we would anticipate that every ingredient (i.e., argument to the fitness function) would exert a qualitatively similar influence on fitness, as would other continuous measures such as oven temperature. If we had only two ingredients, the resulting fitness space would look like a mountain with the peak representing the optimal combination of the two ingredients. With more ingredients, the precise shape of the function is harder to visualize but, conceptually, can still be thought of as a peak.

The recipe example is also useful because it can be used to illustrate the notion of migration to peak fitness. Presuming the recipe has been taken from a time-tested cookbook, there is an excellent chance that many different versions of the recipe—featuring different proportions of the relevant ingredients and different treatments (e.g., baking times, temperatures)—were tested and the one chosen for publication reflected that combination maximizing fitness as assessed by the recipe testers. In the broader fitness landscape of all recipes, we would then say that our particular recipe occupies a local fitness peak. This means that while other recipes may have higher fitness (our testers may actually prefer chocolate cake to yellow cake), there is no incremental change to our yellow cake recipe that makes it more fit—all such changes reduce fitness. In many disciplines, such as economics, it is axiomatic that observed decisions collected from a fitness landscape represent the results of individual attempts to maximize fitness.

Landscape Shapes

The preceding recipe example demonstrates how a fitness landscape can have a peak. That is, however, only one type of peak that may be present in a fitness landscape. To better understand landscape shape, we need to examine the concepts of decomposability and ruggedness.

The decomposability of a fitness function defines the degree to which the impact of each individual attribute upon fitness is independent of the values of the other attributes. For example, suppose a particular fitness function can be represented as:

$$F = f(x_1, x_2, \dots, x_N)$$

That function is fully decomposable if we can also represent it as:

$$F = y_1(x_1) + y_2(x_2) + \dots + y_N(x_N)$$

where y_1 through y_N are functions that transform the raw x values into their marginal contribution to fitness. For example, the y_{SUGAR} in the recipe example would take into account the peaked shape of sugar's impact on taste. For simplicity's sake, we can abbreviate the functions in a decomposable landscape as follows:

$$F = y_1 + y_2 + \dots + y_N$$

At the other extreme, a fitness function may be completely non-decomposable—leading to a *maximally rugged fitness landscape*, later referred to as a *chaotic landscape*. What this means is that the contribution of a particular attribute to fitness cannot be determined without knowing the value of the other attributes. To use a somewhat contrived example as an illustration, suppose you were a participant in a game show where you could win: 1) a week-long vacation at a ski resort OR a week-long vacation at an ocean resort (each valued at \$2000), 2) a plane ticket to the ski resort OR a plane ticket to the ocean resort (each valued at \$750), 3) a week long rental of ski gear OR a week long rental of scuba gear (each valued at \$400), and 4) \$800 in cash. In addition, for the sake of the example, assume that you are not able to trade or sell any of your prizes AND that you were unwilling to spend any of your existing resources during the vacation. In this case, you might have a utility function such as the one presented below:

Table 1: Non-Decomposable Utility Matrix

0 = Ski Resort 1 = Beach Resort	0 = Ticket to Mountains 1 = Ticket to Ocean	0 = Ski Gear 1 = Scuba Gear	Utility
0	0	0	1.0
0	0	1	0.75
0	1	0	0.5
0	1	1	0.25
1	0	0	0.25
1	0	1	0.48
1	1	0	0.73
1	1	1	0.95

In the top and bottom cases, perfect fit is achieved and utility is maximized (two peaks), with skiing being slightly preferred. In the case where only the gear is mismatched, positive utility is achieved that is reduced by the need to spend prize money to cover the gear costs. Where resort and gear are consistent, utility is still lower since nearly all your cash would be drained by the plane fare. In the remaining two cases, where the plane ticket matches the gear but does not match resort, the prize has a value equal only to the utility resulting from the amount of cash pro-

vided since you cannot afford the resort and are therefore unable to take advantage of any of the other prizes.

Kauffman's NK Model

Kauffman's (1993) NK landscape model provides a tool for characterizing the decomposability of fitness functions such as the example presented in Table 1. Originally developed to simulate the fitness of a chromosome, the N refers to the number of genes. The K, in turn, refers to the average number of other genes whose values must be ascertained before the contribution of a particular gene to fitness can be determined. It is, therefore, a measure of interdependence between arguments. The model has two extreme points:

- N, 0: At this point each gene contributes to fitness independently, leading to a *fully decomposable landscape*.
- N, N-1: At this point the impact of a given characteristic on fitness can only be determined by considering the value of every other characteristic. As a consequence of this complete interdependency, no meaningful estimate of fitness can be made without knowing the values of all N characteristics. We'll refer to this landscape as the *chaotic landscape*.

For a decomposable (N,0) fitness landscape, there will be a single fitness peak at which point the fitness values y_1 through y_N are individually maximized. Somewhat less immediately obvious, and at the other extreme, the chaotic (N, N-1) fitness landscape can, for all intents and purposes, be modeled as a set of random numbers (Kauffman, 1993), thereby ensuring that no separable relationships between a subset of elements and fitness are likely to occur.

To fully understand the nature of the chaotic landscape, we need to realize that in order to model it *accurately* with a tool such as multiple linear regression, we would need to create a separate interaction term for every possible combination of values—meaning that there would be $2^N - 1$ coefficients plus a constant. Assuming that we had enough observations so that there was at least one in every cell and assuming minimal error, we would then be able to estimate the fitness of each cell in the landscape. If, on the other hand, we attempt to fit that N,N-1 landscape with a decomposable model (i.e., with N coefficients), the only significances that *should be* observed would be coincidental.

If a chaotic fitness landscape is modeled as a field of random numbers, it follows mathematically that such landscapes will necessarily have a large number of local fitness peaks (i.e., combinations of x_1, x_2, \dots, x_N where changing any single value will lead to a decline in fitness). Specifically, when only moves to adjacent fitness sites values are considered, the estimated number of these peaks will be given by a formula (Kauffman, 1993, p. 47):

$$2^N / (N+1)$$

An example of this is presented in Figure 2, illustrating an NK space of dimension 6,5. The peak locations are indicated by the values in parentheses. To interpret this figure, the combination of the left column and top row represent the 6 argument values to the fitness function, while the values in the central cells provide the related (randomly generated) fitness values. For example:

$$F(0,1,1,1,0,1) \equiv .771, \text{ which happens to be a peak}$$

To determine that a value is a peak, the six adjacent cells need to be examined. In the case of 011101 (commas omitted), these would be 111101 (0.114), 001101 (0.198), 010101 (0.181), 011001 (0.658), 011111 (0.059), and 011100 (0.090). The estimated number of peaks for a 6,5 NK landscape would be roughly 9 ($2^6 / (6+1) = 64/7$), with the actual number found in the example being 10. (A more detailed look at constructing a rugged fitness landscape using a spreadsheet is

presented in Gill & Sincich, 2008). As the value of K declines, the number of peaks in the landscape will also decline until, when K=0, a single peak remains.

	0,0,0	0,0,1	0,1,0	0,1,1	1,0,0	1,0,1	1,1,0	1,1,1
0,0,0	0.811	0.058	0.265	0.433	(.865)	0.462	0.352	0.258
0,0,1	0.756	0.658	0.720	0.753	0.154	0.198	0.170	0.362
0,1,0	0.143	0.518	0.488	0.503	(.905)	0.181	0.868	(.937)
0,1,1	0.729	0.057	(.795)	0.230	0.090	(.771)	0.084	0.059
1,0,0	0.239	0.841	0.820	0.656	0.413	0.559	(.96)	0.581
1,0,1	0.178	(.852)	0.770	0.842	0.468	0.843	0.672	0.162
1,1,0	(.807)	0.263	0.776	0.120	0.252	(.728)	0.228	0.935
1,1,1	0.495	0.115	0.446	(.926)	0.219	0.114	(.679)	0.328

Figure 2: Local fitness peaks (in parentheses) on a randomly generated NK fitness landscape of dimension 6,5. Row headers indicate the first three attribute values, column headers the last three.

Increasing Fitness

An obvious motivation for conducting research that leads to a better understanding of a given fitness landscape is to identify actions that will lead to improved fitness. Landscape shape in general—and the number of local fitness peaks in particular—exert a major influence on what we can reasonably expect to achieve from such research.

Where the fitness landscape is decomposable, the process by which fitness can be increased is relatively straightforward. Because each characteristic that contributes to fitness does so independently, we can examine each characteristic independently *and*, once we understand how that characteristic contributes to fitness, we will not need to revisit it.

To provide a concrete example, suppose a national magazine publishes a list that ranks universities. Further suppose, as is commonly the case, that such a list is constructed by taking a set of attributes (x_1 through x_N), each weighted by some undisclosed factor (a_i) and then summed, to compute a score (S) for every university, e.g.,

$$S = a_1x_1 + a_2x_2 + \dots + a_Nx_N$$

This formula would obviously meet the criteria of a fully decomposable fitness function since the contribution of a particular attribute—e.g., whether or not the institution has a football team—would be the same amount no matter what other attributes the institution has or doesn't have. To determine the impact of the football team attribute, an institution would simply need to add a football team and see what happened to its ranking. If fielding a team seemed too expensive for the sake of information gathering, the researcher might attempt to find two institutions whose characteristics differed only in the presence/absence of a football team and observe how their rankings differed on the list. If such a comparison were not available, the researcher might gather the characteristics of a sample of universities and then use a statistical tool—such as multiple linear regression analysis—to estimate the coefficient weights for each characteristic, including the presence/absence of a football team. Any of these techniques would work because each attribute's impact is entirely independent of the impact of any other attribute.

Even where a fitness function is fully decomposable, the process of achieving optimal fitness may prove far from trivial. There may, for example, be constraints—such as the availability of budgetary resources—that prevent an entity from optimizing all attributes simultaneously. The fitness benefits of a football team may need to be weighed against the construction of a new science building, since sufficient funds are not available for both. Where decomposability is present, however, algorithmic approaches to optimizing fitness—such as linear or integer programming—

are often available that dramatically reduce the amount of searching that must be done in order to optimize fitness given a particular set of constraints. Thus, it is acquiring an understanding of the underlying nature of the global fitness function that is the principal challenge in a decomposable landscape; once the function is understood, the process of moving towards high, indeed optimal, fitness can be relatively mechanical.

The situation is entirely different for a rugged landscape. As the number of characteristics relevant to fitness grows, algorithmic or exhaustive search becomes impossible. There are simply too many possible combinations. To justify this assertion, let us imagine that the fitness of a university was, in fact, governed by a chaotic fitness function. As it happens, in the late 1980s the author developed college search software that attempted to capture the main attributes of each institution. Roughly 500 bits were used to encode relevant characteristics, such as the presence or absence of several hundred undergraduate majors, the presence or absence of different sports, average entering test scores, geographic location, campus setting, and size. If a chaotic fitness function existed, the landscape would be 500,499 in NK model terms, meaning that 2^{500} different combinations could be generated, each with its own fitness, and roughly 2^{491} ($\sim 10^{150}$) local fitness peaks could be anticipated based on the formula $2^N/N+1$. Since both these numbers are unimaginably larger than the number of atoms in the universe, any attempt to search the fitness landscape by brute force would be certain to fail. But, in a chaotic landscape, brute force is the only technique guaranteeing that a fitness maximum will be achieved. Thus, the best one can hope to achieve is a deep understanding of how fitness behaves near a particular local peak or a small number of alternative peaks.

