

Illusions of Significance in a Rugged Landscape

T. Grandon Gill and Terry L. Sincich
University of South Florida, Tampa, Florida, USA

ggill@coba.usf.edu, tsincich@coba.usf.edu

Abstract

A great deal of academic research, particularly in the social sciences, makes an implicit assumption that the processes being investigated are largely decomposable, meaning that the effects of individual variables identified in one context are likely to be observed in other contexts as well. Many of the statistical tools employed in analyzing social science data, such as regression and structural equation modeling, implicitly depend on such decomposability. A companion paper (Gill, 2008) proposes that many informing system situations may actually exist on rugged fitness landscapes for which the decomposability assumption is unrealistic. What this paper demonstrates is that applying statistical tests that assume decomposability on rugged landscapes may lead to statistical illusions of significance. These illusions, in turn, may convince the researcher that the decomposability assumption is valid, posing a serious threat to rigor by vastly overstating the statistical significances of observed relationships and by producing spurious significances. These significances, moreover, will not be readily detected by statistical tests that are commonly used to check the validity of the regression assumptions, such as the normality of error terms. This paper also explores the underlying reasons for such illusions and under what circumstances they can be avoided.

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

Introduction

Suppose you had collected, or had been given, a set of data summarizing the characteristics of a great many informing systems, including information relating to the sender, the client, the delivery system, and the associated task for each system. Further suppose that you also had, for each system, one or more measures of system success, such as user satisfaction, degree of use, and contribution to profitability. Under such circumstances, it would be perfectly natural to want to explore the degree to which different informing system characteristics contribute to success. Furthermore, if you had the proper training—such as a doctorate in MIS research—you would have at your disposal a variety of statistical tools, such as multiple regression, that could be employed to perform such analysis.

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.

The use of such tools involves an implicit assumption. Specifically, they assume that the underlying process that generated the data is decomposable. What this means in the case of a basic multiple regression model (i.e., a model with a single term for each independent variable—often called a first-order model) is that each variable in the model impacts the success measure in a man-

ner that is independent of the values of the remaining variables. Where such decomposability is not predicted to be present, it is possible to incorporate specific interaction terms, thereby modeling the effects of combinations of more than one independent variable value. The potential number of such terms, however, is extraordinarily large for all but the most trivial set of characteristics. Thus, decomposability tends to be assumed unless strong evidence to the contrary exists.

In a companion paper (Gill, 2008), a conceptual argument is presented that the assumption of decomposability is unlikely to be valid where the characteristics that drive the success of an informing system are being studied. Instead, it proposes that the relationship between system characteristics and system success is likely to be better described as a rugged fitness landscape, a conceptual framework developed for evolutionary biology (Kauffman, 1993). In rugged landscapes, a particular characteristic's impact on fitness cannot be determined without knowing the values of several more attributes. For example, an informing system that is useful, easy to use, and impacts a minor aspect of its user's job may be greeted with enthusiasm, signifying a high value on the "user acceptance" fitness scale. On the other hand, if a nearly identical system impacts a significant part of the user's job, the user's reaction could be resistance stemming from concerns about job loss. Moreover, the level of that resistance might further depend on the strength of various user motivational drives (e.g., Lawrence & Nohria, 2002). Users with a strong drive to bond might well place organizational needs on a par with their personal needs and thus welcome the system; users with a strong defensive drive—which manifests itself as the individual's need to protect possessions and status—might go so far as resorting to sabotage.

A compelling argument that can be advanced against the rugged fitness landscape model of the companion paper is the large body of research that has found significant relationships between system characteristics and fitness using statistical techniques that assume decomposability. Indeed, one of the authors has advanced such models himself in the past (e.g., Gill, 1996). Multiple regression, for example, provides an estimate of the significance for each model coefficient that it estimates. As researchers, we typically ignore any relationship that could happen by chance more than one time in twenty ($p >= 0.05$). Frequently, our tests tell us that what we observe could be explained as the result of random variations in the data less than one time in a hundred ($p < 0.01$). If the assumption of decomposability is invalid, how is it possible that significance tests so often reveal relationships that are so unlikely to have been caused by chance?

In the present paper, we address the following question: Under what circumstances can a rugged fitness landscape produce statistical illusions of significance that would tend to reinforce a researcher's assumption of decomposability? We begin by introducing the concept of a fitness landscape and Kauffman's (1993) NK model as a tool for modeling such landscapes. We then proceed through a series of simulations that demonstrate how the use of multiple regression to analyze observations from such landscapes can produce strong illusory significances, both overstating the statistical significances of some coefficients and identifying relationships that are not part of the underlying process used to generate the data as being significant. These problems are of particular concern because they do not stem from typical departures from the regression assumptions—such as the independence of adjacent observations and the normality of error terms. As a consequence, the problems and their underlying source could be easily overlooked when results of analysis are inspected, particularly since research findings that identify significant relationships can offer considerable career benefits to the researcher. We also describe the underlying statistical causes of these illusions. Finally, we identify the circumstances under which problems such as the ones that we have identified could prove to be a serious threat to research validity.

We have intentionally kept the focus of the article very narrow, intending that it serve as a complement to the companion article (Gill, 2008) which considers the implications of landscape ruggedness for informing system research from a much broader perspective. To keep the body of the

paper from becoming too technical, details relating to the construction of the various simulations described are presented in an appendix.

Fitness Functions and Landscapes

As described in the companion paper (Gill, 2008), 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 allocating income so as to purchase that basket of goods and services that maximizes utility.

In both of these cases, we would not expect entities (e.g., informing systems) existing on such landscapes to be distributed randomly. Where fitness signifies survivability, we expect low fitness entities to perish, leaving a distribution dominated by higher fitness entities. Where fitness drives choice, we would expect that higher fitness combinations would be observed more commonly than lower fitness combinations. This is, of course, the fundamental governing principle of evolutionary theory: that species migrate towards increased fitness over time.

Kauffman’s NK Landscape Model

Kauffman’s (1993) NK landscape model is a mathematical framework that can be used to describe the qualitative characteristics of a fitness function. It originated as means of characterizing the fitness of a chromosome, with the N referring to the number of genes. The K , in turn, refers to the 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 an overall fitness model that can be presented in the form:

$$y_1 + y_2 + \dots + y_N$$

where y_i is some transformation $T_i(x_i)$ that returns the impact of x_i on fitness. In a basic linear regression model, the most common type employed in social science research, $T_i(x_i)$ would simply be $a_i x_i$, where a_i is a multiplicative coefficient. That is, the underlying regression model has the form: $F = c + a_1 x_1 + a_2 x_2 + \dots + a_N x_N$

We refer to this landscape as the *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. In this case of complete interdependency, no meaningful estimate of fitness can be made without knowing the val-

ues of all N characteristics. We refer to this landscape as the *chaotic landscape*, or *maximally rugged landscape*. In a regression context, the underlying model for the rugged landscape would include all possible interactions among the N independent variables: $F = c + a_1x_1 + \dots + a_Nx_N + a_{1+N}x_1x_2 + a_{2+N}x_1x_3 + \dots + a_{2^N-1}x_1x_2x_3\dots x_N$

For a decomposable $(N,0)$ fitness landscape, there will be a single fitness peak where 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 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 (allowing for a base case) plus a constant. Assuming that we had enough observations so that there was a 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 the basic regression model that includes only N coefficients: one for each of the independent variables), 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 characteristic values are considered, the estimated number of peaks will be given by the formula (Kauffman, 1993, p. 47):

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

An example of this, for an NK space of dimension 6,5, is presented in Figure 1, where the peaks are indicated by the values in parentheses. In this case, the estimate would be for roughly 9 peaks (64/7), with the actual number being 10. In interpreting Figure 1, it is important to realize that adjacency in the tabular display does not necessarily imply adjacency as defined in the NK model, which refers to bitwise adjacency. For example, $F(1,0,1,0,0,1)$ has a value of 0.852 and is a fitness peak. This can be determined by comparing its value to the 6 adjacent cells $F(0,0,1,0,0,1)$, $F(1,1,1,0,0,1)$, $F(1,0,0,0,0,1)$, $F(1,0,1,1,0,1)$, $F(1,0,1,0,1,1)$, and $F(1,0,1,0,0,0)$ whose values (lightly shaded) are 0.658, 0.115, 0.841, 0.843, 0.842 and 0.178, respectively. (A more detailed look at constructing a rugged fitness landscape using a spreadsheet is presented in the appendix.)

	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 1: Local fitness peaks (in parentheses) on a randomly generated NK fitness landscape of dimension 6,5. Row labels represent the first 3 characteristic values, column labels represent the remaining 3 values.

As the value of K declines, we should expect the number of peaks in the space to decline. This is most easily demonstrated for situations where our fitness space is partitioned between those attributes that act decomposably and attributes that act interdependently, e.g.,

$$F = f(x_1, x_2, \dots, x_K) + y_{1+K} + y_{2+K} + \dots + y_N$$

In this case, we have a chaotic fitness landscape of $K, K+1$ combined with $N-K$ decomposable characteristics. Since all local peaks will occur where the decomposable portion is maximized, it follows that the multiple peaks will all be the result of the interactions between x_1 through x_K , meaning we would expect $2^K/(K+1)$ peaks rather than the larger $2^N/(N+1)$ number of peaks we would have estimated if the entire landscape were chaotic.

Experimental Simulations

The reasons that a rugged landscape might be a good model for informing system fitness are presented in the companion paper (Gill, 2008). What we now consider is the outcome of using multiple regression to analyze observations taken from an underlying landscape that is rugged.

