Bios Data Analyzer

H. Sabelli\textsuperscript{1}, Chicago Center for Creative Development
A. Sugerman, Chicago Center for Creative Development
L. Kovacevic, Chicago Center for Creative Development
L. Kauffman, University of Illinois at Chicago
L. Carlson-Sabelli, Rush University
M. Patel, University of Illinois at Chicago
J. Konecki, Chicago Center for Creative Development

Abstract: The Bios Data Analyzer (BDA) is a set of computer programs (CD-ROM, in Sabelli et al., Bios: A Study of Creation, 2005) for new time series analyses that detects and measures creative phenomena, namely diversification, novelty, complexes, nonrandom complexity. We define a process as creative when its time series displays these properties. They are found in heartbeat interval series, the exemplar of bios just as turbulence is the exemplar of chaos, in many other empirical series (galactic distributions, meteorological, economic and physiological series), in biotic series generated mathematically by the bipolar feedback, and in stochastic noise, but not in chaotic attractors. Differencing, consecutive recurrence and partial autocorrelation indicate nonrandom causation, thereby distinguishing chaos and bios from random and random walk. Embedding plots distinguish causal creative processes (e.g. bios) that include both simple and complex components of variation from stochastic processes (e.g. Brownian noise) that include only complex components, and from chaotic processes that decay from order to randomness as the number of dimensions is increased. Varying bin and dimensionality show that entropy measures symmetry and variety, and that complexity is associated with asymmetry. Trigonometric transformations measure coexisting opposites in time series and demonstrate bipolar, partial, and uncorrelated opposites in empirical processes and bios, supporting the hypothesis that bios is generated by bipolar feedback, a concept which is at variance with standard concepts of polar and complementary opposites.

Key Words: bios, time series analysis, creativity, economics, heart rate variation

\textsuperscript{1} Correspondence address: Hector Sabelli, 2400 N. Lakeview #2802 Chicago, IL 60614, USA. E-mail: Hector_Sabelli@rush.edu
INTRODUCTION

Time series analysis (Hamilton, 1994; Kantz and Schreiber, 1997; Kendall, 1973; Sprott & Rowlands, 1995; Webber & Zbilut, 1996) becomes a focus of scientific methodology when we recognize that the universe is in continual evolution (according to notable authors such as Lao-tzu, Heraclitus, Bruno, Hegel, Lamarck, Darwin, Engels, Whitehead, Chardin, Hubble, Thom, and Prigogine). Simple processes generate new, diverse and complex processes. To develop a science of creative processes, we construct methods to identify and measure creative phenomena. In this manner, we may also develop an operational definition of creative processes. Through the study of empirical time series of empirical processes, we have advanced a definition of creative phenomena (Sabelli, Carlson-Sabelli, Patel, & Sugerman, 1997; Sabelli & Abouzeid, 2003) as characterized by the coexistence of simple and complex components of variation (pointing to the generation of complexity by simpler processes), episodic pattern (instead of uniformity), diversification (increase in variance rather than convergence to an attractor), novelty (innovation greater than in random series with the same variance), nonrandom complexity (as contrasted to algorithmic complexity that is maximal in random series), and irreversibility. Creative processes are by necessity aperiodic and nonstationary. Notwithstanding, following mechanical and statistical worldviews, most analytic methods focus on stable patterns (static, periodic or chaotic) and require stationary data. Aperiodic patterns include random distributions, stochastic processes such as random walks, chaos (Glass & Mackey, 1988) and bios (Kauffman & Sabelli, 1998).

Bios is exemplary of creative processes. Bios is a pattern found in heartbeat interval series (Fig. 1), which are exemplary of complex, creative and vital processes. Their pattern continually varies because they reflect ever-changing central nervous activity; thus, sequences of heartbeat intervals are always unique. Heart rate variation is a significant indicator of cardiac health (Malik & Camm, 1995). As turbulence is considered paradigmatic of chaos, we regard heartbeat variation as paradigmatic of bios (Sabelli & Kauffman, 1999; Sabelli, 2005). The definition of bios is discussed from a mathematical perspective elsewhere (Kauffman & Sabelli, 2003). The empirical characterization of bios and its differences with chaos will become evident through the analyses described in this article. Time graphs (Fig. 1) show continuity between consecutive terms of biotic series, and extreme changes, often equal to the range of the entire series, in the case of chaos.
Biotic patterns similar to those found in the heart can also be generated by recursions of trigonometric functions (Kauffman & Sabelli, 1998; Sabelli, 1999), such as

\[ A_{t+1} = A_t + k \times t \times \sin(A_t), \]
\[ A_{t+1} = A_t + k \times t \times \sin(A_t \times k \times t). \]