Even when the value of K is much closer to 0 than it is to N, we quickly reach a point where achieving fitness increases through manipulating attributes becomes much more difficult than it is for the fully decomposable case. Continuing with our preceding example, suppose university rankings were computed in an entirely different manner: by estimating the number of students for whom each institution represented the “best choice”, subsequently referred to as best choice fitness. Although no existing college ranking system—to the best of the author’s knowledge—is actually constructed in this manner, we may reasonably speculate as to how it might behave and as to how it would differ from more traditional rankings. To begin with, the fitness score for a university would necessarily depend upon its ability to best meet the needs of particular subsets of the population of potential students. Many attributes that would independently contribute to fitness in traditional ranking systems, such as scores on standardized college entry tests, would impact the “best choice” metric differently. For example, elite universities often boast S.A.T. scores in the top 1% of all students; under the “best choice” system such a criterion would eliminate 99% of the population of potential students from consideration, potentially reducing fitness. Similarly, some attributes that would not impact rankings under normal systems, such as whether an institution was liberal or conservative in its political leanings, might exert a great impact on determining whether or not it was a good fit for a particular student—for one group of students a left-wing outlook might increase fit, while for others it would reduce it. Another aspect of the best choice ranking system would be a strong motivation for institutions to migrate towards customized missions that target particular clusters of students. In the traditional ranking system, universities would tend to maximize fitness in the same way, since fitness for each is determined using the same set of attributes. Under best choice fitness, institutions would benefit from continuously searching for large subsets of the potential student universe whose needs were not being well met by other universities that also happened to be a reasonably close match to the university’s existing characteristics.

It should be evident that the principal challenge presented by best choice fitness stems from the lack of decomposability in the relationship between institutional attributes and student characteristics. For example, universities targeting students who want to leave home for college might do

best by investing money in dormitories and emphasizing classes that meet during the day so as to match the desires of full-time students; universities targeting the local population, particularly the local working population, might emphasize night classes and part time programs. In regions serving the economically disadvantaged, low cost community colleges might exhibit far greater fitness than more traditional, higher-priced schools that focus on a national pool of affluent students by providing luxurious surroundings at a high price tag. On the other hand, because we assume the landscape is not chaotic, and that K is probably much closer to 0 than to N , we may also reasonably expect that some attributes may contribute to fitness relatively independently of other attributes. The ephemeral “quality of teaching” might be an example of such an attribute—although a counter-argument could also be made that teaching quality will have less of an impact on fitness in institutions whose target clientele is high achieving self-motivated students than for institutions targeting first-in-family college students.

Increasing fitness in a rugged landscape is vastly more difficult than it is for decomposable landscapes. This difficulty stems from two sources. First, it is much harder to assess if fitness changes that are observed in one entity are going to generalize to another. For example, the observed positive fitness impact of adding a football program to small full time college may be quite different than it would be for a large urban university whose student base consists mainly of part time commuters. Second, although experimenting with individual characteristics one at a time can lead to incremental fitness increases, eventually the entity will reach a point at which every change leads to declines in fitness. At that point, the entity has reached a *local fitness peak*. The problem is that such peaks will not necessarily exist at very high fitness levels. The analogy here is that of climbing a mountain by ensuring that every step you take is in a direction that leads to an increase in altitude. Eventually, you will reach a peak. It may, however, be the top of a foothill rather than a mountain summit. Returning to our university illustration, some missions—no matter how perfectly they are carried out—may never be the best fit for a significant number of students. Thus, if acceptable fitness levels are to be achieved, an entity may, from time to time, need to consciously jump from one place in the landscape to another in the quest for alternative (higher) fitness peaks.

A variety of mechanisms for changing attributes in the quest to increase fitness have been observed in the study of genetics. Many of these, including sexual reproduction, cross-over, inversion and mutation, have been adapted to other search situations, such as genetic algorithms (Holland, 1992). The challenge of designing such search is in balancing the need to increase fitness through incremental changes (e.g., mutation) with the need to continue a wider exploration of the landscape so as to avoid settling for an unacceptably low peak (e.g., sexual reproduction, cross-over, and inversion). In addition, analogous to the peaks in a mountain range, the process of moving incrementally from one local peak to another often requires transiting through values of lower fitness. In some cases, these valleys in fitness may be so low that the entity fails to survive the trip.

Dynamic Fitness Landscapes

Up to this point, the model presented assumes that attempts to increase fitness are taking place on a fixed landscape. This would, of course, be a severe limitation for any model intended to be realistic. Transformations to the fitness landscape are presumed to come from two sources: adaptation of existing entities on the landscape and interaction with coupled, coevolving landscapes. We will now consider each of these.

Although we have treated fitness as a function of a set of situational characteristics, in many cases it will also depend on the characteristics of other entities occupying the same fitness landscape. This is likely to be particularly true for rugged landscapes. For example, the number of students for whom a university is the “best choice” will depend not only upon the characteristics of the university but also upon what other universities are attempting to attract the same target group of

students. In such a fitness space, we would expect continual adaptation and refinement by the associated entities. In some cases, an equilibrium state might be achieved where all entities reach local peaks. In others, adaptive cycling might continue indefinitely.

A considerable number of examples, drawn from many domains, suggest that the adaptation process generally follows a characteristic pattern: extended periods of relative stability interrupted by short bursts of rapid transformation. Per Bak, a colleague of Kauffman's, studied the phenomenon extensively, referring to it as punctuated equilibrium (Bak, 1996). The magnitude and frequency of the sharp transitions are distributed roughly according to a power law (the logarithm of magnitude plotted against the logarithm of frequency is a straight line), whereby smaller disruptions occur more frequently than larger disruptions. Such a law appears to govern many unrelated phenomena, such as avalanches on a sand pile, earthquakes and biological extinction events. A number of researchers in business have observed a similar pattern of sharp changes in business environments as well (e.g., Gersick, 1991; Gill, 1995; Handy, 1990). As a general rule, the frequency of transitions tends to grow with the ruggedness of the fitness landscape. Thus, we would expect that the ability of entities to adapt would contribute more significantly to survival on such landscapes.

Kauffman's (1993) NK landscape model takes another view of dynamic fitness landscapes, attributing changes principally to coevolution. The coevolution model assumes two or more separate fitness landscapes are coupled together as a result of interdependencies. For example, the fitness landscape of a particular plant species might depend upon the population of a particular animal that eats its fruit and later evacuates its seeds in distant locations. The fitness of that animal species might, in turn, depend upon the availability of that plant species. Thus, they might contribute to each other in a synergistic way. It might also be the case, however, that the same animal will chew on the plant's leaves—thereby reducing its fitness—if insufficient fruit is available. Thus, a delicate dance between the fitness of one landscape and another could ensue.

In our university example, demographic changes that result from generational decisions regarding whether or not to have children can dramatically impact the size of the student pool. Economic changes, both relating to geographic regions and impacting parental willingness to invest in the education of such children, would similarly impact the best choice fitness metric. Thus, a university that pursues best choice fitness would face a continuously changing landscape. In order to survive, it once again follows that adaptability would be an important characteristic for such an institution to cultivate.

Decomposable landscapes tend to be much more stable than rugged landscapes. To take a simple example, suppose a farmer is trying to select a seed variety for the next year's planting. A number of important characteristics will impact the fitness of a particular seed choice: expected yield, disease resistance, the projected quality of the crop (e.g., protein content), the expected cost of inputs (e.g., fertilizer, pesticides), and so forth. These factors could each be accounted for independently in a profit equation (i.e., fitness) and the "best" seed could then be selected. Assuming farmers in a particular area were limited to a single crop, it would make no difference whether or not all your neighbors also planted the same variety of seed. Although the abundance of yield might impact price adversely—therefore lowering the fitness of the seed—you would still be better off planting the best seed as opposed to a lesser seed. Similarly, if overall demand for the crop dropped (co-evolution), you would still be better off with the best seed than with some other seed associated with the same crop. Although it is true that certain co-evolutionary effects could change fitness to another seed—for example, a sharp spike in input costs could make a seed with lower input requirements and lesser yield the new optimal choice—the same factors that impact one farmer would impact all, leading to a mass migration to the new seed during the subsequent planting season. The orderly nature of the decomposable fitness landscape comes with a hidden cost, however. The fact that all entities are drawn to a single peak means that should transforma-

tions occur that dramatically reduce the fitness of that peak, the survival rate of entities may be very low. In our agricultural example, depending solely on a single seed variety or crop increases vulnerability to an unexpected loss of fitness specific to the choice. An example of the price to be paid for such dependence can be found in the Irish Potato Famine of the mid-1800s, where extreme dependence on a single crop combined with the emergence of a disease that attacked the same crop led to a humanitarian disaster of extraordinary magnitude.

The example just presented illustrates a central theme in Kauffman's (1993) evolutionary theory. Specifically, where a fitness landscape is likely to be subject to transformations, a fundamental tradeoff between efficiency and adaptability is likely to emerge. As we have already seen, in NK landscapes where $K \rightarrow N$, the number of peaks increases; it is also necessarily true that the average fitness of those peaks declines. As a result, entities migrating to peaks in highly chaotic landscapes are unlikely to be particularly fit, adversely affecting the survivability of such landscapes. On the other hand, where significant internal or external forces transform the landscape, the widespread distribution of peaks means that many entities will already be close to post-transformation peaks. Thus, the landscape is very adaptable.

In the other direction, as decomposability increases (i.e., $K \rightarrow 0$), so does the average fitness of entities on peaks—the difference between peak and average fitness tending to be greatest in the single peak case of $K=0$. Thus, the ability of entities to achieve peak efficiency tends to be maximized in decomposable landscapes. If a major external force transforms the landscape, however, entities that were formerly at the peak may find themselves a great distance from the post-transformation peak or peaks. That means that there is a strong likelihood that local migration will lead to a suboptimal peak or will require a long migration through low fitness values in order to reach a suitable peak. In nature, and in business, such long migrations to a totally new region of the fitness landscape are likely to be accompanied by considerable risk of extinction. Thus, entities existing on landscapes where $K \rightarrow 0$ will tend to exhibit low adaptability.

What Kauffman (1993) proposes is that for a gene or species to evolve successfully it must achieve a balance between efficiency and adaptability—a boundary he refers to as the edge of order and chaos. For a particular entity, this balance will be impacted by the rate of change in coupled environments that are coevolving. Where coevolution is slow, the boundary moves towards decomposability; where coevolution is rapid, a more rugged landscape is favored.