Simulation Design

The design used for the simulations involved two separate manipulations: 1) nature of the underlying landscape and 2) presence or absence of migration.

With respect to the underlying landscapes, there are many different structures that can lead to “ruggedness”. As a consequence, assuming any particular landscape shape may not prove generalizable. For this reason, we test three alternative landscapes. These are defined as follows:

- *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). This is equivalent to Kauffman’s $N, N-1$ landscape, with fitness values generated randomly, and represents the case where virtually no significant coefficients should be found when coefficient values for a standard linear model of the form $c_0 + a_1x_1 + a_2x_2 + \dots + a_Nx_N$ are estimated using multiple linear regression.
- *Mixed*: Fitness was determined in two parts: by a linear part involving a subset of the attributes, $a_1x_1 + a_2x_2 + \dots + a_Dx_D$, and by a non-decomposable function of *all* the relevant attributes, $f(x_1, x_2, \dots, x_N)$, where N is the total number of attributes. This form is not well described in N, K terms (since there are interactions between all variables in addition to an independent component for D of the variables). It is, however, the basis of a commonly used regression technique in which the coefficients a_1 through a_D represent *main* effects, while separately constructed terms are employed to capture *interaction* effects. Our design, however, presumes that the researcher believes the fitness space to be decomposable, so no interaction terms are included in the analysis. In the absence of contrary theory predicting interactions, assuming no interaction would not be unreasonable from a practical standpoint since the number of possible interaction terms is huge. For example, if x_1 through x_N are all binary (0,1) variables, consistent with Kauffman’s model, to capture all the N -way interactions between the variables would require 2^N separate terms.
- *Partitioned*: The underlying function used to create fitness values was in two parts: a non-decomposable function of K attributes, $f(x_1, x_2, \dots, x_K)$, and a linear function of the remaining attributes, $a_{1+K} y_{1+K} + a_{2+K} y_{2+K} + \dots + a_N y_N$. It is intended to capture situations where variables *either* contribute decomposably *or* through interactions, but not both ways.

With respect to making migration our second manipulation, we propose the following rationale. The main characteristic distinguishing a fitness function from multi-argument functions in general is that the value being returned has an associated desirability or survivability trait associated with it. If the entities on the landscape are intelligent and adaptable, it would then follow that they should attempt to reposition themselves on the landscape so as to achieve maximum fitness. This is, of course, the justification of axioms that provides the mathematical basis for most of classical economics (utility maximization) and much of contemporary finance (maximizing shareholder wealth). It would therefore follow that any quantitative analysis of observations drawn from a fitness landscape must consider the possibility that entities will not be randomly distributed over the fitness landscape. Instead, we would expect to see them occupying, or migrating to, local peaks or global peaks. As a consequence, each landscape received two treatments: the first assuming the entities we were observing were uniformly distributed across the landscape (in the 8 variable landscapes we used for our base case, that meant we used 2^8 or 256 observations representing each possible combination of 0s and 1s for all variables), the second assuming that entities had climbed to their local fitness peak. As part of our second analysis, we also performed sensitivity analysis that examined intermediate results during periods while entities were migrating to fitness, but had yet to reach their local peak.

Simulation Hypotheses

The underlying scenario of our experiments is that a researcher has acquired observational data from a rugged fitness landscape that includes a single fitness value (dependent variable) and multiple explanatory characteristics (independent variables) for each observation. We further assume that the researcher is unfamiliar with the structure of the landscape and therefore assumes that characteristics contribute to fitness in decomposable fashion, making a basic multiple regression model, $F = c + a_1x_1 + a_2x_2 + \dots + a_Nx_N$, an obvious tool to apply. The companion paper (Gill, 2008) explains why it would be in the researcher's best interest to use such an assumption as a starting point.

This scenario motivates our main research question: *Can applying multiple linear regression analysis to observations drawn from a rugged fitness landscape produce misleading results?*

We can operationalize the term “misleading” in the question with two specific hypotheses:

- H1. Using regression analysis to estimate the process underlying a rugged landscape will produce coefficients that are reflective of the underlying process for explanatory variables that contribute decomposably.
- H2. Using regression analysis to estimate the process underlying a rugged landscape will produce non-significant coefficients for explanatory variables that only contribute through interactions with other variables in the underlying process.

The rationale behind these hypotheses is simple. For the first hypothesis, if regression analysis is an appropriate tool for rugged landscapes, we would expect it to be able to extract sensible values for those coefficients that are, in fact, contributing in the linearly additive manner that regression presupposes. For the second hypothesis, where strong interactions between variables are present, we would want our regression to indicate this lack of fit by producing non-significant coefficient estimates. We state the hypotheses so that each null hypothesis represents the case where misleading values are obtained because, through experimental design, we know that the processes we are trying to fit do not match the structure of the regression equation. Thus, it makes sense to make our base case the assumption that errors will be introduced.

Experiment #1: Chaotic Landscape, No Migration

This particular case is one of two trivial cases that can be addressed without extensive reporting of experimental results (see appendix Example 1, Exhibit 5). Because fitness values were randomly assigned to each possible combination of variables, when you perform a regression to estimate fitness using the characteristics as explanatory variables, you get very low R^2 - values and statistically non-significant beta coefficients. Only the constant value is statistically significant and reflects the average of fitness for all observations.

Thus, we find H2 is supported (H1 cannot be tested since there are no decomposable independent variables). In other words, in the absence of migration, using regression to analyze observations drawn from a chaotic fitness space does not produce any misleading results. The only danger such a landscape would present would be encouraging the researcher to conclude that the characteristics involved do not impact fitness—which is patently false. The correct conclusion would be that the impact of the characteristics involved on fitness is not well described by a decomposable linear model.

Experiment #2: Mixed Landscape, No Migration

This particular case is second trivial case. Because fitness values were randomly assigned to each possible combination of variables, that portion of fitness that results from interactions between variables is treated by the equation as an error term. As a consequence, the analysis produces reasonable estimates for the decomposable coefficients and non-significant estimates for the remaining coefficients. Example 3, Exhibit 8 in the appendix illustrates this for a simulation with 8 variables, 4 of which also exert a decomposable impact on fitness.

Here, we find both H1 and H2 are supported. Once again, the only danger here from a research standpoint is that the researcher might conclude, based upon the non-significance of the non-decomposable coefficients, that these variables do not contribute to fitness. That would be a mistake, of course, since they do contribute—but only in a very complex way (i.e., through interactions).

Experiment #3: Partitioned Landscape, No Migration

On the surface, the partitioned landscape seems similar to the mixed landscape. Actually, it is subtly different in a way that turns out to make a big difference to our results. Consider the partitioned process:

$$F = f(x_1, x_2, \dots, x_K) + y_{1+K} + y_{2+K} + \dots + y_N$$

Because the chaotic and mixed cases assume that all variables interact, in each simulation we assign a different random number to each combination. For the partitioned case, however, we only assign random values to each possible combination of non-decomposable values. If, for example, we have 8 variables in total and 2 are decomposable, the remaining 6 variables (K in the equation above) contribute through interactions with each other. That means that only 64 random numbers (2^6) need to be generated to simulate the interaction, and that the each value will be used 4 times. The results of 10 simulations, created using a process described in appendix Example 4, are shown below, in Table 1.

Table 1: Regression results for migrated peaks in partitioned space with 2 decomposable variables and 6 interacting variables including all 256 observations

	Peaks	Adj. R ²	V1	V2	V3	V4	V5	V6	V7	V8	# SIG
1	12	0.73	D	D	N	N	N	*****	***	*****	3
2	13	0.64	D	D	N	N	N	N	***	*	2
3	7	0.60	D	D	N	N	N	N	N	N	0
4	9	0.62	D	D	N	N	N	N	*	**	2
5	9	0.64	D	D	*	***	*	N	*	N	4
6	9	0.63	D	D	***	N	N	N	N	*	2
7	10	0.63	D	D	*****	N	N	N	N	N	1
8	11	0.59	D	D	N	N	N	N	N	N	0
9	6	0.59	D	D	*	N	*	N	N	N	2
10	13	0.63	D	D	N	*	N	*	N	N	2
	9.9	0.63	Averages								1.8
Legend: * → p<0.05, ** → p<0.01, *** → p<0.001, **** → p<0.00001, ***** → p<0.000001 N = Not significant, N/A = Not available (model is over fitted or no variation encountered). D = Decomposable, with exact fit, E = Decomposable, with error in coefficient estimate											

What these results show is that 1.8 of the non-decomposable variables, on average, showed significance of p<0.05 or better—often much better—for 10 separate replications of the experiment. Also, because we constructed the underlying process so that half the variance would be explained by the decomposable part, we expect R² to be about 0.5. Instead, the value averages 0.13 higher, implying that the (randomly generated) non-decomposable portion of the landscape explains the additional variance—which should not happen. Moreover, sampling error does not seem to be the culprit, since we generated an observation for each of the 256 possible combinations of independent variables; there was no sampling involved.

What is particularly remarkable about these results is the number of high significances encountered in variables that contribute only through interactions. As a consequence, we must reject H2. Using regression to analyze a partitioned fitness landscapes can produce erroneous significances. Moreover, we replicated these results on fitness landscapes of up to 12 variables (4096 observations) and the trend of large numbers or erroneous significances persisted. Hypothesis H1, on the other hand, was fully supported. The regressions did, in fact, produce extremely precise estimates for the coefficients of the decomposable variables. Unfortunately, such accuracy might serve to build the researcher’s confidence in the non-decomposable estimates as well.