These recursions represent bipolar feedback, i.e. a feedback that is sometimes positive and sometimes negative; the feedback gain \( g \) is given by the product \( k \times t \). They generate a series which progresses from convergence to \( \pi \) to a cascade of bifurcations generating periods \( 2, 4, 8 \ldots 2n \), chaos, and bios, as \( t \) increases. The transition from chaos to bios occurs when the gain \( g \) is equal or larger than 4.6035... Figure 1 illustrates the difference between chaotic and biotic series generated by these equations. Trigonometric recursions without a conserved term, such as

\[ A_{t+1} = k \times t \times \sin A_t \]

generate chaos but not bios (Sabelli, 2001a). This is significant regarding what processes are necessary for creative features to emerge in nature. We use the term “biotic equations” for those recursions that generate bios at some values of their parameters. “Process chaos” is the chaotic series generated by biotic equations.

Fig. 1. Biotic and chaotic series. Left: Time graph of heartbeat intervals computed by measuring the interval between \( R \) in the electrocardiogram (RRI). Middle: Biotic series generated with the diversifying equation \( A_{t+1} = A_t + k \times t \times \sin(A_t \times k \times t) \). Right: Chaotic series generated with the same equation.

Bios analysis focuses on its three defining characteristics: (a) Causation: chaos and bios differ from random and random walk in that they are generated causally. (b) Bipolarity: the BDA focuses on measures of opposition because bios is generated mathematically by recursions of bipolar feedback, and bipolar feedback also appears to be crucial in the generation of biotic patterns in heartbeat series (sympathetic vs.
parasympathetic nerves) and in economic series (supply vs. demand). (c) Creativity: biotic series and random walks differ from chaotic attractors and random distributions in displaying features to be expected from creative processes: nonstationarity, diversification, novelty, nonrandom complexity, asymmetry, and irreversibility (Table 1).

**Table 1.** Bios Analysis Differentiates Four Elementary Types of Aperiodic Series.

<table>
<thead>
<tr>
<th>Empirical measures</th>
<th>Stochastic processes</th>
<th>Causal processes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Causation:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonrandom series of differences</td>
<td>Random, independent</td>
<td>Causal, connected actions</td>
</tr>
<tr>
<td>Partial autocorrelation</td>
<td>events</td>
<td></td>
</tr>
<tr>
<td>Consecutive isometry at low embeddings</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stability:</strong></td>
<td>RANDOM</td>
<td>CHAOS</td>
</tr>
<tr>
<td>Stationary variance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recurrence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random complexity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uniform patterns</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Creativity:</strong></td>
<td>RANDOM WALK</td>
<td>BIOS</td>
</tr>
<tr>
<td>Diversification (increasing variance)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novelty (less recurrent than random)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonrandom complexity (arrangement)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Episodic patterns (complexes)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These properties can be measured in time series (Carlson-Sabelli & Sabelli, 1992; Carlson-Sabelli, Sabelli, Patel, & Holm 1992; Carlson-Sabelli et al., 1995, 1997; Patel & Sabelli, 2003, 2004; Sabelli, 1999, 2000a, 2001b, 2002; Sabelli et al., 1994; 1995a; 1997; Sabelli, Carlson-Sabelli, Patel, Levy, & Diez-Martin, 1995b; 1997) provided they are not removed by differencing or detrending, as is often done to render the data stationary. These methods have been developed through the study of biological (Carlson-Sabelli et al., 1994, 1995, 1997; Sabelli et al., 1994, 1995a, 1995b, 1997; Sabelli & Carlson-Sabelli, 1999), economic (Sabelli & Kauffman, 1999; Sugerman et al., 1999; Levy-Carciente, Sabelli & Jaffe, 2004), meteorological (Sabelli, 2000), cosmological (Sabelli & Kovacevic, 2004), and other time series (Sabelli, 2005), as well as series generated by mathematical recursions. These are the data considered in this study, which will be referred to as “empirical series” for the sake of brevity. These analytic methods are embodied in a series of programs, the Bios Data Analyzer (BDA) (Sugerman et al.; CD-ROM enclosed in Sabelli, 2005).
This article presents an overview of the BDA. Its main tools are shuffling (to eliminate temporal order while leaving the statistical distribution unchanged), differencing (that reduces bios to chaos), integrating (that turns chaos into bios and random into random walks), embedding (that enhances many patterns), and trigonometric transformations (that reveal opposites within a single time series). It follows that the BDA presupposes the use of original data. Data transformation prior to analysis (e.g. detrending, converting integer into non-integer data, as when heartbeat intervals are presented as “instantaneous heart rate”) is likely to distort the existing patterns, and as a rule prevents the detection of bios.