The Fitness Continuum from Science to Art

The concept of a fitness landscape can also be useful in considering how science and art can be distinguished. Both science and art can benefit from research—the nature and objectives of that research differ significantly, however. This becomes important when we later consider how ruggedness might impact the manner in which we design research within the informing sciences.

As we have already emphasized, fitness is often interpreted in terms of the likelihood that an entity will survive from generation to generation. There is no particular reason that the relevant entity could not be a theory or equation. Viewed in this context, nearly all research in the “hard sciences” can be cast in fitness terms, where an important contributor to the fitness of a particular theory is likely to be the degree to which its predictions match observed outcomes. A simplistic expression of such fitness might be:

$$F = -|O(x_1, x_2, \dots, x_N) - P(x_1, x_2, \dots, x_N)|$$

where F is the fitness of our theory, O is an observed outcome for attributes x_1, x_2, \dots, x_N and P is the theory-predicted outcome for attributes x_1, x_2, \dots, x_N . Theories that are highly fit will tend to produce differences that are consistently 0 or very small. Additional contributors to theory fitness would include:

- *Compactness*: for a given level of predictive accuracy, smaller numbers of variables (N , in the above notation) would be more fit than larger numbers. This is, essentially, a re-statement of the principle known as Occam's Razor.
- *Robustness*: The greater the domain of predictive accuracy (i.e., the range of x_1 , through x_N over which error is low), the more fit the theory.
- *Reliability*: The degree to which a given set of values for x_1, x_2, \dots, x_N produces consistent predictions. Where consistency is not present the implication would be that x_1, x_2, \dots, x_N may never be sufficient to explain observed behaviors. Challenges to reliability often stem from inability to estimate or measure tacit attributes accurately. They may also be the result of embedded characteristics within the theory such as irreducible uncertainty (e.g., in quantum mechanics electron positions are described by a probability cloud) or sensitive dependence on initial conditions (e.g., chaos theory).

Although it would be a serious mistake to underestimate the impact of social forces in the adoption of a particular theory (e.g., Kuhn, 1970), once established a highly fit theory can be expected to survive for a very long time.

The source of a particular theory in the sciences may be a mathematical derivation or a synthesis of empirical observations. Sometimes, in fields such as economics, it is a combination of the two—with the value of unknown parameters appearing in mathematical formulations being estimated through statistical techniques (e.g., multiple regression, structural equation modeling, and factor analysis) applied to observations. The suitability of such techniques for analyzing rugged fitness landscapes will be summarized later in the paper and is the subject of the companion paper (Gill & Sincich, 2008).

The concept of a fitness landscape is as applicable to works of art as it is to scientific equations. Some works of art survive for centuries, even millennia. Others are quickly discarded or forgotten. Interestingly, it is relatively easy to describe the functional form and arguments for many types of original art works. For example:

- *Marble Sculpture*: The arguments to the fitness function would be an array of three dimensional pixels, of a resolution just below the threshold of the human eye's ability to detect granularity, each of which is a bit for which 0 is no rock, 1 is rock. For such a function, N might be $10,000 \times 10,000 \times 10,000$, or 10^{12} .
- *Oil Painting*: The argument of the fitness function would be two dimensional pixels, each of which would have a color attribute (20 bits would probably be sufficient) and a depth attribute (to allow textures to be represented, perhaps another 10 bits). For such a function, N might be $10,000 \times 10,000 \times 30$, or 3×10^9 .
- *A Simple Tune*. 32 measures (2^5) could be time sliced into 64^{th} intervals (2^6), each of which might have 16 (2^4) possible pitch values (so as not to exceed the vocal range of inexperienced singers) plus some additional attributes signifying how the note was to be played and if it was a continuation of the previous note (2^6). For such a function, N might be $2^5 \times 2^6 \times 2^4 \times 2^6 = \sim 2 \times 10^6$.

In considering these functions, there are a number of important points that need to be emphasized. First, the N values specified represent the number of arguments, not the number of combinations. The number of possible combinations is 2^N , creating a set of possibilities that is infinite in the practical sense. Second, the value of K is definitely not 0—that would imply, for example, that there is an “optimal” color for each dot in a painting that does not change from subject to subject. Indeed, the level of interdependence between arguments is probably quite high. If Leonardo Da Vinci had decided to make one of the Mona Lisa's eyes blue and one brown, it is doubtful that the

painting would have survived the test of time as well as it has in its more consistent form. Similarly, even if you have never heard a tune before, it is often possible to detect a sour note that detracts from the overall performance.

The effectively infinite number of possible works of art in each category is not the only factor that inhibits the development of predictive scientific theory in the arts. Replicability, as previously noted, is an important contributor to the fitness of theory. In the arts, however, differing tastes across individuals leads to high variability in observed fitness for a given work—and taste changes over time can substantially influence the acceptability of an individual work. A composer writing in the Baroque style would have far more difficulty finding broad acceptance today than in the Baroque period, although it is hard to argue that the evolution from Baroque music to grunge rock actually represents an improvement in any objective measure of fitness. Moreover, the artistic fitness functions previously described apply only to original works of art; a perfect copy of a masterpiece will generally exhibit far lower fitness than the original. Thus, for the fitness function to be complete, it would need to include arguments identifying all previously created works in the same category. A history of past creations would also be useful because, under many circumstances, originality may be an important contributor to the fitness of an art work.

For all the reasons stated, no serious researcher is ever likely to attempt to determine the precise fitness function for any category of art. That does not mean, however, that research and theory do not play a role in the arts. Rather, it means that such theory tends to be directed towards achieving better understanding of the heuristic techniques that can be employed to improve fitness. Examples of areas where an artist might benefit from being informed include the properties of the medium (e.g., the behavior of stone, the underlying chemistry behind different types of cooking techniques), the characteristics of different subjects and tools (e.g., anatomy, the sounds that can be produced by different musical instruments and their associated range), approaches that can be used to create certain effects (e.g., the use of perspective in drawing, narrative devices in literature), and stylistic conventions that should nearly always be adhered to (e.g., proper spelling) along with those that should be occasionally violated for the sake of achieving realism or impact (e.g., rules of grammar).

Nowhere is the distinction between science and art more apparent than with respect to what is meant by the term “experiment”. In science, the experiment is nearly always constructed in order to test the fitness of a particular theory by allowing observations and predictions to be compared. In the arts, the “experiment” involves employing an unfamiliar technique or subject to produce a work that is often radically at odds with existing works and, perhaps, existing notions of fitness as well.

Another implication of the highly rugged nature of the fitness landscape for different types of art relates to how art is studied. In the sciences, particularly where decomposability is present, phenomena can be examined in isolation and small, carefully constructed demonstrations can often be just as informative as real-world observations. When Galileo demonstrated the principles of uniform acceleration of objects by dropping two balls of different size off the Leaning Tower of Pisa then showing that they landed at the same time, the simplicity of the experiment was no obstacle to the proof of the concept—indeed, its simplicity and replicability made it all the more convincing. In the arts, however, concepts are far more likely to be illustrated in the context of a complete or major section of a work created by a master of the craft. The utility of techniques to enhance fitness is generally best demonstrated by showing how they can contribute to a complete work of very high fitness.

Table 2: Differences between art and “hard” science fitness landscapes

Characteristic	“Hard” Science	Art
Number of attributes	Small	Very Large
Decomposability	High	Low
Replicability of observations	High	Low
Variability of fitness over time	Low	High
Focus of experimentation	Confirm or refute existing theory	Exploration of new techniques or heuristics
Objectives of research	Understanding and predicting behavior	Identifying heuristics for assessing or improving fitness

Key differences between “hard” sciences and the arts are summarized in Table 2. In interpreting the table, we must recognize that the two columns represent the extreme end points of a continuum. While the science column might represent a plausible characterization of some of the physical sciences, as we move into the life sciences and then into the social sciences, the number of attributes grows, replicability declines, and fitness may well vary over time. Similarly, while the art column may be representative of some arts, such as painting, many artistic endeavors, such as architecture, may require equal parts engineering science and creative vision. Understanding where a particular area of inquiry, such as the informing sciences, fits on this continuum is likely to provide important insights into the types of research that are, and are not, likely to produce valuable results. We therefore now turn to the question of how the ruggedness of a particular fitness landscape may be predicted.

Complexity and the Rugged Fitness Landscape

Synthesizing the discussions of the previous section, we might expect to see environments better described by a rugged fitness landscape than a decomposable fitness function in situations where the following characteristics are present:

1. *Numerous attributes appear to impact fitness.* Where fitness appears to be determined by a small number of attributes, the opportunities for the development of multiple peaks is limited.
2. *Attributes appear to have strong interrelationships.* Where attributes do not appear to exert the same influence on fitness in different regions of the fitness landscape, the opportunities for alternative solutions that are local maxima grow.
3. *Entities on the landscape continually adapt to improve fitness and linkages to dynamic external landscapes are present.* As noted earlier, these types of dynamics tend to favor environments where entities gravitate towards multiple peaks so as to reduce the risk of major declines in peak fitness.

Interestingly, these three characteristics closely parallel the most widely used task complexity definition in the management literature (Gill & Hicks, 2006), being essentially identical to the set of characteristics that Wood (1986) proposes as the source of objective task complexity. This convergence is more than coincidental; achieving a resolution to a complex task can, itself, be visualized as finding an appropriate point on a fitness landscape—where fitness is represented by how effectively the solution meets the requirements that the task performer has been given. Thus, task performance can be viewed as a special case of the more general problem of exploring a fitness landscape. The relationship between the first two characteristics and complexity is also consistent with the writings of Herbert Simon (1981), who characterizes decomposability and complexity as polar opposites.

The Informing System Fitness Landscape

The preceding examination of landscape shapes led to two important conclusions: 1) that certain entity characteristics (i.e., many attributes, many interdependencies between attributes, dynamically changing fitness) tend to promote rugged fitness landscapes, and 2) that the nature of research in highly rugged landscapes (e.g., the arts) will necessarily differ from that suitable for decomposable landscapes (e.g., the physical sciences). In this section, we ask the question: how rugged is the fitness landscape for an informing system likely to be? The subsequent section then considers the implications for informing science research.

We begin with two assumptions. The first is that there are four key components to a basic informing system: the sender, the client, the delivery system, and the task to be performed (Cohen, 1999). It has been noted elsewhere (e.g., Gill & Bhattacharjee, 2007) that numerous variations to such systems exist, some of which appear to be substantially more complex (e.g., multiple senders, multiple clients, and/or multiple channels). It stands to reason, however, that if fitness for the basic informing system is rugged, then more complex variations are likely to be rugged as well.