Although these results appear bewildering at first glance, the explanation for them is quite simple. Consider the underlying process we used to generate the observations:

$$F = f(x_1, x_2, \dots, x_K) + y_{1+K} + y_{2+K} + \dots + y_N$$

As expected, the regression algorithm used to estimate coefficients has no trouble detecting the significance of the decomposable coefficients (1+K through N). Since the t-tests on the beta coefficients in regression test the effect of each independent variable after accounting for the effects of the other independent variables in the model, essentially the algorithm removes the decomposable variables and the variance they explain from the equation. The remaining (non-decomposable) variables, 1 through K, are then used to explain the remaining variance. In our example, that would mean that we have effectively modeled a chaotic landscape using 6 terms. The problem is that this regression now has 4 copies of each observation. Unfortunately, such duplication of observations creates an artificially inflated sample size, which in turn causes the regression equation to yield significant t-tests on negligible terms. This is demonstrated in Table

2, which reports the results from partitioning our process into 4 decomposable variables and 4 interacting variables. With this process, there are only 16 (2^4) possible combinations of the interacting variables, which means they would be repeated 16 times in the 256 observation sample. As should be evident from the table, nearly 3 out of 4 estimated coefficients for the non-decomposable variables are significant across 10 replications. In fact, they tend to be *highly* significant.

Table 2: Regression results for fitness in partitioned space with 4 decomposable variables and 4 interacting variables including all 256 observations

	Peaks	Adj. R ²	V1	V2	V3	V4	V5	V6	V7	V8	# SIG
1	2	0.65	D	D	D	D	*****	N	***	***	3
2	3	0.62	D	D	D	D	*	N	*****	*****	2
3	2	0.66	D	D	D	D	N	*****	N	**	2
4	4	0.49	D	D	D	D	*****	***	*	N	3
5	4	0.62	D	D	D	D	*****	***	*****	N	3
6	2	0.53	D	D	D	D	*	*	*****	N	3
7	2	0.51	D	D	D	D	N	*****	**	*****	3
8	2	0.62	D	D	D	D	****	*****	N	N	2
9	4	0.66	D	D	D	D	*****	**	*****	***	4
10	5	0.59	D	D	D	D	*****	***	*****	**	4
	3.0	0.60	Averages								2.9
Legend: *→p<0.05, **→p<0.01, ***→p<0.001, ****→p<0.00001, *****→p<0.000001 N = Not significant, N/A = Not available (model is over fitted or no variation encountered). D = Decomposable, with exact fit, E = Decomposable, with error in coefficient estimate											

The fact that the source of the problem is straightforward does not make it any less significant. Unless the researcher happens to be very suspicious of his or her results—as a consequence of a strongly grounded theory being tested, for example—there would be little reason to dispute the results of a regression analysis that otherwise looks outstanding. Moreover, incorporating interaction terms would not be a trivial matter. In our example with 2 decomposable variables, 64 separate terms would be required. If, however, we also allow for the possibility of 2-way, 3-way and 4-way interactions, the number of terms grows even larger and the difficulty of identifying the most appropriate interactions grows as well. Thus, we perceive the problem of erroneous significances in partitioned landscapes to be a potentially serious threat to empirical research.

Experiment #4: Chaotic Landscape, With Migration

The underlying chaotic landscape ceases to be a trivial case when entities are allowed to migrate to their local peaks. The technique used to simulate migration is discussed in Example 1 of the appendix and essentially involves entities travelling to peaks by altering the characteristic offering the greatest fitness gain for each position until a peak was attained (a method also used by Kauffman, 1993). Our 8,7 landscape would offer roughly 28.4 such peaks ($2^8/9$) according to Kauffman's (1993) formula. The results of the analysis for 10 simulations are presented in Table 3.

Table 3: Regression results for migrated peaks in 8,7 space

	Peaks	Adj. R ²	V1	V2	V3	V4	V5	V6	V7	V8	# SIG	
1	26	0.70	N	*****	*****	****	*****	*****	*	**	7	
2	24	0.51	*****	*****	*****	N	N	N	N	*	4	
3	31	0.70	N	*****	*****	****	*****	*****	*****	**	7	
4	31	0.21	N	*	N	***	N	*****	***	N	4	
5	30	0.49	N	*	N	*****	*****	*****	**	*****	6	
6	33	0.15	N	N	N	**	N	**	N	*****	3	
7	24	0.51	*****	**	*****	*****	N	*****	***	*****	7	
8	32	0.19	N	N	N	*****	N	*	N	N	2	
9	30	0.15	N	*	N	**	**	*****	N	**	5	
10	32	0.13	N	N	**	***	*	N	N	*	4	
	29.3	0.37	Averages									4.9
Legend: *→p<0.05, **→p<0.01, ***→p<0.001, ****→ p<0.00001, *****→p<0.000001 N = Not significant, N/A = Not available (model is over fitted or no variation encountered).												

These results provide a strong basis for rejecting H2 (again, H1 cannot be tested owing to lack of decomposability). Over half the interacting independent variables show misleading statistical significances. What makes them particularly misleading is the fact that they were drawn from peaks. In a completely decomposable process, if a coefficient is positive with a high degree of significance and represents a controllable variable (e.g., top management support in a system implementation), it would make sense for the researcher to recommend changing that characteristic in any entity for which it was set to a 0 value. If, however, all our entities have migrated to peaks, then changing the value to 1 on any peak where the value is 0 will lead to a decline in fitness; that is how a “peak” is defined.

The explanation for these significances is, essentially, the same as it was for Experiment 3. Through migration, what were originally 256 different observations have become roughly 29 clusters of identical observations. This clustering has the effect of creating lots of duplicates that artificially inflate the sample size; this, in turn, causes the regression algorithm to yield statistically significant coefficients for terms in the model that are really negligible (spurious significances). These significant coefficients could, in turn, ultimately entice the unaware researcher to draw erroneous conclusions about the variables that really effect fitness. The nature of the problem caused by convergence is illustrated in Figure 2. In this illustration, the X variable has a different impact on fitness at each of the two possible values of Y (y_0 and y_1). At x_0, y_0 there is a peak where fitness reaches a maximum of f_0 . At x_1, y_1 there is a second peak where fitness reaches a maximum of f_1 . Assuming observations (shown as squares) cluster around peaks with some error, the resulting regression line (shown with arrows at both ends) would show a high R-square—since a lot of variation in fitness is explained by the differences between x_0 and x_1 , and y_0 and y_1 —and yet the line would tell us absolutely nothing useful about the underlying process. With only two variables involved, the clustering of the data leading to this effect would be fairly obvious. With many interacting variables, it would be much harder to recognize. Also, as illustrated in the appendix Example 2, these types of coefficient significances remain relatively robust as random error is introduced into both dependent and independent variables.

Another possible objection to the observed significances might be that all entities are unlikely to have reached peaks. In Table 4, we test this for a sample 8,7 landscape. The *Start* column signifies the initial landscape, while the number columns indicate the regression values after each step in the regression, until all entities have reached peaks in step 4 (same as the *Final* column).

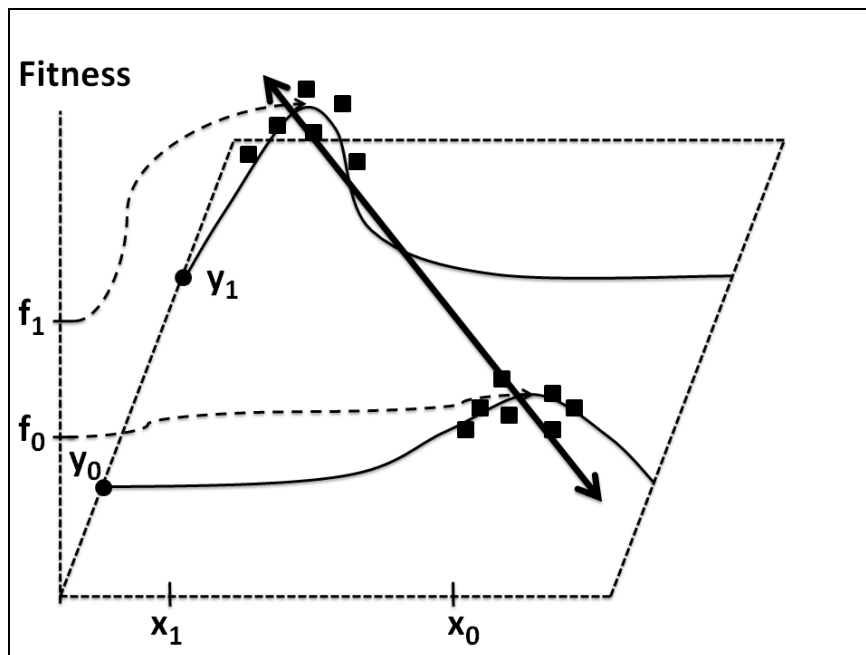


Figure 2: Illustration of peak convergence problem

The bottom row of the table shows the count of significant coefficients ($p < 0.05$). What this indicates is that as soon as a single step towards convergence is taken by entities on our fitness landscape, spurious significances start to arise. After just 3 steps, the number reaches its ultimate peak of 5, including many extremely significant coefficients. Thus, we must reject H2 even where migration is far from complete.