**ISOMETRY AND SIMILARITY RECURRENCE PLOTS**

The quantification of isometries is the core of the analysis of creative processes, allowing one to measure novelty, nonrandom complexity and temporal changes in pattern the defining features of creativity, and to demonstrate causation, thereby distinguishing biotic from stochastic processes. Recurrence methods reveal temporal patterns by comparing sequences of successive members of a time series (embeddings), and finding, plotting (Eckmann, Kaphorst, & Ruelle, 1987) and quantifying (Webber & Zbilut, 1994, 1998; Zbilut & Webber, 1992, 1998; Zbilut et al., 1998) recurrences among these vectors. All recurrence quantification programs sample the time series; the BDA allows the user to choose how many pairs of vectors will be compared.

There are several ways in which recurrence can be defined and measured. We have introduced the terms isometry and similarity to identify two types of recurrence measurements (Sabelli, 2001b; Sabelli & Abouzeid, 2003) that have been used in existing recurrence programs but not explicitly distinguished. Isometry refers to the recurrence of vectors of the same length. Vector length is measured by calculating its Euclidean norm (the square root of the sum of squares of its members). These norms are then compared to one another. Two vectors are considered isometric if the absolute value of the difference between their Euclidean norms is smaller than some cutoff radius chosen by the user. When two vectors are isometric (i.e., they are equal within the tolerance determined by the chosen cutoff radius), a recurrence is counted and plotted. Similarity recurrence requires that two vectors be alike in length and direction: two vectors are recurrent when the Euclidean norm of the distance between the vectors (the square root of the sum of squares of the differences between their members) falls below the selected cutoff radius (0.1 to 10 % in the studies discussed here). This is the method used to
quantify recurrence adopted by other recurrence programs. The BDA quantifies these two types of recurrences (isometry and similarity), plots them as a function of time (recurrence time graphs, Fig. 2) and dimensionality (embedding plots, Figs. 4, 5, and 6).

**HEARTBEAT INTERVALS N=1000**

**SHUFFLED**

**ISOMETRIC**

**SIMILARITY**

**Fig. 2.** Isometric and similarity recurrence plots of heartbeat intervals and their shuffled copies. Complexes are alike in both types of plots. Shuffling the data of RRI series increases the number of isometry recurrences, while it decreases the number of similarity recurrences.

A second distinctive feature of the BDA is allowing the use of very long time series in order to study how patterns change during the evolution of a process. Although the recurrence technique allows one to study transient changes and non-stationary patterns, it is often restricted to the analysis of stationary epochs. For this reason, relatively short
samples are chosen. The process perspective suggests analyzing sufficiently long samples instead of stationary periods and interpreting recurrence plots as time graphs that portray the transformation of one episodic pattern (complex) into another (Fig. 2). The time graphs of similarity and isometry recurrences are alike, indicating that both describe the same process. Complexes are observed in human heartbeat intervals, economic series, meteorological series, DNA sequences, some literary texts, as well as in biotic series generated by recursions of bipolar feedback, and in stochastic noise. All these processes are assumed to be creative processes. In contrast, uniform plots are found in random and chaotic series.

A third distinctive feature of the BDA is the quantification of recurrences at a wide range of embeddings (embedding plot, Figs. 4, 5, and 6) in order to consider both the complex components created by the process and the simpler components that generate them (causation). In contrast, recurrence analyses are often performed at a fixed, relatively low embedding dimension (e.g., 3 to 10), in fear that at higher dimensions noise might inflate and swamp out the dynamics under study. The choice of a low embedding dimension is justified by the assumption that dissipative processes converge to a lower dimensional attractor. The number of recurrences increases with embedding in many random and chaotic series but it actually the decreases with embedding for most empirical series and for bios.