The second assumption is that informing system fitness landscapes follow the same principles identified for generic NK landscapes (e.g., Kauffman, 1993) and complexity in general (e.g., Wood, 1986). In other words, where the fitness function has many attributes, a high degree of interaction between attributes, and is subject to changes over time, we may expect that the resulting landscape to be quite rugged.

Number of Attributes

Suppose we were to attempt to define a general fitness function for informing systems of the form:

$$F = f(x_1, x_2, \dots, x_N)$$

An obvious question to ask would be: What is the value of N ? Since this question could probably not be answered authoritatively without knowing the precise nature of the function—at which point the question of the fitness of informing systems would have been “solved”—the best we can hope for is to establish an approximate lower bound. It seems very unlikely that the number of attributes impacting informing system fitness could be less than a few hundred although, realistically, it is probably much, much larger. This lower bound estimate was derived by considering the types of factors that could exert an impact on informing effectiveness for each entity in the informing system.

For the client entity, there are many areas that seem highly likely to impact the effectiveness of an informing system. These areas, in turn, each involve multiple attributes. To give some examples:

- *Motivation.* The ability to inform a client effectively will almost certainly depend on his or her desire to be informed. Such desire, in turn, would involve the satisfaction of some underlying drive or need on the part of the client. Research in the area of motivation has produced estimates ranging from 4 (e.g., Lawrence & Nohria, 2002) to 16 (e.g., Reiss, 2000) of these underlying drives or desires. Examples include the drive to acquire, to bond, to learn, and to defend (Lawrence & Nohria, 2002), as well as desires that include seeking power, independence, curiosity, acceptance, and order (Reiss, 2000), to name just some. Individuals are found to have unique individual desire profiles, specifying the degree to which each drive/desire is important to that particular person. We would therefore expect that each motivational factor would represent a separate argument to the fitness function.
- *Prior knowledge.* What a client already knows would necessarily impact the fitness of the informing system. With insufficient prior knowledge, the data that is being transmitted is un-

likely to be assimilated as information. With excessive prior knowledge, the informing system is unlikely to produce much useful impact. Unlike motivation, there do not appear to be any ready-made frameworks for characterizing prior knowledge. We would certainly expect a large number of attributes would be required, such as scalar values reflecting knowledge of relevant jargon, knowledge of system operation, practical task experience, and knowledge of task concepts.

- *Cognitive preferences.* A growing literature suggests that different individuals have different preferences with respect to how they learn. Since many informing systems are likely to include client learning among their objectives, the preferences exhibited by an individual client could certainly impact the overall fitness of a particular informing system. A recent survey of learning style research identified five different frameworks that collectively would require nearly 30 distinct attributes (Hawk & Shah, 2007, p. 12-13).

Arbitrarily guessing that it would take at least as many attributes to adequately characterize the client's prior knowledge as it would to characterize each of the other two areas, it is hard to imagine that the client entity could be characterized with fewer than 60 attributes. One might reasonably expect the value to be vastly larger.

A similarly large number of task attributes could potentially impact the informing fitness. Some examples include:

- *Complexity.* Presumably, the complexity of the task component of the system would impact the fitness of a particular informing system. A recent review found at least 13 distinct task complexity constructs that had at least some support in the literature (Gill & Hicks, 2006). Some of these were further broken down into a larger number of individual attributes, suggesting a total number of distinct attributes that is probably closer to 30.
- *Job design and enrichment.* The job design literature examines how the design of a task impacts employee motivation and performance. At least 5 distinct attributes (variety, identity, significance, autonomy, and feedback) are identified (Hackman & Oldham, 1976). The ability of information technology to impact these attributes in a manner that subsequently impacts system fitness has already been reported in the literature (e.g., Gill, 1996).

With respect to the delivery system, a vast number of characteristics could be required in order to capture performance (e.g., bandwidth, speed), reliability (e.g., availability, vulnerability), information formatting (e.g., text, audio, video), and interactivity (e.g., directionality of information flow, interface design). Many of these would, in turn, require many sub-attributes in order to achieve a full characterization of the system.

On the sender side, we would anticipate that a collection of attributes characterizing sender motivation to inform could potentially impact system fitness. We would also expect that a sender's *awareness of* client motivation, prior knowledge, and cognitive preferences could positively impact effectiveness in many informing situations. Thus, characterizing the sender is likely to require at least as many attributes as characterizing the client.

Beyond these component-specific attributes, we could also expect a number of general attributes might apply to the system as a whole. For example, the locus of control for informing might reside with the sender, the client, or even the delivery system (e.g., a cable TV network decides what programming to carry without being the originator of the content or the client).

What the preceding informal exercise demonstrates is that the number of attributes that could plausibly impact the informing system fitness landscape is quite large. Indeed, the number is sufficiently large—and the associated constructs sufficiently fuzzy and in need of better definition—that decades of research would be required to achieve an acceptable understanding of the inform-

ing system landscape even if it is fully decomposable. We therefore now turn to the next criteria for ruggedness: interdependency of attributes.

Interdependency of Attributes

Much like the number of attributes, precisely determining the level of interdependency between attributes in an informing system landscape requires complete knowledge of the fitness function. Indirect evidence that such interdependency could be present and an important contributor to fitness can arise from a number of sources, however. Among these:

- *Theoretical arguments.* Where theory suggests that proposed contributors to fitness should interact, then it would make sense to hypothesize such interactions as part of empirical research.
- *Empirical examples.* In depth analyses of different informing situations may lead to evidence of particular attributes contributing to fitness in different ways. For example, a variable that contributes positively in one context may contribute negatively in another or a variable that exerts strong influence upon fitness in a particular situation may have no influence in another situation.
- *Observed ruggedness.* Where the entities on a fitness space are adaptable, we would expect to see evidence of very different approaches to achieving fitness emerging, as different entities migrate towards local peaks. In a completely decomposable landscape, on the other hand, we would expect to see entities all striving towards the same peak.

We will now consider how these might apply to informing systems.

One of the strongest theoretical arguments that can be made for interdependency of attributes is the presence of variables that need to be consistent with other variables to achieve the desired effect. For example, the use of baking soda to make a batter rise requires the presence of an acidic ingredient, such as lemon juice. Many of the attributes presented in the previous section would be hypothesized to require such consistency. For example, the cognitive preferences of the client (e.g., which include attributes such as preferences for visual vs. verbal information; Hawk & Shah, 2007) should, according to theory, be matched with the information format provided by the delivery system. Indeed, nearly all of the learning style preferences would, in theory, relate to some corresponding delivery system or sender characteristic or set of characteristics. Similarly, we would expect the relative strength of different client motivational characteristics to interact with the job design attributes of the task. For example, if the client has a strong drive for control (e.g., Gilbert, 2007), we might expect that the high interactivity designed into the system would impact fitness far more positively than it would for individuals with a lesser drive for control. In the area of client prior knowledge, we could predict, with a high level of confidence, that a severe mismatch between the sender's perception of client knowledge and actual client knowledge would dramatically hamper the effectiveness of informing. Indeed, it is difficult to think of any attribute governing the fitness of an informing system that could not plausibly interact with other attributes in determining system effectiveness. Thus, on a theoretical basis, we would expect the informing system landscape to be quite rugged.

Providing a full catalog of empirical examples of reported interactions between informing system characteristics would be far beyond the scope of the current paper, which is principally focused on considering how research in a rugged fitness landscape needs to be directed. Instead, we will need to content ourselves with a two case studies that illustrate what such examples might look like.

Both illustrations involve a unique undergraduate programming course (see Gill & Holton, 2006). The course was extremely unusual in its design, being entirely self-paced, making extensive use

of multimedia and incorporating a strong peer-based learning component. In analyzing the course outcomes, two significant departures from widely observed fitness outcomes were observed. Specifically, in the research literature describing student performance in introductory programming courses two factors consistently impact learning: prior programming experience (e.g., Hagan & Markam, 2000; Holden & Weeden, 2003; Wilson & Shrock, 2001) and gender (e.g., Goold & Rimmer, 2000; Sackrowitz & Parelius, 1996). Despite having extensive data on 254 students previously enrolled in the course, absolutely no evidence of such effects could be detected for the self-paced course. Thus, the researchers concluded that nature of the course structure somehow interacted with the experience and gender characteristics so as to render them less relevant than they were for more conventionally designed courses (Gill & Holton, 2006).

A second example involves a direct comparison of the self-paced programming course with two other courses: another version of the same course programming taught by another instructor and a different course taught by the original instructor, all offered during the fall semester of 2007 (Gill & Jones, 2008). The characteristics of the three courses are presented in Table 3.

Table 3: Cross-Course Comparison (from Gill and Jones, 2008)

	Ism3232.A	Ism3232.B	Ism6155.A
Classroom Lectures	No	Yes	Minimal
Multimedia Lectures	Yes	No	No
Moderated Classroom Discussions	Optional	No	Yes
Paired Student Problem-solving	No	Yes	No
Student Presentations	No	No	Yes
Deadline Flexibility	Yes	No	No
Mandatory Attendance	No	Yes	Yes
Examinations	No	Yes	No
Outside Class Projects	Yes	No	Yes
Level of Performance Feedback	High	High	Low
Grade Subjectivity	Low	Low	High
Student Level	Undergraduate	Undergraduate	Graduate
Source	Evolved	Designed	Designed
Instructor	Instructor A	Instructor B	Instructor A
Instructor Experience with Course Subject Matter	High	Low	High
Evaluations	Outstanding	Outstanding	Outstanding

Particularly noteworthy in Table 3 are two things. First, all three of the courses were perceived to be outstanding in their effectiveness. The two courses taught by the original instructor, Ism3232.A and Ism6155.A, had both won the Decision Science Institute’s Innovative Curriculum competition (in 2007 and 2005 respectively). The third example, Ism3232.B, received the second highest student evaluations in the history of the course (for which roughly 70 sections had been taught since the course’s inception) the very first time that particular instructor taught it. Second, of all the course attributes in the table, there is not a single example of consistency across all three courses. In a decomposable model, this would provide support for one of two hypotheses: that none of the attributes are important or that they balance each other out—meaning that an “optimal” course could be constructed by choosing the better value for each attribute. A far more compelling explanation—supported by the history of Ism3232.A, which had experienced periods of much lower fitness in prior semesters (Gill & Holton, 2006)—is that each course is at, or close to, a distinct fitness peak arising from the interaction of its particular design characteristics.

The final evidence of ruggedness can be acquired from examining how entities are positioned on the fitness landscape. Where entities are adaptable, we would expect them to migrate towards fitness. Where they are not, we would expect less fit entities to disappear from the landscape over time. Where a landscape is fully decomposable, this process should lead to entities gravitating

towards sameness, since decomposable landscapes typically have a single fitness peak. Where the landscape is rugged, we would expect to see a far greater diversity of positioning.