Table 4: Regression results on 8,7 landscape during migration to peaks

	Start	1	2	3	4	Final
R²	<0.000	0.0960	0.2084	0.2497	0.2377	0.2377
Const	0.504	0.934	0.966	0.962	0.964	0.964
Coeff 1	-0.006	-0.013	-0.021	-0.020	-0.019	-0.019
Coeff 2	-0.038	-0.023	-0.012	-0.019	-0.019	-0.019
Coeff 3	0.018	0.029	0.029	0.033	0.032	0.032
Coeff 4	0.048	-0.005	-0.012	-0.005	-0.005	-0.005
Coeff 5	-0.042	-0.049	-0.044	-0.033	-0.033	-0.033
Coeff 6	0.024	-0.016	-0.021	-0.015	-0.016	-0.016
Coeff 7	-0.037	-0.001	-0.009	-0.011	-0.011	-0.011
Coeff 8	-0.025	-0.012	0.005	0.006	0.006	0.006
P Value 1	0.86890	0.27277	0.00712	0.00356	0.00355	0.00355
P Value 2	0.30893	0.04207	0.11219	0.00304	0.00341	0.00341
P Value 3	0.62995	0.00852	0.00006	0.00000	0.00000	0.00000
P Value 4	0.19915	0.68987	0.14024	0.49842	0.47122	0.47122
P Value 5	0.26238	0.00003	0.00000	0.00000	0.00000	0.00000
P Value 6	0.52397	0.14881	0.00775	0.03164	0.02271	0.02271
P Value 7	0.33131	0.96027	0.24959	0.12783	0.14751	0.14751
P Value 8	0.50743	0.28962	0.46492	0.33519	0.30806	0.30806
Count	0	3	4	5	5	5

Experiment #5: Mixed Landscape, With Migration

When migration is incorporated in our mixed landscape, we see the same development of spurious significances that we observed for the 8,7 case with migration. The main difference is that the effect is even stronger, since few peaks develop. Since the results are qualitatively similar to those of the chaotic case, they are omitted.

Of greater interest, in the mixed case, is what happens to the decomposable variables over the course of the migration. Table 5 illustrates this process for a single case, similar to Table 4 presented previously. Based on the process that we established (see Appendix), our estimate for each decomposable coefficient (Coeff1 through Coeff4) should be 0.125 and the Const estimate should be 0.25 (reflecting the mean value of the non-decomposable contribution to fitness). This is approximately the case when the process starts.

Table 5: Migration steps of a mixed landscape from starting point to convergence

	Start	1	2	3	4	5	Final
# Independent	256	91	47	31	24	21	19
R ²	0.37	0.77	0.90	0.94	0.96	0.97	0.97
Const	0.26	0.48	0.49	0.50	0.51	0.52	0.52
Coeff 1	0.13	0.13	0.13	0.12	0.11	0.10	0.10
Coeff 2	0.10	0.08	0.08	0.08	0.08	0.09	0.09
Coeff 3	0.11	0.08	0.08	0.08	0.08	0.08	0.09
Coeff 4	0.11	0.13	0.13	0.13	0.13	0.13	0.13
P Value 1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
P Value 2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
P Value 3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
P Value 4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
P Value 5	0.167492	0.000060	0.000000	0.000000	0.000000	0.000000	0.000000
P Value 6	0.639501	0.000644	0.000000	0.000000	0.000000	0.000000	0.000000
P Value 7	0.429853	0.000363	0.000407	0.000002	0.000000	0.000000	0.000000
P Value 8	0.102342	0.469058	0.539482	0.240816	0.534751	0.416935	0.702114

As the process proceeds—towards a very high adjusted R² of 0.97—we see significances develop for 3 of the 4 interacting variables, leading us to reject H2 once again. We also see, however, a gradual erosion of the accuracy of the estimates for the 4 decomposable coefficients—with p-values suggesting that the estimates are far more accurate than they really are. Thus, the results of the experiment (which we replicated several times) cause us to reject H1 as well. In other words, when a basic linear regression model is fit to data from a mixed rugged fitness landscape with entity migration towards fitness, misleading estimates will not only result for the interacting (nondecomposable) variables, but also for the decomposable variables.

Experiment #6: Partitioned Landscape, With Migration

At the top level, the partitioned landscape can be decomposed into separate rugged, i.e., $f(x_1, x_2, \dots, x_k)$, and further decomposable, i.e., $a_{1+k} y_{1+k} + a_{2+k} y_{2+k} + \dots + a_N y_N$, parts. As a consequence, coefficient estimates for decomposable variables are not impacted by migration except to the extent that each variable quickly takes on a single value throughout the landscape—after which the coefficient values become impossible to determine. We verified this with several tests whose results are omitted. A demonstration that coefficient estimates remain valid as migration to fitness takes place in a purely decomposable process—representing the decomposable portion of the partitioned space—is presented in appendix Example 5. The only major concern surfaced by the experiment is that significances for the decomposable coefficient estimates proved to be vast-

ly overstated. Thus, H1 can be accepted for a partitioned decomposable process, with appropriate qualifications relating to significance over-estimates, while H2 remains rejected.

Summary of Experiments

The results of the experiments are summarized in Table 6. The shaded areas, Experiments 3 through 6, represent processes where H1 rejection (i.e., estimates of decomposable variable coefficients are misleading) and/or H2 rejection (i.e., spurious significances are attributed to interacting variable coefficients) occur as a consequence of the rugged fitness landscape characteristics of the underlying process.

Table 6: Summary of Experimental Results

	<i>No Migration</i>	<i>Migration</i>
<i>Chaotic</i>	Experiment 1 H1 Not tested H2 Supported	Experiment 4 H1 Not tested H2 Rejected
<i>Mixed</i>	Experiment 2 H1 Supported H2 Supported	Experiment 5 H1 Rejected H2 Rejected
<i>Partitioned</i>	Experiment 3 H1 Supported H2 Rejected	Experiment 6 H1 Qualified Support H2 Rejected

Discussion

As summarized in Table 6, the statistical illusions found in the six experimental simulations were concentrated in those processes where entities were undergoing migration towards fitness. What was not addressed, however, was the potential impact of the evolutionary process that involves failure of less fit entities, often referred to as “survival of the fittest”. In this section, we discuss this process from a conceptual standpoint. We then turn to the broader question of the likely prerequisites for the illusions we have discussed.

Survival of the Fittest

The mechanism for migration applied in Experiments 4 through 6 involved entities changing attributes so as to increase fitness, thereby reaching a local peak. This is consistent with the perspective taken by many business disciplines—most notably economics and finance—for which utility maximization is a central axiom. It is also consistent with many cognitive models of fitness-related activities, such as choosing what move to make in a game. Evolutionary models, on the other hand, show fitness increasing by a somewhat different mechanism. While search towards fitness does occur—at the gene level, employing various reproductive techniques such as crossing-over, inversion, and mutation (Holland, 1992)—the survival of high fitness entities (and non-survival of less fit entities) is the fundamental mechanism through which overall and individual entity fitness gradually increases from generation to generation.

Because it is not clear that low-level techniques for generating new gene combinations would have much applicability to informing situations involving intelligent entities, it would be hard to develop convincing simulation models of the statistical impact of such processes. We can, however, predict that the non-survivability of certain attribute combinations could potentially cause serious sampling issues that might interfere with the validity of research conclusions.

The problem occurs when some combinations of attributes in a rugged fitness landscape produce fitness values so low that the associated entity fails to survive. Where this occurs, subsequent

samples of the landscape will be unable to include the non-surviving entities attribute values. That, in turn, may lead to conclusions that are skewed with respect to these attributes.

To illustrate the process, we draw upon a concrete example taken from the human resources literature. Recently, the *Academy of Management Journal* (AMJ) ran a special issue (Volume 50, No. 5) on evidence-based management in the human resources field. Of particular concern to researchers was the degree to which practitioners were unaware of, or chose to ignore, some of the most significant findings of academic research. At the head of the AMJ list was the finding that “Intelligence predicts job performance better than conscientiousness (Schmidt & Hunter, 1998),” which elicited a particularly high level of disagreement from practitioners (Rynes, Giluk, & Brown, 2007, p. 988).

Two general comments can be made about the specific finding. First, the dependent variable—job performance—is clearly a fitness measure. Indeed, in the military, annual employee evaluations are referred to as “fitness reports”. Second, we may reasonably conclude that the performance landscape is rugged. This can be asserted rather than conjectured because it is relatively easy to identify instances where the needs of the organization would have been better served by an employee screening process that placed a higher premium on conscientiousness relative to intelligence. For example:

1. *Barings* (“Barings debacle,” 2008). A trader who failed to follow established trading guidelines took a firm that had survived the Napoleonic Wars and put it out of business.
2. *Societe Generale* (“French bank blames,” 2007). An employee who failed to establish trading guidelines lost \$7 billion of the firm’s money, crippling it.
3. *Chernobyl* (INSAG-7, 1992). Although the principal cause of the worst nuclear accident in history was an appallingly unstable reactor design, a lackadaisical attitude towards testing procedures on the part of operators was a significant contributor to the actual disaster.

Thus—bearing in mind that a large body of research found intelligence correlated with job performance (e.g., Schmidt & Hunter, 1998)—we may conclude that neither intelligence nor conscientiousness is fully decomposable. Their relative contribution to fitness will vary according to the values of other characteristics related to the job and, possibly, upon each other—with the highly intelligent/highly non-conscientious combination seeming to pose a particularly serious threat.

From a statistical standpoint, the fact that we have a fitness variable where insufficient fitness can lead to non-survival presents a serious empirical sampling issue. In our example, it would have been impossible to include either Barings or Chernobyl in any post-event study of the intelligence vs. conscientiousness tradeoff since one organization was disbanded and the referenced employees in the other all died. Thus, if conscientiousness-related failures occur at a rate lower than that of intelligence-related failures—and we make no assertions or conjectures on this point—the relative contribution of conscientiousness to survival could be severely under-sampled. Were this to be the case, employers with a sensible level of risk-aversion might be quite rational in preferring conscientiousness to intelligence in spite of the substantial body of empirical evidence that seems to support the benefits of intelligence.