**NOVELTY AND ISOMETRY HISTOGRAMS**

The BDA calculates and compares recurrences in the original series and in shuffled copies, a sine qua non step to measure novelty and nonrandom complexity. Similarity recurrences decrease with shuffling in all the time series we have examined. Many natural processes regarded as creative, as well as mathematical bios, have less isometry than their shuffled copies (except at very low embeddings). This property, exhibiting less recurrence than random, defines novelty, an essential characteristic of creative processes (bios and stochastic noise). Periodic processes show more isometry than their shuffled copy when the embedding coincides with the period. Thus, creativity and order are opposite departures from randomness. Chaotic series also display more isometry than their shuffled copies resembling periodic rather than biotic pattern. Novelty and diversification (see later) may be regarded as a most fundamental feature of biological processes; for instance, sexual reproduction, meiosis, and crossover are mechanisms that increase genetic variation beyond that expected from random mutation.
Fig. 3. Novelty, flux and order in empirical and mathematical series. Top: The calculation of these three values is illustrated for daily average atmospheric temperatures. Bottom: Comparison between series sorted by novelty.
Deterministic causality, stochastic change and nonrandom creation most likely coexist in many physical, biological and economic processes. We have developed a simple method to estimate their relative contribution (Sugerman & Sabelli, 2003). We compute the Euclidean norm for 2, 3, 4, 6, ...23...50 embeddings. Isometries are measured in the histogram of Euclidean norms. Each series $A_i$ is compared with 5 shuffled copies $A_s$, and all results refer to these comparisons. A Pareto distribution of frequencies allows one to measure the change in isometry between the original data and the average of its shuffled copies $A_s$. We compare the frequencies $F$ of $A_i$ and $A_s$ by bin in an effort to distinguish ordered, creative and random components of the process under study (Fig. 3). $F(A_i) > F(A_s)$ defines order; in fact, we find that $F(A_i) > F(A_s)$ in periodic series. $F(A_i) = F(A_s)$ implies random flux since shuffling does not change the frequency. A chaotic series is the same as a random flux by this measure. $F(A_i) < F(A_s)$ defines novelty. We use the specific frequency of $F(A_i)$ as the measure of order, novelty, or flux for each bin, and their sums for the series as a whole. This rough method complements the results obtained with recurrence analysis. It allows one to evaluate to what extent order and novelty coexist in a given series. It also permits examining the entire series rather than sampling it as in recurrence analysis.

CONSECUTIVE RECURRENCE AS A MEASURE OF CAUSATION

Consecutive recurrence is the number of recurrences that follow one other; e.g., if vector $y_t$ is recurrent with vector $y_{t+m}$, then vector $y_{t+1}$ is recurrent with vector $y_{t+m+1}$. (Webber and Zbilut label consecutive recurrence as “determinism”, in keeping with the usage of this term by some mathematicians; physicists and philosophers use the term “determinism” to refer to a theory. High consecutive recurrence also occurs in pi digits, and they denote deterministic causation only when higher than shuffled. Also, higher than shuffled consecutive recurrence occurs in processes that are creative rather than mechanically deterministic). Consecutive recurrence beyond that observed in shuffled data indicates nonrandom order, which may be deterministic or creative. Consecutive recurrence may be observed (a) at low embedding dimensions, demonstrating causation (thus, we observe it in many empirical series and in mathematical chaos and bios, but not in stochastic processes); (b) periodically, peaking when the embedding dimension corresponds to the period; and (c) at high embeddings, indicating integration, i.e., the accumulation of actions, as observed in many empirical series, random walks and in mathematical bios but not in chaotic series or in 1/f noise (Figs. 4
and 5). Natural and mathematical biotic series thus show consecutive recurrence at both low and high embeddings, with a minimum at intermediate embeddings. Measures of consecutive recurrence thereby distinguish between causal and stochastic creative processes. Consecutive recurrence unambiguously distinguishes heartbeat intervals and many other empirical series from 1/f noise. Distinguishing non-random series, such as deterministic chaos from colored noise, is one important problem in time series analysis for which several algorithms have been proposed (Chang, Schiff, & Sauer, 1994; Kaplan & Glass, 1992; Schreiber & Schmitz, 1997; Sigeti & Horsthemke, 1987; Sugihara, Grenfell, & May, 1990). In our studies, we find partial autocorrelation useful but less powerful than the quantification of consecutive recurrence, a point of practical significance regarding economic data.