Once again, this paper makes no attempt to survey all informing system fitness landscapes. Instead, we return to our university example. It has been noted that universities are, in fact, part of the academic informing system (Gill & Bhattacharjee, 2007). If the fitness landscape for universities were decomposable—which would be approximately true if magazine-created rankings were accurate measures of university fitness—we would expect them to be very similar and would also expect that similarity to be increasing over time as institutions migrate towards greater fitness. If the landscape is rugged—as it would be for the hypothetical best choice fitness function as well as for other alternative functions that could readily be imagined—we would expect to see a broad distribution of missions and structures among institutions present in the landscape. Just a casual glance at the landscape of universities suggests that the second characterization is far more apt than the first. Furthermore, one can point to the growth of non-traditional institutions such as the University of Phoenix (the largest private university in the U.S. according to their web site) as evidence that diversity in the higher education space is actually growing, rather than shrinking. In fact, one would probably do better to view decomposable fitness rankings, such as those produced by *US News and World Reports*, as potential contributors to fitness—boosting student applications to those institutions exhibiting ranking improvements—rather than as measures of fitness.

Dynamics of Fitness Landscape

The final factor that is expected to contribute to ruggedness is the degree to which the fitness landscape is continuously changing. The substance of the argument is that such a landscape encourages entities to migrate to diverse fitness peaks for the sake of adaptability. The fitness benefits of changing peaks so as to achieve “optimum” fitness are also reduced, since what is optimum can be expected to change continuously.

Two important forces support the view that the fitness landscape of the typical informing system is likely to be quite dynamic. Information technology (IT) is often central to the delivery system component of an informing system. The reader should not need convincing that the pace of IT change over the past five decades has been extraordinary. Just a decade ago, transmitting video over the Internet was largely impractical except in university labs and corporate boardrooms, today it is routine. Interactive social environments, such as Second Life, providing an opportunity for rich interactions between senders and clients, are still in their relative infancy. Enabling technologies, such as XML and web services, making linkages between senders and clients much easier to establish and increasing their flexibility, have only recently begun to achieve widespread acceptance. In short, in the past and for the foreseeable future, we can expect to see technology play an important transformative role in reshaping the informing system fitness landscape.

Globalization is the second important force that can potentially transform the informing system fitness landscape. Its impact is felt particularly by the task, sender, and client entities. Offshoring and outsourcing frequently mean that systems that once needed to manage product or service production tasks must be modified towards coordinating external organizations that have taken on the production role. Perhaps even more significantly, there are many personality and motivational characteristics that appear to vary considerably across cultures (Hofstede, 2001). As informing systems expand their informing activities across borders, these factors cannot help but influence the shape of overall fitness through changing the distribution of client and sender characteristics.

Thus, we find that the typical informing system is likely to exist on a fitness landscape that is both subject to continuous change and a function of multiple attributes of high interdependence; in other words, the informing system landscape meets all the prerequisites of ruggedness.

Researching the Rugged Landscape

The arguments of the preceding section were intended to convince the reader of the distinct possibility that the informing system fitness landscape is quite rugged in shape. In this section, we address two questions relating to research in the rugged fitness landscape: (1) What happens if you make the assumption of decomposability while researching a landscape that is actually rugged? and (2) What types of research are most suitable for researching a landscape that you believe to be rugged?

The Assumption of Decomposability

From the earlier discussion of the continuum of science and art fitness spaces, it can be concluded that a decomposable fitness space offers numerous rewards to the researcher. Among these are:

- The likelihood that research will yield “attractive” theory, which is to say theory that is very compact in size and large in its domain of applicability.
- The ability to devise experiments that examine one variable at a time, while retaining both rigor and generalizability.
- The opportunity to take many observations, each consisting of multiple independent variables, and separate out the influence of each using statistical techniques that also provide confidence estimates for each influence, most notably those techniques derived from multiple regression.

The last of these is particularly attractive in a field research setting because such techniques can be used either to test existing theory or to discover relationships using data that can be acquired cheaply from existing sources (e.g., financial databases) or through instruments such as surveys.

As a quick review, techniques such as multiple linear regression analysis make the assumption that the fitness value—referred to as the dependent variable (F)—is generated from N characteristics through a decomposable linear process of the form:

$$F = c_0 + a_1x_1 + a_2x_2 + \dots + a_Nx_N + \varepsilon$$

The ε value represents error that cannot be explained even when the values of c_0 and a_1 through a_N are correctly established. Multiple regression serves to perform a mathematical computation that estimates c_0 and a_1 through a_N so as to minimize the residual error when the formula is fit to a set of actual observations.

The regression equation presumes that a_1 through a_N contribute to fitness decomposably. Where specific interactions between variables are believed to be likely, corresponding interaction terms may be introduced into the estimating equation. These are normally added only as a consequence of strong prior belief about the process. Arbitrarily introducing such terms for every possible interaction can rapidly lead to an estimating equation with so many unknowns to be computed that it exceeds the estimating power (i.e., degrees of freedom) of the observations available.

The power of multiple linear regression and related techniques makes them a staple of research in the social sciences, where developing a mathematical derivation of theory is much harder than in the physical sciences. Even where mathematically-based theory can be proposed—as is often the case in economics—the regression technique is often used to estimate actual values for the parameters contained in a particular theory. Thus, it is one of a family of statistical tools taught to nearly every social science doctoral student and its use is commonplace throughout the social science research literature.

Naturally, if a researcher strongly believes a fitness landscape to be rugged, he or she would be ill-advised to employ multiple regression analysis to determine the best linear decomposable fit to

the observations. What point is there in achieving a mathematical best fit with an equation of the wrong form? Suppose, for example, you had a fitness function that mapped the list of ingredients to an objectively determined measure of “taste fitness” for all the recipes in a cookbook. If you were to do a regression on taste (dependent variable) using the ingredients (independent variables), you might find—for instance—that garlic shows a high positive significance. What would that tell you (other than, possibly, that the individuals rating the recipes enjoyed garlic)? What it would *definitely not tell you* is that you could improve your recipe for angel cake by adding garlic to it. Indeed, the whole notion of applying a technique that assumes linear decomposability to a fitness landscape that is so obviously not decomposable is preposterous.

But what about the fitness space that might be decomposable (you’re not sure) or which you believe to be partially decomposable (you think, or theory tells you, that some variables will contribute to fitness uniformly across the entire fitness landscape)? In this case, the notion of employing regression-related multivariate techniques in an exploratory fashion does not seem so far-fetched.

From a researcher’s point of view, what would be ideal is if we could apply low-cost techniques such as multiple linear regression to observations drawn from a particular fitness landscape and, based upon the results, make an informed judgment as to whether or not the assumption of landscape decomposability was justified. For example, we might identify attributes that we expected—either intuitively or based upon theory—to impact fitness and gather observations of these attributes with known or estimated fitness values. We could then regress the attributes (independent variables) against the associated fitness value (dependent variable). If the analysis produced statistically significant regression coefficients, we might then conclude that the decomposability assumption was valid. If not, we would be forced to conclude that the landscape was too rugged for these techniques to be useful—as was obviously the case for our recipe example.

Unfortunately, experimental evidence suggests that the results of applying techniques such as multiple regression analysis may be very misleading when it comes to judging the shape of a fitness landscape. In a companion paper to the current paper (Gill & Sincich, 2008), a series of experiments was devised in which NK fitness landscapes of known shapes were created and then analyzed using multiple linear regression. Three different landscape shapes were considered:

- *Partitioned*: The underlying function used to create fitness values was in two parts: a linear part involving a subset of the attributes (i.e., $a_1x_1 + a_2x_2 + \dots + a_Dx_D$) and a non-decomposable function using the remaining attributes (i.e., $f(x_{D+1}, x_{D+2}, \dots, x_N)$, where N is the total number of attributes). In this model, variables either contributed decomposably to fitness (variables 1 through D) or were completely interrelated with other non-decomposable variables (D+1 through N)
- *Chaotic*: Fitness was determined by a non-decomposable function of all the relevant attributes (i.e., $f(x_1, x_2, \dots, x_N)$, where N is the total number of attributes).
- *Mixed*: Fitness was determined in two parts: by a linear part involving a subset of the attributes (i.e., $a_1x_1 + a_2x_2 + \dots + a_Dx_D$) and by a non-decomposable function of all the relevant attributes (i.e., $f(x_1, x_2, \dots, x_N)$, where N is the total number of attributes). This differed from the partitioned model in that certain variables could contribute to fitness both decomposably (1 through D) and through their interaction with other variables.

The paper examined the results of regressions of all attributes against the generated fitness value (no error term was used) in two conditions: 1) using original fitness values, and 2) allowing entities to migrate towards local peaks in the fitness space (as Kauffman assumed they would do). A summary of the results of these experiments is presented in Table 4.

Table 4: Summary of experimental regression results for different fitness landscape shapes (Gill & Sincich, 2008)

Landscape	Description	Initial	During Migration To Higher Local Fitness
Partitioned	Independent variables <i>either</i> contributed to fitness decomposably <i>or</i> through non-decomposable relationships with other variables.	A large number of spurious statistically significant relationships among the interrelated variables were detected. Decomposable coefficients reproduced with perfect accuracy.	Spurious significances grew. Decomposable coefficient estimates remained accurate.
Chaotic	All relationships were non-decomposable.	No spurious coefficient estimates were detected beyond those likely to occur by chance.	Spurious coefficient estimates emerge, often highly significant.
Mixed	Some variables contribute to fitness decomposably. <i>All</i> variables participate in non-decomposable contribution to fitness.	Decomposable coefficient estimates of reasonable, but not perfect, accuracy are obtained with high significance. No spurious coefficient estimates for non-decomposable variables were detected beyond those likely to occur by chance.	Spurious coefficient estimates emerge for non-decomposable variables, often highly significant. The quality of decomposable variable estimates declines.

These results suggest that landscape ruggedness could present a serious barrier to the validity of research conducted under the assumption that the underlying landscapes are decomposable. The problem is particularly severe where the underlying fitness function is partitioned such that variables exclusively contribute decomposably or through interaction. In such cases, erroneous significances on interacting variables appear side by side with accurate estimates of decomposable variables in all cases. Without knowing the underlying process, it would be easy to interpret the entire underlying process as decomposable and the spurious significances as valid. The companion paper (Gill & Sincich, 2008) explains the mathematical source of these errors and why the results cannot be treated as meaningful. The good news is that, in the absence of migration towards higher fitness, the chaotic and hybrid landscapes do not produce similar errors in estimating significance.

As soon as entities are allowed to migrate towards higher fitness, all three landscapes exhibit major errors in coefficient estimates and significances. As explained in the companion paper (Gill & Sincich, 2008), this is largely a consequence of such migration's impact on the underlying assumption of independent observations. If, however, we accept the argument that entities on a fitness landscape will tend to migrate towards higher fitness (or that less fit entities will not survive), these findings suggest that any results derived from the application of techniques such as multiple regression to observations drawn from rugged landscapes must be viewed with great skepticism.