Although they have not been framed in terms of a rugged fitness landscape model, concerns of the type we have expressed have been raised in the research literature. Consider, for instance, a statement made in the context of the example we have been discussing:

...concepts such as contingencies, configurations, complexity, “equifinality” and trade-offs all raise questions about the extent to which “average” findings can be generalized. (Rynes et al., 2007, p. 1001)

It would be hard to devise a clearer statement of the care that must be taken in making practical application of statistical findings derived from the analysis of observations taken from a rugged fitness landscape.

Prerequisites for Statistical Illusions

Are the statistical anomalies that have been identified in Experiments 3 through 6 and discussed in our “survival of the fittest” example potential threats to the validity of research in the informing sciences? In order for such research to be at risk, several preconditions must be met:

1. *The landscape we are trying to explain must be a consequence of a process of unknown structure.* If we know the structure of the underlying process, then we will not try to fit it with an inappropriate model. In the example offered by the companion paper (Gill, 2008), common sense alone would keep us from trying to perform a regression analysis of cookbook recipe ingredients on the judged tastiness of the prepared dishes. If we did, how would we interpret our results? Suppose we found that garlic was a statistically significant contributor to tastiness? Surely that wouldn't mean that we could improve our recipe for butterscotch brownies by adding garlic to it. Landscapes that we intuitively know to be rugged are not the problem; it is the landscapes of undetermined structure that are of concern.
2. *The variable that we are trying to explain needs to be some sort of fitness measure.* Unless the dependent variable involves desirability or survivability, there is no reason to anticipate that migration of entities towards fitness will occur. That immediately eliminates the scenarios posing the greatest threat to statistical validity.
3. *The entities involved in the process need to be intelligent and adaptable.* Similar to (2), if entities cannot adapt, there is no reason to expect migration to fitness to occur except through a much more drawn out process of survival of the fittest.
4. *Observational data needs to be an important contributor to our understanding of the landscape.* In some disciplines, such as economics, theories of fitness can be entirely mathematical in their derivation. In these cases, our tests of theory can be independent of the process by which theory is created—meaning the illusions we have presented may cause us to reject a theory that does not predict them (which is reasonable), but will not lead us to propose an erroneous theory based on the illusions we have observed. The threat to validity presents itself in those situations where our theory originates from our attempts to explain observations drawn from the same fitness landscape that we later use to test our theory. In this process, the illusions of significance that we have described could well make their way into theory. Should that happen, then our equally ill-conceived subsequent tests will tend to support our observation-derived theory.

In addition to these four prerequisites, we would generally expect that a rugged fitness landscape would exhibit certain characteristic features. Among these:

5. *Plausible examples inconsistent with the general conclusions produced by a decomposable model are observed or can easily be constructed.* The case of conscientiousness, presented in the previous section, would be typical of a rugged landscape, where simple rules (e.g., intelligent employees are better performers than conscientious employees) almost never apply across the entire fitness landscape.

6. *A distribution of entities that includes a very wide range of controllable characteristics is observed.* Where a landscape is decomposable, migration towards fitness will tend to produce convergence of characteristics, as a characteristic that enhances fitness of one entity enhances fitness of every other entity. In particular, we would expect that any characteristic that can be controlled by an entity will increasingly converge to the same value across all entities. Where a landscape is rugged, however, a diversity of characteristics across entities will be present—as local fitness peaks will demand different combinations.

If items (1) through (6) are all present, then there is ample reason to believe that the fitness landscape is rugged and that statistical techniques assuming decomposability should only be applied with the greatest of care. Even if migration is not occurring, we must still recognize that the characteristics of the underlying landscape are inconsistent with the structure of the equation we are attempting to model it with. Under such circumstances, the best we can hope for is that we do not obtain spurious significances as part of our results.

Conclusions

Within the social science literature, particularly the business research literature, the empirical research paper generally follows a familiar pattern. A problem is stated. The relevant literature is reviewed. A theory, based upon the literature, is presented and hypotheses consistent with that theory are proposed. A methodology through which empirical data was gathered is described. Analysis of the data is performed and the results are contrasted with the proposed hypotheses. Conclusions are then presented.

Over the years, concerns regarding the practical relevance of the research conducted in this manner have often been raised (e.g., Pfeffer, 2007). Researchers, however, have generally been far more complacent with respect to the rigor of that same research. We are, for example, very careful not to draw conclusions from any evidence that could have occurred by chance more than 1 time in 20 and often estimate the odds of our findings being coincidental to be far more remote.

The power of our statistical tests, however, depends upon the degree to which the underlying process that generates our data conforms to the assumptions built into the analytical tool we employ. In the case of multiple regression with a linear model, and its close cousins factor analysis and structural equation modeling, the key assumption is that the explanatory variables contribute to the dependent variables decomposably. What the present paper has demonstrated is that, where this assumption is incorrect, serious inconsistencies between the results of the analysis and the actual process used to generate the data can emerge. In particular, estimates of significance that appear phenomenally good can be extremely misleading. They are, in fact, statistical illusions.

To understand the potential impact of these illusions, it is critical that we recognize that *they are NOT the result of randomness*. Although we have generated random numbers to simulate variable interactions for the purpose of our experiments, within the actual processes that we are mimicking the interdependencies will be quite real and consistent over time or, at least, consistent until forces such as coevolution (Kauffman, 1993) change the fundamentals of the fitness landscape. What this means is that illusions present in one set of observations are likely to be present in other sets of observations gathered independently. The fundamental problem is the lack of fit between the process and the tool used to analyze the process, not sampling issues or random variations. The practical consequence of this would be that erroneous models of process introduced early in a stream of research could easily be replicated and reinforced by later research. All the while, the perception of rigor is retained.

It is also important to note that the illusions discussed here particularly relate to the use of empirical data to develop and test theory. If the researcher's goal in using linear regression is simply to predict fitness values for observations, then the analyses performed may be perfectly appropriate

for establishing a “best guess”—particularly if a lot of data has been acquired. For example, if an individual likes garlic, then it is perfectly reasonable to consider the presence or absence of garlic in a recipe in predicting whether or not the individual will enjoy it. Where the problem arises is when that data is also used as a basis for theoretical conclusions about the underlying process that causes individuals to like or dislike particular recipes. Unlike the decomposable case, in a rugged landscape, the finding that the presence of garlic contributes to an individual’s preference for a recipe is not equivalent to predicting that adding garlic to a recipe will improve the same individual’s preference for that particular recipe.

In the companion paper (Gill, 2008), the argument has been presented that the types of processes that can create such illusions—situations where our dependent variable is some form of fitness and the impact of our independent variables on fitness is interdependent—are likely to be quite common for research efforts involving informing systems. As a consequence, we need to be very cautious in attributing significance to the statistical results based upon the analysis of observational findings. In particular, we need to be extraordinarily careful about developing “theory” based upon these results. The late Jens Mende (2005) has already warned the informing sciences about the dangers of empiricism from a philosophical perspective, asserting:

[A] discipline only qualifies for the status of a science after it has progressed beyond empirical generalisations to explanatory theories; but although empirical methods are useful for discovering the former, they are inherently useless for creating the latter. (Mende, p. 189)

Our findings, however, go a step further. Where a rugged fitness landscape is involved, casual analysis of empirical data without careful consideration of the underlying process may not even serve us well in achieving valid empirical generalizations.

References

- Barings debacle. (2008). Retrieved on 4/22/2008 from http://www.riskglossary.com/link/barings_debacle.htm
- French bank blames trader for \$7 billion fraud. (2007). *Associated Press*. Retrieved on 4/22/2008 from <http://www.msnbc.msn.com/id/22818054>
- Gill, T. G. (1996). Expert systems: Task change and intrinsic motivation. *MIS Quarterly*, 20(3), 301-329.
- Gill, T. G. (2008). Reflections on researching the rugged fitness landscape. *Informing Science: the International Journal of an Emerging Transdiscipline*, 11, 165-196. Retrieved from <http://inform.nu/Articles/Vol11/ISJv11p165-196Gill219.pdf>
- Holland, J. H. (1992). *Adaptation in natural and artificial systems*. Cambridge, MA: MIT Press.
- INSAG-7. (1992). *The Chernobyl accident: Updating of INSAG-1 -- A report by the International Nuclear Safety Advisory Group*. Vienna, Austria: International Atomic Energy Agency.
- Kauffman, S. A. (1993). *Origins of order: Self organization and selection in evolution*. Oxford, U.K.: Oxford University Press.
- Lawrence, P. R., & Nohria, N. (2002). *Driven: How human nature shapes our choices*. San Francisco, CA: Josie-Bass.
- Mende, J. (2005). The poverty of empiricism. *Informing Science: the International Journal of an Emerging Transdiscipline*, 8, 189-210. Retrieved from <http://inform.nu/Articles/Vol8/v8p189-210Mende.pdf>
- 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.

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.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262-274.

Appendix: Spreadsheet Examples of a Rugged Fitness Landscape

This appendix provides an outline for constructing a rugged fitness landscape using a spreadsheet, which should allow the reader to confirm or extend the results presented in the mathematical analysis portions of the paper.

Kauffman’s (1993) NK model is presented in terms of binary strings of 0 and 1. As a result, independent values landscape where N=8 will be represented as an 8 bit string, e.g., 01100101. It necessarily follows, that there will be 256 (2⁸) possible combinations for which fitness values need to be assigned, starting at 00000000 and ending at 11111111.