Summarizing, consecutive recurrence portrays causation including periodicity (consecutive recurrence at periodic intervals), causality (low dimension consecutive recurrence), and integration (high embedding dimensions). Chaos is causal, Brownian noise is integrative, and cardiac and mathematical bios are both. The quantification of isometric vectors demonstrates creativity (novelty + causality) and differentiates it from order (causation without novelty) and random innovation (novelty without causation).

NONRANDOM COMPLEXITY: ARRANGEMENT

There are multiple definitions of complexity; many of them are counterintuitive and non-empirical. Confusion stems from use of the term “complexity” to mean both organization and randomness (Crutchfield, 2003). Gell-Mann (1994) questions the validity of measures of complexity, such as algorithmic complexity, which assigns highest complexity to random processes. Creative processes generate complex organization of a different kind—typewriting monkeys hardly ever write Shakespeare. Theoretically, nonrandom complexity may be expected to increase with novelty and with causation. We thus measure the ratio of consecutive isometries over the number of all isometries, which we call arrangement (Sabelli, 2002) as a possible measure of nonrandom complexity. We have found that arrangement is high in time series of physiological recordings (electroencephalogram, electromyogram, respiration), economic processes (Dow-Jones Industrial Average; prices for crude oil, corn, gold, silver; exchange rates for currencies [Sabelli, 2002]), DNA sequences, complex literary texts, and in biotic series generated by recursions of bipolar feedback (Sabelli, 2005). Arrangement is low in random, periodic and chaotic series. Thus,
arrangement correlates with an intuitive notion of nonrandom complexity. This conclusion is validated by the fact that arrangement increases with the number of different periodicities combined in a multiperiodic series, which is the most straightforward definition of complexity for an engineer. It is also validated clinically: arrangement is lower in heartbeat interval series recorded from cardiovascular and psychiatric patients than in those obtained from healthy persons. Disease, being an unquestionable instance of decay, is simpler than health.

By way of comparison, the Lempel-Ziv complexity also differentiates complex natural processes and biotic series from chaos and $1/f$ noise. But the Lempel-Ziv complexity is very low in biotic series (e.g., 0.09 for $g = 4.65$), much higher in chaos (process chaos, 0.69 for $g = 4.2$; Ikeda 0.64; logistic 1.05), $1/f$ noise (0.69) and random data (1.04). These numerical values do not correspond with our intuition of complexity.

In summary, the quantification of isometric vectors demonstrates creativity (novelty + causality) and differentiates it from order (causation without novelty) and from random innovation (novelty without causation). Lower isometry recurrence than random defines novelty, the central characteristic of creative processes (bios and stochastic noise), in contrast to order as defined by greater recurrence (periodic and chaotic series). Consecutive recurrence portrays causation including periodicity (consecutive recurrence at periodic intervals), causality (low dimension consecutive recurrence), and integration (high embedding dimensions). Chaos is causal, Brownian noise is integrative, and cardiac and mathematical bios are both.

SIMPLICITY AND COMPLEXITY: EMBEDDING PLOTS

An evolutionary perspective leads one to focus on the creation of complexity rather than on complexity itself. In creative evolution, simple low dimensional processes generate complex high dimensional patterns. The time series of processes that evolve from simple to complex may be expected to contain both high dimensional complex components and the simpler low dimensional processes that generate them. In contrast, deterministic processes conserve their original form and dimensionality, so mechanical as well as chaotic processes have only simple low dimensional components. Random processes are highly complex in the sense that each change represents an independent event; thus, stochastic series in which complex organization is generated by random change display only complex components. To study a creative system, it is necessary to measure both its simple and complex components, rather than only complexity.
To study both simple and complex components of a process from a single time series, we plot isometries as a function of the duration of the vector (embedding plots). Periodic series show 100% isometry when the vector is an integer multiple of the period, indicating that the embedding dimension represents the time dimension of the component being examined. It is on this basis that we regard low dimensional consecutive recurrence as a marker of simple linear causation, as contrasted to high dimensional consecutive recurrence. Embedding plots differentiate five types of aperiodic series (Figs. 6 to 8): random (indistinguishable from its shuffled copies), chaotic (low dimensional consecutive recurrence and high dimensional randomness), stochastic (low dimensional randomness and high dimensional novelty and consecutive recurrence), biotic (high dimensional novelty and high dimensional consecutive recurrence at all embeddings), and prebiotic (novelty without consecutive recurrence). In this manner, bios is unambiguously distinguished from chaos (Fig. 4) and Brownian noise (Figs. 6 and 7). Biotic patterns are shown to widespread in nature, including a wide variety of natural processes (Fig. 6).