The severity of the threat to research validity posed by these findings depends, to a great extent, on the origins of the theory. In some fields, particularly economics and finance, we have already noted that a great deal of the applicable theory has mathematical origins. Where statistical analysis of the sort described above is used to test such theory, we may presume that the theory assumes decomposability (since it would not make sense to apply such analysis to a landscape presumed to be rugged). This means that if the underlying landscape is actually rugged—thereby causing the results to fail to confirm precisely to the theory—the theory will tend to be rejected. Furthermore, since the very presence of ruggedness implies the theory is wrong (since the theory assumed decomposability), it seems that no particular harm is done. It might also be commented that in these disciplines, where empirical results fail to confirm mathematical predictions—as is

notably the case for models of individual utility (Gill, 2008)—the mathematical theory tends to exhibit greater survivability.

A much graver threat to validity occurs in situations where theory tends to emerge as a consequence of observations, rather than from formal derivation. This approach to theory-building—which would be typical in social science disciplines such as MIS, management, and, presumably, the informing sciences—views theory development as an iterative process in which observations are used as a basis of establishing the same theory that is later tested with empirical research. The problem arising from this process is that analytical results obtained under the mistaken assumption of decomposability will become incorporated into a body of theory. That theory will then subsequently be tested using a different set of observations gathered from the same landscape under the same mistaken assumption of decomposability; as a consequence, such tests will most likely confirm the mistaken theory since ruggedness is not the same as randomness—two independent samples drawn from the same landscape will tend to support the same conclusions.

To illustrate the problem by revisiting a previous example, suppose a researcher performing a regression on ingredient variables in recipes sampled from a cookbook discovers that the garlic ingredient variable makes a particularly significant contribution to taste fitness. Based on that exploratory research, the same researcher publishes the Garlic Acceptance Model (GAM), which proposes a causal relationship between garlic and taste fitness, hypothesizing that recipe fitness can be enhanced by adding garlic—a model totally consistent with the assumption of landscape decomposability and made even more credible should it happen to be true that the investigator loves garlic. Presuming the original sampling was done correctly, a subsequent researcher drawing another large sample of observations from the same recipe landscape in order to test the GAM will confirm the original finding, thereby adding further credence to the GAM. As this process continues, the GAM will become widely accepted. In effect, what has happened is that heuristic rules acquired through observation have attained the status of theory.

The statistical problem posed by ruggedness becomes particularly acute when the underlying fitness space is partitioned or mixed in its structure. In this case, statistically significant results are likely to be obtained for fully or partially decomposable attributes that seem quite plausible. These attributes can be characterized as the *low hanging fruit* of such research since their decomposability makes them easy to find in almost any setting. Further, statistical confirmation of their presence would tend to add credibility (at least in our own minds) to our initial results and our underlying assumption of decomposability. For example, suppose we were studying examples information system adoption and the underlying fitness landscape was rugged-mixed, with the characteristic “usefulness” contributing to fitness decomposably. If a statistical test of our observations finds that a usefulness construct is predictive of the degree to which a system is used, consistent with our expectations, we might naturally become more confident that the other significant variables we identified were equally decomposable. While our conclusions might be valid, those observed significances might also be an illusion.

As was previously stated, there is little danger of the GAM becoming a reality because the fitness landscape for recipes is so obviously and intuitively rugged *and* because it would be quite easy to devise laboratory experiments to refute the GAM by adding garlic to other recipes and observing consequent fitness. But the same cannot be said about many of the landscapes we research in the social sciences, where reproducing real world phenomena in a laboratory setting is often impractical (e.g., the likelihood of acceptance of a large information system within an organization) and both our independent attributes and our fitness measures are commonly tacit rather than being directly observable. Furthermore, to emphasize what was stated earlier, decomposable processes lead to *attractive theory*: compact, generalizable, replicable. Highly rugged landscapes, on the other hand, lead to *ugly theory*: large in size, filled with caveats, and breathtakingly hard to reproduce. Attractive theory leads to influential research publications; ugly theory—however valid

it proves to be—looks as if you are making it up as you go along. The ambitious researcher would therefore be well advised to hope that the landscape that he or she is studying is largely decomposable. But what is to be done when he or she becomes convinced that the underlying landscape being studied is actually quite rugged? We now turn to that question.

Empirical Research on the Rugged Fitness Landscape

The science-art continuum presented earlier in the paper illustrates how, at the extremes, the nature of the fitness landscape being investigated can exert a sizeable impact on the objectives and conduct of research. One of the first questions a researcher therefore needs to answer about a research domain is: How rugged is it? Where the domain is extremely rugged, as it is for the arts, it is likely to be pointless to attempt to create any general theory of the domain. Better instead to focus research efforts on identifying heuristic techniques for improving fitness—as is done in the arts.

Where ruggedness is confined to a small number of identifiable peaks, on the other hand, it may be possible to develop attractive theory that is applicable to each peak. Generally speaking, any domain that has many fitness-relevant attributes and few peaks is likely to exhibit a high level of decomposability. If the researcher ensures observations being aggregated are not drawn from alternative peaks—or, to be more precise, from the fitness wells that lead migrating entities towards alternative peaks—then many of the misgivings that have been raised regarding misleading statistical results will be of less concern.

Given these considerations, an overarching goal of research on a rugged fitness landscape needs to be establishing a broad picture of the territory. As suggested by the earlier attempt to characterize the informing system landscape, this will not necessarily prove easy to do. Among the associated challenges are included:

- Identifying the set of plausible attributes impacting fitness
- Identifying the level of interrelatedness between the attributes
- Identifying possible peaks (or at least high points) in the fitness landscape
- Identifying regions of the fitness space where important variables may be treated decomposably (which may or may not correspond to peaks).

With such a map in place, the researcher can begin to answer questions regarding the generalizability of findings from a given region, the likelihood that attractive theory will emerge from broad investigations, and, perhaps, the degree to which entities have already attained fitness peaks (or perceive themselves to have attained them). This last issue is particularly important from a motivational perspective. If intelligent entities on the fitness landscape—be they individuals, groups, or organizations—perceive that they are already operating at high fitness levels, the likelihood that they will be interested in research findings related to their specific fitness function will be very limited. As a consequence, the researcher's only audience is likely to be other researchers.

The difference in approaches between decomposable and rugged landscape research includes placing fundamentally different priorities on different research activities. One taxonomy for classifying such activities—with a particular focus on their theoretical contribution—uses the dimensions of “building new theory” and “testing existing theory” (Colquitt & Zapata-Phelan, 2007). That taxonomy is illustrated in Figure 3.

The taxonomy identifies five basic types of contribution: *builders* develop new constructs and explore relationships, *expanders* take existing theory and build upon it, *testers* verify existing theory, *qualifiers* explore the limitations of existing theory and modify the theory as necessary,

and *reporters* gather observations. It is interesting to consider how the nature of the underlying fitness landscape could influence the relative importance of these contributions.

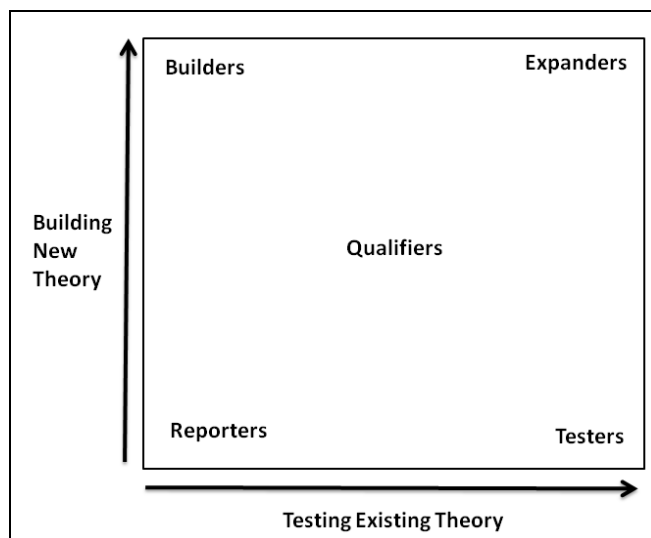


Figure 3: Theoretical contributions of empirical research
(from Colquitt & Zapata-Phelan, 2007, p. 1283)

As has been previously noted, one of the most significant differences between decomposable and rugged landscapes is in regard to the nature of the theory that is likely to emerge. We would expect the perceived value of researcher roles to be influenced by a discipline's assessment of its own landscape. In decomposable landscapes, we would expect roles influential in the development and refinement of theory—e.g., builders, expanders, and testers (Colquitt & Zapata-Phelan, 2007, p. 1283)—to be particularly emphasized, since the completed theory is likely to be attractive. In the rugged landscape, we'd expect priorities to be somewhat reversed. Since such a landscape is likely to produce ugly theory anyway, those roles that seek out and describe interesting areas of the landscape—most particularly, the reporter—should be of greater significance. For example, consider the importance of the role of the reporter, also known as the critic, in the arts.

The other justification for a strong reporter role in rugged landscape research can be made in terms of our appetite for observations. As fitness landscapes become increasingly rugged, observations from one area of the landscape tell us increasingly less about what is going in other areas of the landscape. For example in an NK space of with 12 binary attributes, there are 4096 combinations leading to fitness (2^{12}). If the space is fully decomposable (12,0), using multiple regression we will likely have sufficient degrees of freedom to establish a complete and accurate map of the landscape with under 100 observations. If it is chaotic (12,11) on the other hand, we need to acquire observations of all 4096 combinations to develop a complete map—since each observation tells us nothing about similar (but not identical) combinations of attributes. In addition, if the landscape is dynamically changing, it will not be sufficient to survey the landscape once. It must be surveyed continuously if an accurate map of fitness is to be maintained.

As a consequence of the influence of ruggedness on our need for observations, we would expect that the reporter role to remain critical in such research for as long as such research is being conducted. Random reporting, however, will not be sufficient to provide useful insights into such spaces. It is quite plausible to imagine that our landscapes could be huge. Consider how our previous informal calculation suggested that defining the informing system fitness function would require a few hundred attributes at a minimum; the actual number could just as easily run into the thousands. With a domain of that magnitude, even carefully designed random sampling is unlike-

ly to produce important insights into overall fitness. Thus, our reporter needs to be very efficient in selecting and conducting observations. We now consider that topic in greater detail.

Reporting the Rugged Landscape

One of the most important roles a reporter can undertake in a rugged fitness landscape is to identify peaks. As previously noted, the challenge that ruggedness often presents is a number of potential observations that is too large to contemplate. On the other hand, outside of the arts, we are likely to be looking at landscapes with interrelationship levels (e.g., K in the NK model) that are far less than chaotic, meaning that the number of expected fitness peaks would be well under the theoretical limit (e.g., $2^N/N+1$ for NK landscapes). Indeed, their number may be sufficiently manageable to allow all major peaks to be catalogued. Furthermore, the activity of identifying achievable fitness peaks can be very useful in situations where entity characteristics can be altered to improve fitness, as would certainly tend to be the case for the task and delivery system components of an informing system.