Random N,N-1 Landscapes

To simulate the 8,7 landscape (any other N,N-1 landscape can be simulated by extending the technique), we created an 8 column range with all possible bit permutations in order then, in the adjacent column, copied 256 random number values. These numbers were generated on a separate page using the RAND() function, then pasted using Paste Values options (thereby preventing numbers from regenerating every time the spreadsheet recalculates). Next to which we placed the row number (using the Excel ROW() function) as shown below as Exhibit 1:

	A	B	C	D	E	F	G	H	I	J
1	0	0	0	0	0	0	0	0	0.366569	1
2	0	0	0	0	0	0	0	1	0.748031	2
3	0	0	0	0	0	0	1	0	0.324232	3
4	0	0	0	0	0	0	1	1	0.979208	4
5	0	0	0	0	0	1	0	0	0.65176	5
6	0	0	0	0	0	1	0	1	0.815595	6
7	0	0	0	0	0	1	1	0	0.046888	7
8	0	0	0	0	0	1	1	1	0.665309	8
9	0	0	0	0	1	0	0	0	0.897561	9
10	0	0	0	0	1	0	0	1	0.27449	10
11	0	0	0	0	1	0	1	0	0.497506	11
12	0	0	0	0	1	0	1	1	0.907134	12
13	0	0	0	0	1	1	0	0	0.117364	13
14	0	0	0	0	1	1	0	1	0.983412	14
15	0	0	0	0	1	1	1	0	0.169216	15
16	0	0	0	0	1	1	1	1	0.419149	16
17	0	0	0	1	0	0	0	0	0.516019	17
18	0	0	0	1	0	0	0	1	0.348522	18

Exhibit 1: Example of fitness values (column I) and bit values (columns A through H).

Because of the way we created the bit permutations, the 8 bits (Columns A-H) were actually the binary equivalent of the row number -1. For example, Row 14 consists of the bits for the binary

value of 13, i.e., 00001101. Identifying fitness values by their row number is useful in performing the next step, the search for local peaks.

To conduct a search for local peaks, individual entities on the fitness space transition to their best neighbor, unless their fitness is higher than all their neighbors, in which case the entity is on a local fitness peak (and therefore remains in position). The state space stabilizes at the point where every entity is on a local fitness peak.

In the N-K model, neighbors of a given entity are defined as the set of positions that are 1 bit different. For example, the position 00001111 has eight neighbors, 10001111, 01001111, 00101111, 00011111, 00000111, 00001011, 00001101, and 00001110. To determine if an entity needed to change positions, we first compute the maximum of the entity's fitness and that of its 8 neighbors. This is done using the formula in Exhibit 2:

f_k	<code>=MAX(INDIRECT(ADDRESS(J1,9)),INDIRECT(ADDRESS(J1+1-2*MOD(J1-1,2),9)), INDIRECT(ADDRESS(J1+2-4*MOD(INT((J1-1)/2),2),9)),INDIRECT(ADDRESS(J1+4-8*MOD(INT((J1-1)/4),2),9)),INDIRECT(ADDRESS(J1+8-16*MOD(INT((J1-1)/8),2),9)),INDIRECT(ADDRESS(J1+16-32*MOD(INT((J1-1)/16),2),9)), INDIRECT(ADDRESS(J1+32-64*MOD(INT((J1-1)/32),2),9)),INDIRECT(ADDRESS(J1+64-128*MOD(INT((J1-1)/64),2),9)),INDIRECT(ADDRESS(J1+128-256*MOD(INT((J1-1)/128),2),9)))</code>								
	G	H	I	J	K	L	M	N	O
	0	0	0.366569	1	0.921703385	65	0.921703	65	0.9217

Exhibit 2: Formula for computing maximum adjacent value

For those not familiar with some of the functions used, the ADDRESS function returns the address of a particular row and column of the spreadsheet. In this case, since I is the 9th letter of the alphabet, column 9 contains the random fitness value. The J column (as mentioned earlier) contains the row number. The INT function truncates a value to its integer equivalent, while the MOD function returns the remainder of its first argument divided by its second argument. Thus, MOD(J1-1,2) would return the rightmost bit of an integer, since the remainder of any integer divided by two is either 0 or 1. To get the next bit, you use the formula MOD(INT((J1-1)/2),2). To get the next bit, you use MOD(INT((J1-1)/4),2) and continue the process by dividing by the remaining powers of 2, i.e., 8, 16, 32, 64 and, finally, 128.

Once you know if a particular bit is on or off, you know whether that particular neighbor is on a higher or lower row. For example, if the rightmost bit is 0, the neighbor is on the next row. If the first bit is 1, the neighbor is on the preceding row. More generally, if we number the bits from 0 to 7, going from right to left, the general formula for finding the row for the neighbor at bit position k is:

$$\text{OriginalRow (J1 in the example)} + 2^k - 2^{k+1} * (0 \text{ or } 1, \text{ depending if the bit is on})$$

Or, more specifically:

$$J1 + 2^k - 2^{k+1} * (\text{MOD}(\text{INT}((J1-1)/2^k),2))$$

with 2^k being replaced by the corresponding values of 2,4,8,...,128. Under this formula, you move forward 2^k when the bit is 0, backwards by 2^k when it is 1.

Illusions of Significance

Once we've determined the neighbor's row, the ADDRESS(row,9) returns the cell address. The INDIRECT function then takes that address and returns its value. The outer MAX function then returns the maximum of these values.

Unfortunately, using the MAX function gives us the value but not where it is located in the fitness space. So, in the next cell, we match the neighbors to the maximum values and then return the row where the match was found. This is done with the highly nested IF statement shown in Exhibit 3.

H	I	J	K	L	M	N	O
0	0.366569	1	0.921703385	65	0.921703	65	0.9217
1	0.748031	2	0.979208426	4	0.979208	4	0.9792
0	0.324232	3	0.979208426	4	0.979208	4	0.9792

Exhibit 3: Formula for determining row source of maximum adjacent value

Each IF clause compares the neighbor value to the computed maximum. If it matches, the computed row value for the neighbor is returned. If not, the value of the nested IF looking at the next set of values is returned.

Once we've got the value, row pair of cells, we can copy them into successive columns until the row values stop changing. (For example, row 2 moved to row 4 in column L, but stayed the same in column N, meaning that 4 is a peak (its value happens to be .979 in the example, as seen in the original diagram). According to Kauffman (1993, p. 49), the number of steps before convergence is reached in a random space is on the order of $\log_2(\text{Number of Neighbors} - 1)$, of 3 in our example. To be safe, we set up 8 separate value-column pairs then tested for convergence (which always had occurred).

Once the converged set of fitness values had been obtained, the number of peaks was counted by copying the converged values from the fitness level and row position columns (all 256 rows) to another worksheet page, then using Excel's Remove Duplicates function to establish the unique peaks.

To perform the regression analysis, we used the bit values (Columns A through H) of the peak cell as independent variables and the converged fitness values column as the dependent variable.

Non-Chaotic Landscapes

There are many possible ways that N,K landscapes with $K < N-1$ could be modeled. Since this paper was not intended for use by evolutionary biologists but rather for purposes of qualitative analysis, we modeled them by separating out values that always contribute to fitness from those that interact—the latter continuing to be simulated with a random number. For example, to simulate a *mixed landscape* with 4 variables that contribute decomposably, as well as interactively, we used:

$$0.125 * (\text{Bit 7} + \text{Bit 6} + \text{Bit 5} + \text{Bit 4}) + 0.5 * \text{Random Number}$$

A *partitioned landscape* has a similar structure, but the randomness is different. Specifically, to simulate a partitioned landscape, we needed to ensure that the interacting portion of the landscape function always contributed the same value each time the same set of variable values was provided. To accomplish this in the spreadsheet, we placed all the interacting variables in the right-most columns (e.g., E through H in Exhibit 1) and then used a formula that caused the same sequence to repeat once we had cycled through all combinations. An example of this formula is provided in Exhibit 4, where 4 variables are decomposable and 4 variables interact.

```
=0.125*(INT((ROW(A1)-1)/128)+MOD(INT((ROW(A1)-1)/64),2)+MOD(INT((ROW(A1)-1)/32),2)+MOD(INT((ROW(A1)-1)/16),2))+0.5*INDEX($A$1:$A$64,MOD(ROW(B1)-1,16)+1)
```

Exhibit 4: 8 characteristic partitioned process with 4 interacting variables

In this example, the range \$A\$1:\$A\$64 contains random numbers. Only 16 of these (2^4) are used, reflecting the 16 combinations of decomposable variables. Rather than having explicit characteristic columns (e.g., A through H in Exhibit 1) the values of the characteristics are computed using the row number.

Example 1: Regression Results for 8,7

To demonstrate the impact of fitness peaks of regression results, the output of 3 different regressions are presented. In the first example, we first show the regression of the 256 raw (unmigrated) random fitness values against the 8 bit values. As shown as Exhibit 5, the results show no significant patterns with a constant value around 0.5 (the mean value of our random number generator) and no significant coefficients—which is what we'd hope for given that the dependent variable values were entirely random.

Illusions of Significance

SUMMARY OUTPUT				
<i>Regression Statistics</i>				
Multiple R	0.185086008			
R Square	0.03425683			
Adjusted R Square	0.002977699			
Standard Error	0.295656129			
Observations	256			
<i>ANOVA</i>				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	8	0.765872115	0.095734014	1.095197638
Residual	247	21.59089895	0.087412546	
Total	255	22.35677107		
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0.525528336	0.055435524	9.479992195	2.2072E-18
X Variable 1	0.012768747	0.036957016	0.34550265	0.730010584
X Variable 2	-0.042239267	0.036957016	-1.142929586	0.254174424
X Variable 3	-0.039908829	0.036957016	-1.079871514	0.281252714
X Variable 4	-0.031348028	0.036957016	-0.848229421	0.39713146
X Variable 5	0.051162533	0.036957016	1.384379424	0.167491715
X Variable 6	-0.017331902	0.036957016	-0.468974599	0.639501454
X Variable 7	-0.029223486	0.036957016	-0.790742579	0.429852803
X Variable 8	0.060598102	0.036957016	1.639691422	0.102342288

Exhibit 5: Regression of bit variable values against 256 random fitness values

In the next regression, Exhibit 6, the same 256 random values are used, but values have migrated to the 26 local fitness peaks, meaning that the 256 observations are drawn from a pool of only 26 fitness values. As a result, patterns that would not be particularly significant in 26 observations appear vastly more significant when assumed to have come from 256 observations. In this case, an adjusted R^2 of 0.70 has been observed and 7 of the 8 bit variable values are significant—6 of them at the $p < 0.001$.

SUMMARY OUTPUT				
<i>Regression Statistics</i>				
Multiple R	0.840829254			
R Square	0.706993835			
Adjusted R Square	0.697503756			
Standard Error	0.032071754			
Observations	256			
<i>ANOVA</i>				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	8	0.61302933	0.076628666	74.49820936
Residual	247	0.254063564	0.001028597	
Total	255	0.867092893		
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0.983561449	0.006459558	152.2645021	3.4296E-246
X Variable 1	0.004559779	0.004314654	1.056812285	0.291629778
X Variable 2	-0.064802382	0.004410095	-14.69410152	1.50407E-35
X Variable 3	-0.046229009	0.004227078	-10.93639757	5.81366E-23
X Variable 4	0.019345259	0.004318042	4.480099494	1.14151E-05
X Variable 5	0.02037802	0.004359393	4.674508058	4.84926E-06
X Variable 6	-0.02907322	0.004341858	-6.696032108	1.43233E-10
X Variable 7	-0.031598475	0.004167566	-7.581997459	6.87611E-13
X Variable 8	0.013012702	0.00410026	3.173628732	0.001696391

Exhibit 6: Regression of NK 8,7 values after they have migrated their peaks

Example 2: 8,7 Landscape with Likert Independent Variables and Random Error

To determine if the large variable significances observed were purely a function of binary independent variables and the limited dependent variables (i.e., only peak values were represented), we took an 8,7 run (Exhibit 6) and gradually transformed it as follows:

1. Input variables were transformed to Likert style variables on a 7 point scale. This was done with the transformation: Binary Value * 3 + Integer Value (4*Random Number between 0 and 1). This allowed for potential overlap at the 3 value, adding error to the input.
2. A random error component was added to the fitness value. Because all fully migrated observations had fitness between 0.7 and 1.00, and 99% had fitness between 0.75 and 1.00, the dependent range was treated as 0.25. The following error distributions were tested:
 - a. 0 random dependent error
 - b. Average 5% random dependent error: computed using $0.025*(0.5-\text{Random Number between } 0 \text{ and } 1)$.

Illusions of Significance

- c. Average 10% random dependent error: computed using $0.05 \times (0.5 - \text{Random Number between 0 and 1})$.
 - d. Average 20% random dependent error: computed using $0.10 \times (0.5 - \text{Random Number between 0 and 1})$.
3. As a further test, the original binary independent variables were regressed against the 20% transformation of the dependent variable

The results of these tests are presented in Exhibit 7.

	8,7 Run 1	Likert Only	Likert + 5%	Likert + 10%	Likert + 20%	Binary + 20%
R ²	0.6975	0.4377	0.4238	0.4191	0.2986	0.5556
Const	0.9836	1.0003	0.9979	1.0021	0.9986	0.9862
Coeff 1	0.0046	0.0008	0.0007	0.0017	0.0003	0.0062
Coeff 2	-0.0648	-0.0135	-0.0134	-0.0138	-0.0132	-0.0678
Coeff 3	-0.0462	-0.0098	-0.0097	-0.0095	-0.0090	-0.0446
Coeff 4	0.0193	0.0050	0.0049	0.0043	0.0063	0.0182
Coeff 5	0.0204	0.0038	0.0041	0.0038	0.0039	0.0210
Coeff 6	-0.0291	-0.0059	-0.0057	-0.0063	-0.0058	-0.0307
Coeff 7	-0.0316	-0.0072	-0.0073	-0.0075	-0.0071	-0.0321
Coeff 8	0.0130	0.0050	0.0052	0.0047	0.0032	0.0099
P Value 1	0.29163	0.57308	0.63607	0.27652	0.88149	0.29642
P Value 2	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
P Value 3	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
P Value 4	0.00001	0.00099	0.00139	0.00621	0.00096	0.00238
P Value 5	0.00000	0.01273	0.00823	0.01885	0.04142	0.00052
P Value 6	0.00000	0.00012	0.00027	0.00009	0.00268	0.00000
P Value 7	0.00000	0.00000	0.00000	0.00000	0.00019	0.00000
P Value 8	0.00170	0.00083	0.00065	0.00266	0.08981	0.07905

Exhibit 7: Tests of random transformations applied to 8,7 results

What these results demonstrate is the robustness of the observed significances in the face of added randomness. Although the transformation of the input variables to Likert scales added considerable error (e.g., roughly 25% of all observations were expected to be 3 for a given variable—which was an ambiguous value) that was reflected in the reduced adjusted R², the pattern of variable significances remained strong. Only at the 20% error case did significances start to drop out (see Variable 8), but this was true whether or not the Likert input scale was used.

What this demonstrates is that the significances produced by the migration process cannot necessarily be detected through duplicate observations. Just incorporating randomness in the transformation of binary variables to a Likert scale eliminated duplicate rows. Adding any randomness to the fitness value then eliminated duplication of dependent variable values. Thus, although the underlying pattern of significances certainly was a result of observations clustering around peaks, such clustering would not be detectable through a casual inspection of the data when error in measurement is present.

Example 3: Regression Results for Mixed Landscape

In the next example, mixed landscape is simulated using the formula presented earlier in the appendix, relabeled to reflect the output variable names, i.e.:

$$0.125*(X \text{ Variable } 1 + X \text{ Variable } 2 + X \text{ Variable } 3 + X \text{ Variable } 4) + 0.5*\text{Random Number}$$

The same set of 256 random values used in Exhibits 5 and 6 were reused for this example. For the first regression we use the 256 raw (unmigrated) fitness values. Based on the model used to produce the data, the underlying process should produce an R^2 value of 0.5 (since half our variance is actually predicted by variables 1 through 4), a constant value of 0.25 (the mean of the random component of fitness), and individual coefficient values of 0.125 for each of the first four values, with insignificant coefficients for the remaining 4—which make no contribution to fitness beyond coincidence. As shown in Exhibit 8, our results largely conform to our expectations. (Although the R^2 is a bit low and coefficient estimates are not exact, all values are within reasonable bounds. Tests on other runs often led to even closer estimates).

SUMMARY OUTPUT				
<i>Regression Statistics</i>				
Multiple R	0.620855146			
R Square	0.385461112			
Adjusted R Square	0.365557018			
Standard Error	0.147828064			
Observations	256			
<i>ANOVA</i>				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	8	3.38564901	0.423206126	19.36592143
Residual	247	5.397724738	0.021853137	
Total	255	8.783373748		
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0.262764168	0.027717762	9.479992195	2.2072E-18
X Variable 1	0.131384373	0.018478508	7.110118049	1.24352E-11
X Variable 2	0.103880366	0.018478508	5.621685813	5.09723E-08
X Variable 3	0.105045586	0.018478508	5.684743885	3.68321E-08
X Variable 4	0.109325986	0.018478508	5.916385977	1.09219E-08
X Variable 5	0.025581266	0.018478508	1.384379424	0.167491715
X Variable 6	-0.00866595	0.018478508	-0.468974599	0.639501454
X Variable 7	-0.01461174	0.018478508	-0.790742579	0.429852803
X Variable 8	0.030299051	0.018478508	1.639691422	0.102342288

Exhibit 8: Regression results for mixed case with 4 decomposable variables on raw (unmigrated) values

The final example, presented in Exhibit 9, shows the same model presented in Exhibit 8 after the observations have migrated to the 19 available peaks (the number of peaks declining, as predicted, with the drop in K value from 7 to 3). A number of interesting comparisons to Exhibit 8

can be made. First, the significances of the fundamental underlying model (for variables 1 through 4) remain intact. Second, the R^2 value is far too high (as a consequence of the declining number of independent observations). Third, we have picked up illusory significances on 3 of the 4 remaining variables—although the coefficient values are lower than those for Variables 1 through 4. Fourth, the constant (intercept) is almost twice what it should be.

What this example should make clear is the degree to which attempting to analyze a multi-peaked evolutionary process using a linear model that assumes decomposability can lead to misleading results that imply significant relationships that don't exist.