The obvious challenge in making and reporting observations from a rugged landscape is telling whether or not a particular entity is at or close to a fitness peak. Sometimes the entities being studied can provide useful insights. Where the entities on the landscapes being modeled are individuals, groups, or organizations, we would anticipate conscious migration towards higher fitness should frequently occur. As a result, the reporter may well be able to assess proximity to a local peak by examining the nature of the search process through which the entity reached its current fitness state. This means that a longitudinal investigation of entity fitness—acquired through examination of archival data, interviews, and other sources—may be extremely valuable. Stated another way, the study of history necessarily plays a critical role in rugged landscape research—another close parallel to the arts. It is possible to test the fitness of the predictions from a mathematically derived theory in the hard sciences without knowing how that theory evolved; the same can only rarely be said for entities on a rugged landscape.

In order to make such observations effectively, the reporter would benefit substantially from practical expertise in the landscape domain being investigated. Not only are the attributes that must be observed likely to be numerous, many will prove difficult to observe directly, such as those relating to client and sender motivation in the case of an informing system. The reporter would also need to have considerable expertise in interpreting the opinions of individuals participating in the entity being observed. The subjective opinions of individuals on matters such as causality tend to be heavily discounted where landscapes are presumed to be decomposable. Such discounting makes sense under those circumstances since decomposability implies that relationships will generalize from one situation to another. Thus, the investigator studying entities across a domain is likely to have a far more objective perspective than the individual participant. Where ruggedness is present, on the other hand, participant opinions relating to the underlying factors relevant to a particular situation are likely to be at least as valid as those of the investigator attempting to generalize from entities observed in other regions of the fitness landscape. Thus, participant opinions—however subjective—need to be elicited and carefully considered if the local fitness landscape for a particular entity is to be understood.

Another aspect of reporting that differs between decomposable to rugged landscapes is the nature of what the research is trying to accomplish. Previously, we observed that on the continuum from science to art, the emphasis of research shifts from understanding the fitness function (science) to exploring techniques whereby fitness can be improved (art). That shift, however, becomes evident long before we reach what would normally be considered arts. Looking at fields such as evolutionary biology and genetics, we see at least as much interest in investigating how migration towards fitness takes place as we do in understanding the fitness function itself. Indeed, the process of migration towards fitness is central to much of the landmark research in these domains,

from Darwin's original work to Kauffman's (1993) NK landscapes and Holland's (1992) genetic algorithms. Thus, the reporter in the rugged landscape needs to be attuned both to identifying those characteristics that lead to fitness in a particular region of the landscape and to identifying those techniques that can be used to search for states of higher fitness and to transition to those states.

Other Researcher Roles in the Rugged Landscape

The preceding discussion of the importance of the reporter role might seem to imply that theory plays minimal role in a rugged landscape and that theory-intensive roles such as the builder, qualifier, and expander are therefore of little importance. That conclusion would be far from the truth. While it is true that informed observations make critical contributions to rugged fitness research, synthesis of these observations is necessary if better understanding of such landscapes is to be achieved.

Consider, for example, medicine. Although classified as a science by almost any definition, its principal domain of study—the human body—clearly exhibits the prerequisites of ruggedness: many entities (e.g., systems in the human body, each made up of huge numbers of cells), high level of interaction between these entities, and dynamically changing fitness (e.g., through the action of aging and the environment). Based upon this, we would expect to find no compact and generalizable “theory” of the human body—nor does such a theory exist. Nonetheless, great progress has been made in the field through the development of theory relating to individual body subsystems, through the direct and indirect (e.g., survey) observation of individual patients and experimental subjects, through studies of the effectiveness of past practices, through insights acquired from observing other systems, such as Fleming's observation that mold in Petri dishes was inhibiting the growth of microorganisms that led to the discovery of penicillin, and through the study of other species, such as the use of laboratory animals for experimental purposes.

One aspect of medical research that is extraordinarily different from research in many of the social sciences is an extreme reluctance to draw conclusions from data that would require making the assumption of underlying decomposability. Medicine has long recognized that interactions between attributes that impact fitness are the rule rather than the exception and that such interactions can make drawing conclusions from statistical data very risky. As an example, consider the deceptively simple question, “Is coffee good for you?” According to a WebMD article (Kirchheimer, 2004), one survey conducted over the course of 18 years involving 126,000 people found that men who had six cups per day or more experienced a reduced risk for type II diabetes of 54% (there was, however, an interaction with sex, however, since the reduction was only 30% for women). Despite the presence of incredibly strong statistical significance, the overriding conclusion was that more research was necessary. That conclusion was, in itself, amazing because the WebMD article also noted “In recent decades, some 19,000 studies have been done examining coffee's impact on health.” To put this number in perspective, that single question has been addressed in a number of studies comparable to the estimated number of articles published globally in business and management over the course of an entire year (AACSB International, 2008, p. 10). The fact that such an extensive research effort has not led to definitive recommendations—despite the presence of high statistical significances—is a testament to the need for caution in drawing conclusions from observations gathered from a rugged fitness landscape. This reluctance also suggests a lesson for disciplines whose resources for research are vastly lower than those available for medical research (which would certainly include the informing sciences): brute force empirical investigations of phenomena that rely heavily on statistical tests of significance are unlikely to make major contributions to our understanding of fitness where the underlying landscape is rugged; moreover such an approach may well lead to misleading conclusions.

What may be a more effective approach to research in domains of high expected ruggedness and low research budgets is to focus on mapping out the rough underlying structure of the fitness domain, using both heuristic rules acquired through observations and logical or mathematic reasoning about the various situations observed. The mapping process may include identifying both sets of characteristics associated with apparent fitness peaks and sets of characteristics that partition a space into fundamentally different fitness functions. As an example of the latter, consider the business strategy research of Porter (1980, 1985), who proposes that three generic competitive strategies exist—cost leadership, differentiation, and segmentation—and asserts that the rules for effective competition (i.e., fitness) are fundamentally different depending upon which of these is chosen.

There is, however, another interesting insight that can be gained from the coffee example. The huge investment required for such a major, and largely non-theoretical, correlation-based study of health factors demonstrates a willingness to engage in experimentation as a tool for discovering new paths to fitness and not merely for confirming or refuting a theory. In this respect, medicine has more in common with the arts than the hard sciences. On the other hand, mainstream medicine is almost totally unwilling to make assertions of causality until the underlying physical mechanism that produces the observed correlation is determined. In this respect, it is totally aligned with the hard sciences. Thus we see how the rugged fitness landscape of medicine has led to the development of an extraordinarily large portfolio of acceptable research techniques.

Social Consequences of Rugged Landscape Research

Even if a researcher accepts that a particular domain being researched is rugged, adopting research methods appropriate to the domain may entail considerable social cost if that perception is not shared by the discipline as a whole. Within the sciences, particularly high levels of prestige accompany the development of attractive new theory. Accepting ruggedness is tantamount to conceding that what theory can be developed is likely to be ugly in its particulars. Such a concession by one individual is unlikely to be greeted enthusiastically by others in the discipline not holding the same perception.

A particularly interesting case study of this phenomenon can be found in the evolution of management research over the course of the past five decades. During the late 1950s, two studies (one conducted by the Ford Foundation, one conducted by the Carnegie Foundation) were sharply critical of U.S. business schools, with particular emphasis on the unscientific nature of their research. At the time, the most common form of research was the in-depth case study—a form of research closely corresponding to the likely output of the previously described reporter research role. The Ford Foundation report, for example, asserted:

Case collection is an important activity for the business school, both because of its contribution to teaching and because of its value as training for the faculty member. But case collection by itself is not research in the usual sense of that term. It can, however, become the raw material for research since, through careful and discriminating analysis, significant generalizations can sometimes be drawn from the study of a large number of cases. (Gordon & Howell, 1959, p. 385)

During the 1960s and 1970s these two reports, augmented by associated funding opportunities, exerted an extraordinary influence on the nature of business education and research (Khurana, 2007; Mintzberg, 2004; Starkey & Tiratsoo, 2007). Of particular note, the twin notions that reporting was not research and that the objective of research should be generalizations (i.e., theory) drawn from many observations—both evident in the preceding quote from the Ford Foundation—were clearly taken to heart. For example, by the year 2002, the reporter-type research that dominated the field in the 1960s (representing roughly 2/3 of articles in 1963 and 100% of articles in

1966) had entirely disappeared from the prestigious *Academy of Management Journal* (Colquitt & Zapata-Phelan, 2007, p. 1291). In gauging the significance of these numbers, it is also worth noting that the Academy of Management's other top-rated academic journal, the *Academy of Management Review*, is exclusively devoted to theory development. In fact, management's obsession with theory-building has become so extreme that even influential researchers in the field have started to complain about it. One such researcher used the example of the epidemiologist who first identified strong correlations between smoking and health problems in the 1930s, asserting that had that epidemiologist been in management field today, he would have been unable to publish his findings owing to their lack of theory-based justification (Hambrick, 2007, p. 1348). In stark contrast, we have already seen that the medical community is perfectly willing to gather and report observed findings and sometimes even recommend actions based upon particularly strong observed associations (e.g., prescribing medications for off-label uses). What they will not do is accord such findings the stature of being theory.

What is particularly extraordinary about the management discipline's devotion to theory is that it coexists with a nearly complete failure to inform practice. For example: (1) on one list identifying the fifty most important management innovations, not one originated from academic research (Pfeffer, 2007, p. 1336), (2) many of the most significant findings of human resources (HR) management research are widely disbelieved by HR managers (Rynes, Giluk, & Brown, 2007, p. 988), (3) managers use far more tools developed by consultants or other companies than tools developed by academics and they are also happier with those non-academic tools (Pfeffer & Fong, 2002, p. 88), and (4) important management ideas are most likely to originate in practice and then flow to academia, rather than the other way round (Barley, Meyer, & Gash, 1988). In short, the concerted attempts by management research to describe its landscape with attractive theory have not offered sufficient benefits to attract the attention of practitioners. That would, of course, be the expected result if the underlying landscape being researched is actually rugged, since attractive theory would probably not be particularly useful in guiding managers towards greater fitness under such circumstances.

Aside from the loss of prestige associated with reporting rather than theory-building, the skills necessary to be an effective reporter-researcher are also quite different. Earlier, it was pointed out that the effective reporter in a rugged domain would benefit from being a practical expert. The core of this argument is that there are likely to be far too many possibly relevant variables to establish values for all of them in a particular setting. Thus, the reporter needs to be selective, much like the doctor deciding what tests to run during the course of a complex medical diagnosis. That implies that researchers must possess considerable expertise in domain practice and also training in research techniques appropriate for in-depth data acquisition in the field. Such training is likely to be very different from that associated with educating researchers for a career of theory building and the analysis of large data sets. The mismatch between acquired and desirable skills represents a formidable barrier to changing research philosophies. It is, perhaps, part of the explanation for why disciplines historically cling to their existing research paradigms long after their shortcomings in explaining their research domain have been exposed (Kuhn, 1970).