SUMMARY OUTPUT				
<i>Regression Statistics</i>				
Multiple R	0.984738246			
R Square	0.969709412			
Adjusted R Square	0.968728341			
Standard Error	0.019128559			
Observations	256			
<i>ANOVA</i>				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	8	2.893312684	0.361664085	988.4185312
Residual	247	0.090377736	0.000365902	
Total	255	2.98369042		
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0.520067032	0.008200864	63.41612487	7.5118E-155
X Variable 1	0.098802841	0.005432513	18.18731801	1.76E-47
X Variable 2	0.086183282	0.004177764	20.62904447	1.19405E-55
X Variable 3	0.08568319	0.003336742	25.67869631	1.10388E-71
X Variable 4	0.132875813	0.003759922	35.34004125	1.38337E-98
X Variable 5	0.047694648	0.002669552	17.86616407	2.16062E-46
X Variable 6	-0.047278336	0.003811452	-12.40428335	8.66883E-28
X Variable 7	-0.028001084	0.003278514	-8.540785135	1.38593E-15
X Variable 8	-0.001145521	0.002991598	-0.38291271	0.702113806

Exhibit 9: Regression of mixed case with 4 decomposable variables after migration to peaks

It needs to be emphasized that the results in Exhibit 9 are entirely a result of the fact that multiple observations are being generated by each peak—confounding the ability of the regression model to accurately estimate significances. This is illustrated in Exhibit 10, where the same data used for Exhibit 9 has been compressed down to 19 observations, i.e., one observation per peak. This eliminates the confounding effect of the number of unique observations being inconsistent with the number of observations used to estimate significances and produces a model that is quite close to the original underlying model, even though our number of observations is well below what is usually recommended for regression analysis.

SUMMARY OUTPUT				
<i>Regression Statistics</i>				
Multiple R	0.972565047			
R Square	0.94588277			
Adjusted R Square	0.902588986			
Standard Error	0.037195336			
Observations	19			
<i>ANOVA</i>				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	8	0.241812488	0.030226561	21.8480041
Residual	10	0.01383493	0.001383493	
Total	18	0.255647418		
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0.472642119	0.03743885	12.62437606	1.8106E-07
X Variable 1	0.113456344	0.025615707	4.429170855	0.001275777
X Variable 2	0.113808045	0.022785668	4.994720597	0.000541532
X Variable 3	0.063809216	0.021968828	2.904534352	0.015711105
X Variable 4	0.143099869	0.018870604	7.583215974	1.8755E-05
X Variable 5	0.031691214	0.021358968	1.483742748	0.168696728
X Variable 6	-0.01300163	0.019775827	-0.657450647	0.525734086
X Variable 7	-0.011148216	0.021252443	-0.524561599	0.611317867
X Variable 8	-0.002508653	0.020082333	-0.124918405	0.903063672

Exhibit 10: Regression of converged mixed case with 19 peaks with all duplicates removed

Example 4: Regression Results for Partitioned Landscape

In the next example, mixed landscape is simulated using the formula presented earlier in the appendix, relabeled to reflect the output variable names, i.e.:

$$0.125*(X \text{ Variable } 1 + X \text{ Variable } 2 + X \text{ Variable } 3 + X \text{ Variable } 4) + 0.5 * \text{Interaction Random Number}$$

Based on the model used to produce the data, the underlying process should produce an R-squared value of 0.5 (since half our variance is actually predicted by variables 1 through 4), a constant value of 0.25 (the mean of the random component of fitness), and individual coefficient values of 0.125 for each of the first four values, with insignificant coefficients for the remaining 4—which make no contribution to fitness beyond coincidence. As shown in Exhibit 11, our results conform to our expectations for the decomposable variables but are wildly different for the interacting variables, with high significances being detected in 3 of the 4 cases.

SUMMARY OUTPUT				
<i>Regression Statistics</i>				
Multiple R	0.813603489			
R Square	0.661950637			
Adjusted R Square	0.65100167			
Standard Error	0.098471462			
Observations	256			
<i>ANOVA</i>				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	8	4.689895985	0.586236998	60.45781521
Residual	247	2.395067338	0.009696629	
Total	255	7.084963323		
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0.269635073	0.018463399	14.60376119	3.06354E-35
X Variable 1	0.125	0.012308933	10.15522646	1.78837E-20
X Variable 2	0.125	0.012308933	10.15522646	1.78837E-20
X Variable 3	0.125	0.012308933	10.15522646	1.78837E-20
X Variable 4	0.125	0.012308933	10.15522646	1.78837E-20
X Variable 5	0.08120404	0.012308933	6.597163292	2.53135E-10
X Variable 6	0.012692333	0.012308933	1.031148105	0.303479987
X Variable 7	-0.0463868	0.012308933	-3.76854758	0.000205417
X Variable 8	-0.04327468	0.012308933	-3.51571374	0.000521651

Exhibit 11: Regression of 8 variable partitioned process with 4 interacting variables

Example 5: Continuous Convergence in 8,0 Landscape

In Example 4 we found that illusions of significance will appear for interacting variables without migration to fitness in partitioned processes (unlike the chaotic and mixed examples). For this reason, in considering the impact of migration to fitness for partitioned processes, what is of particular importance is what happens to the decomposable portion of the fitness function. To test this, we constructed a model of the form:

$$Y = 0.99 * (V1 + V2 + V3 + V4 + V5 + V6 + V7 + V8) + 0.01 * \text{Random Number}$$

The random number was included to prevent all cells from migrating on a particular path (though we later confirmed, in a separate test, that it has no effect; the same phenomenon is observed without it).

The results, presented in Exhibit 12, demonstrate a good estimate for each step (Const should be .005, the mean value of the error term, and each variable coefficient should be 0.12375, which is $0.99 * 0.125$), although some drop off in accuracy of the constant and coefficient estimates is noted as the number of independent observations shrinks. This is to be expected, since each step essentially eliminates all the lowest observations from the sample, causing the slight randomness in the sample to exert greater influence on the estimates (e.g., consider how errors in closely spaced

points located far away from the Y-intercept of a line cause the line to pivot more dramatically than evenly spaced points with some observations close to the Y intercept). In each case, all p values were 0 (meaning their estimated value was below the accuracy threshold used by Excel). Only 4 steps are shown because, by step 5, low fitness 0 values for some bits had been entirely eliminated from the sample as a consequence of convergence.

	Start	1	2	3	4	Final
# Independent	256	106	79	35	19	1
R ²	0.999716	0.999927	0.99991	0.99991548	0.999938	N/A
Const	0.005255	0.01137	0.0144	0.01857462	0.022771	N/A
Coeff 1	0.123878	0.123662	0.12363	0.12358929	0.124457	N/A
Coeff 2	0.123328	0.122631	0.12225	0.12243324	0.12112	N/A
Coeff 3	0.123351	0.122682	0.12247	0.12157904	0.121471	N/A
Coeff 4	0.123437	0.123292	0.12264	0.12226362	0.121248	N/A
Coeff 5	0.124262	0.123399	0.12246	0.12099419	0.120737	N/A
Coeff 6	0.123577	0.122718	0.12204	0.1209377	0.119483	N/A
Coeff 7	0.123458	0.12271	0.12223	0.12196444	0.121796	N/A
Coeff 8	0.124356	0.123083	0.12206	0.12100948	0.119802	N/A

Exhibit 12: Convergence of a decomposable 8,0 landscape

One aspect of the migration that is of concern, even in the 8,0 case, is the vast overestimate of coefficient accuracy that takes place as the peak is approached. In Exhibit 13, for example, we see the full coefficient results for Step 4 of Exhibit 12. What is clear is that every single coefficient is slightly misestimated but not a single 95% range estimate includes the correct 0.12375 value—and only one (Coeff1) comes even close. The conclusion we may draw from this is that even where an underlying process is decomposable, the process of migration towards peak fitness may lead to excessive confidence in estimates.

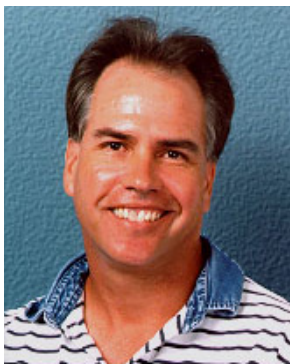
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Const	0.022771	0.000539	42.27015	4.5E-115	0.021709572	0.023832
Coeff 1	0.124457	0.000295	421.5741	0	0.123875812	0.125039
Coeff 2	0.12112	0.000193	626.5277	0	0.120738818	0.1215
Coeff 3	0.121471	0.000249	487.497	0	0.120980429	0.121962
Coeff 4	0.121248	0.000277	438.0766	0	0.120702393	0.121793
Coeff 5	0.120737	0.000204	591.4528	0	0.120334694	0.121139
Coeff 6	0.119483	0.000337	354.38	0	0.118818715	0.120147
Coeff 7	0.121796	0.000201	606.8647	0	0.121401186	0.122192
Coeff 8	0.119802	0.000173	691.3026	0	0.119460611	0.120143

Exhibit 13: Coefficient estimates of 8,0 landscape near convergence

Biographies



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*.



Terry L. Sincich is an Associate Professor in the Information Systems and Decision Sciences Department in the College of Business Administration at the University of South Florida. Dr. Sincich's areas of research interest include applied statistical modeling and analyses. He has co-authored research papers in top-rated journals in a variety of disciplines, including the *Journal of the American Statistical Association*, *Academy of Management Journal*, *Auditing: A Journal of Theory & Practice*, *Research on Accounting Ethics*, *Demography*, and *International Journal of Forecasting*. Dr. Sincich received a Ph.D. in Statistics from the University of Florida. He has held faculty positions at the University of Florida and the University of South Florida, where he has won numerous teaching awards in both graduate and undergraduate statistics classes. Dr. Sincich is a member of the American Statistical Association and the Decision Sciences Institute.