Conclusions

The general goals of this paper have been to argue: 1) that rugged fitness landscapes represent a very different research domain from decomposable landscapes, 2) that numerous attributes, inter-relatedness of attributes, and dynamic changes to fitness all contribute to landscape ruggedness and, 3) that the objectives and approaches to research that are appropriate are heavily influenced by the ruggedness of the fitness landscape. Specifically, decomposable domains lead to attractive theory that compactly describes large regions of the fitness landscape; rugged domains lend

themselves to identifying techniques for improving fitness since the underlying theory describing such domains is ugly—too large and with too many exceptions to be particularly compelling.

In the specific context of the informing sciences, the following arguments have been made:

- The building block of the informing sciences, the informing system, meets all the prerequisites for rugged landscape, i.e., many interacting elements in a dynamically changing environment.
- In the presence of such ruggedness, certain types of research approaches—particularly those that emphasize gathering rich observations of individual systems along with the history of their evolution—are likely to be more valuable to the discipline than a preoccupation with theory building.
- Where disciplines accept the fact that their underlying landscape is very rugged, research techniques consistent with the underlying landscape can be developed and research can make major contributions to practice. Medicine provides an excellent example of such a domain.
- Where disciplines ignore evidence of domain ruggedness and persist in efforts to construct attractive theory, they are likely to experience almost complete disinterest from the practitioner communities that they attempt to study. The field of management may represent one example of this phenomenon. MIS, the discipline most closely related to the informing sciences, has experienced similar disinterest (Gill & Bhattacharjee, 2007).

Despite the challenges presented by the rugged fitness landscape, tailoring a research approach to fit the landscape offers some distinct advantages as well. Whereas empirical research into decomposable systems tends to reward statistical acumen and data collection, the rewards for research in the rugged landscape accrue from immersion in that landscape and its processes; the statistical analysis that occurs once the rugged landscape has been mapped out is largely confirmatory in nature. In order to speed your research program—since time will be very much of the essence in most rugged landscape research—you will need to enlist the active assistance of the participants in the system you are researching; their role will be that of equal partners in your research, rather than subjects. As a natural consequence, the knowledge that you acquire will be of a type that is palatable to the individual, group, or organizational entities that inhabit the landscape being studied. The type of bottom-up research that leads to progress in understanding a rugged landscape produces, as a byproduct, a collection of stories and examples that communicate well to students and experts alike. Thus, there will tend to be an appreciative audience—outside of other researchers—for your findings. Furthermore, once you develop an appreciation for the entirety of a rugged landscape you are unlikely to become fixated upon the behavior of a particular specialized peak. In consequence, you are far less likely to find yourself proposing theories that generalize poorly or which provide highly efficient but brittle solutions that cause entities to fail in the face of a changing landscape.

Studying rugged landscapes, and recognizing them as such, will also tend to chip away at the silos that so often separate research disciplines. Particularly where informing is involved, attributes of different components in the informing system will participate in the non-decomposable relationships that lead to ruggedness. If we are to have any hope of better understanding these processes, it therefore follows that researchers with fundamentally different domains of expertise (e.g., task experts, experts in psychology, experts in technology, and experts in education) will need to work together in understanding how the various attributes interact. A single perspective or research paradigm will simply not be sufficient.

Sadly, the findings of the rugged landscape researcher will not be pretty. Fortunately, their potential for positive impact on the landscape being studied should more than compensate for their ugliness.

References

- AACSB International. (2008). *Report of the AACSB International Impact of Research Task Force*. Retrieved on 5/27/08 from http://www.aacsb.edu/resource_centers/research/Final/Impact_of_Research_Report-FINAL.PDF
- Bak, P. (1996). *How nature works: The science of self-organized criticality*. New York, NY: Copernicus-Springer Verlag.
- Barley, S. R., Meyer, G. W., & Gash, D. C. (1988). Cultures of culture: Academics, practitioners and the pragmatics of normative control. *Administrative Science Quarterly*, 33(1), 24-60.
- Cohen, E. (1999). Reconceptualizing information systems as a field of the transdiscipline informing science: From ugly duckling to swan. *Journal of Computing and Information Technology*, 7(3), 213-219.
- Colquitt, J. A., & Zapata-Phelan, C. P. (2007). Trends in theory building and theory testing: A five decade study of the Academy of Management Journal. *Academy of Management Journal*, 50(6), 1281-1303.
- Gandolfi, A. E., Gandolfi, A. S., & Barash, D. (2002). *Economics as an evolutionary science: From utility to fitness*. New Brunswick, NJ: Transaction Publishers.
- Gersick, C. J. G. (1991). Evolutionary change theories: A multilevel exploration of the punctuated equilibrium paradigm. *Academy of Management Review*, 16(1), 10-36.
- Gilbert, D. (2007). *Stumbling on happiness*. New York: Knopf.
- Gill, T. G. (1995). High-tech hidebound: Case studies of information technologies that inhibited organizational learning. *Accounting, Management & Information Technologies*, 5(1), 41-60.
- Gill, T. G. (1996). Expert systems: Task change and intrinsic motivation. *MIS Quarterly*, 20(3), 301-329.
- Gill, T. G. (2005). *Introduction to programming using Visual C++.NET*. Hoboken, NJ: Wiley.
- Gill, T. G. (2008). A psychologically plausible goal-based utility function. *Proceedings of the DSI Annual Meeting, Baltimore, MD*. 1091-1096.
- Gill, T. G., & Bhattacharjee, A. (2007). The informing sciences at a crossroads: The role of the client. *Informing Science: the International Journal of an Emerging Transdiscipline*, 10, 17-39. Retrieved from <http://inform.nu/Articles/Vol10/ISJv10p017-039Gill317.pdf>
- Gill, T. G. & Hicks, R. (2006). Task complexity and informing science: A synthesis. *Informing Science: the International Journal of an Emerging Transdiscipline*, 9, 1-30. Retrieved from <http://inform.nu/Articles/Vol9/v9p001-030Gill46.pdf>
- Gill, T. G., & Holton, C. F. (2006). A self-paced introductory programming course. *Journal of Information Technology Education*, 5, 95-105. Retrieved from <http://jite.org/documents/Vol5/v5p095-105Gill114.pdf>
- Gill, T. G., & Jones, J. (2008). A tale of three courses: Case studies in course complexity. *Proceedings of the DSI Annual Meeting, Baltimore, MD*, 2091-2096.
- Gill, T. G., & Sincich, A. (2008). Illusions of significance in a rugged landscape. *Informing Science: the International Journal of an Emerging Transdiscipline*, 11, 197-226. Retrieved from <http://inform.nu/Articles/Vol11/ISJv11p197-226GillIllusions.pdf>
- Goold, A., & Rimmer, R. (2000). Factors affecting performance in first-year computing. *ACM SIGCSE Bulletin*, June, 32(2), 39-43.
- Gordon, R. A., & Howell, J. E. (1959). *Higher education for business*. New York, NY: Columbia University Press.
- Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance*, 16(2), 250-279.
- Hagan, D., & Markam, S. (2000). Does it help to have some programming experience before beginning a computing degree program? *ITiCSE Helsinki, Finland*, July, 25-28.

- Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal*, 50(6), 1346-1352.
- Handy, C. (1990). *The age of unreason*. Boston, MA: Harvard Business School Press.
- Hawk, T. F., & Shah, A. J. (2007). Using learning style instruments to enhance student learning. *Decision Sciences Journal of Innovative Education*, 5(1), 1-19.
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks, CA: Sage Publications.
- Holland, J. H. (1992). *Adaptation in natural and artificial systems*. Cambridge, MA: MIT Press.
- Holden, E., & Weeden, E. (2003). Software development: The impact of prior experience in an information technology programming course sequence. *Proceeding of the 4th Conference on Information Technology Education*. Lafayette, Indiana, October 16-18, 41-46.
- Kauffman, S. A. (1993). *Origins of order: Self organization and selection in evolution*. Oxford, U.K.: Oxford University Press.
- Khurana, R. (2007). *From higher aims to hired hands*. Princeton, NJ: Princeton University Press.
- Kirchheimer, S. (2004). Coffee: The new health food? *WebMD*. Accessed on 9/8/2008 from <http://men.webmd.com/features/coffee-new-health-food>.
- Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed., enlarged). Chicago, IL: University of Chicago Press.
- Lawrence, P. R., & Nohria, N. (2002). *Driven: How human nature shapes our choices*. San Francisco, CA: Josie-Bass.
- Mintzberg, H. (2004). *Managers not MBAs*. San Francisco, CA: Berrett-Koehler.
- Pfeffer, J. (2007). A modest proposal: How we might change the process and product of managerial research. *Academy of Management Journal*, 50(6), 1334-1345.
- Pfeffer, J., & Fong, C. (2002). The end of business schools? Less success than meets the eye. *Academy of Management Learning & Education*, 1(1), 78-95.
- Porter, M. E. (1980). *Competitive strategy*. New York, NY: Free Press.
- Porter, M. E. (1985). *Competitive advantage*. New York, NY: Free Press.
- Reiss, S. (2000). *Who am I? The sixteen basic desires that motivate our actions and define our personalities*. New York, NY: Berkley Books.
- Rynes, S. L., Giluk, T. L., & Brown, K.G. (2007). The very separate worlds of academic and practitioner publications in human resources management: Implications for evidence-based management. *Academy of Management Journal*, 50(5), 987-1008.
- Sackrowitz, M., & Parelius, A. (1996). An unlevel playing field: Women in the introductory computer science courses. *ACM SIGCSE Bulletin, Proceedings of the Twenty-Seventh SIGCSE Technical Symposium on Computer Science Education*, March, 28(1), 37-41.
- Simon, H. (1981). *The sciences of the artificial* (2nd ed.). Cambridge, MA: MIT Press.
- Starkey, K., & Tiratsoo, N. (2007). *The business school and the bottom line*. Cambridge, UK: Cambridge University Press.
- Wilson, B., & Shrock, S. (2001). Contributing to success in an introductory computer science course: A study of 12 factors. *SIGCSE, Charlotte, NC*, February, 184-188.
- Wood, R. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, 37, 60-82.

Biography



Grandon Gill is an Associate Professor in the Information Systems and Decision Sciences department at the University of South Florida. He holds a doctorate in Management Information Systems from Harvard Business School, where he also received his M.B.A. His principal research areas are the impacts of complexity on decision-making and IS education, and he has published many articles describing how technologies and innovative pedagogies can be combined to increase the effectiveness of teaching across a broad range of IS topics. Currently, he is an Editor of the *Journal of IT Education* and an Associate Editor for the *Decision Science Journal of Innovative Education*.