Pattern does not consist of a single form with a unique dimension that requires a particular embedding to be detected. Rather, a pattern is a complex form that has different shapes at different durations, and therefore requires a wide range of embeddings in order to be portrayed. Embedding plots allow demonstration of different attractors coexisting in one series (e.g., period 2 and chaos coexist in series generated by the logistic or the process equations, as in Fig. 4). Rössler chaos also shows periodicity (Fig. 5), and likewise biotic and periodic patterns can coexist in empirical time series. To detect periodicity, it is necessary to use the number of embeddings that corresponds to the period. A large number of embeddings is necessary to detect seasonal periodicity. This observation is relevant to often-asked questions regarding the choice of embedding for the quantification of recurrences. There is a legitimate concern that a high number of embeddings may introduce errors. Yet, the percentage of isometry remains very low for heartbeat series, intergalactic distances (Fig. 6), DNA sequences (Fig. 6), stochastic noise (Fig. 5) and mathematical bios (Fig. 4) at very large embeddings. The same patterns of distribution of isometries appear in recurrence plots of time series constructed at low and high embedding dimensions; increasing embedding only sharpens the complexes.
Fig. 4. Embedding plot of time series \( N = 3500 \) generated by the process equation \( A_{t+1} = A_t + g \cdot \sin(A_t) \) at different gains \( g \). Recurrence isometry; cutoff radius 0.1. Observe the qualitative difference between chaos (high isometry, same consecutive isometry and arrangement as random at high embeddings) and bios (low isometry [novelty], high consecutive isometry and high arrangement at low and high embeddings). Comparison of the top three rows shows the coexistence of period 2 and chaos in the time series generated with \( g = 3.6 \).
Fig. 5. Embedding plots of time series generated by chaotic and stochastic series. $N = 3500$, radius 0.1. Brownian noise does not show low dimensional consecutive recurrence. Chaotic attractors such as the logistic show high isometry rather than novelty, and do not show consecutive recurrence at high embeddings. The Rossler attractor generates atypical results because it combines chaos with periodicity. Pink noise ($N = 1000$) shows novelty and arrangement, but no consecutive isometries.
Fig. 6. (continued on next page)
**Fig. 6.** Embedding plots of empirical series. RRI: heartbeat intervals (from \( R \) to \( R \) wave) from a healthy person (CCCD-Rush Psychocardiology Data Base). Data for galactic distances from the Las Campanas Redshift Survey (Shectman et al., 1996) and the 2-degree Field Galaxy Redshift Survey (Colless, 1999).

**DIVERSIFICATION**

Regarding processes as evolutionary rather than stationary, the BDA measures changes in statistical parameters in time series rather than the parameters of statistical distributions construed as static. In healthy persons, the S.D. of heartbeat interval series increases with the duration of the recording (Dalton et al., 1977). Based on this observation, we have introduced two statistical measures of diversification (Sabelli & Abouzeid, 2003) that are illustrated in Fig. 7. Global diversification measures the increase in variation with the duration of the sample. The S.D. or the A.D. is measured for data points 1 to 100, then 1 to 200, 1 to 300, etc., up to \( N \) data points, where \( N \) is the entire time series. The S.D. (or the A.D.) increases with duration in many (but not all) samples of many empirical series and in biotic series generated by the process equation. This indicates global diversification. In contrast, the S.D. decreases or remains flat for chaotic series and uniform random data. Random and chaotic processes change continually, but the more they change, the more they remain the same.

A time series can show global diversification because it evolves from one pattern to another, while each of these patterns is stable. More-
over, such changes may actually result in a decrease in variance. For this reason, we also measure local diversification by calculating the S.D. (or A.D.) for sets (“embeddings”) of 2, 3, ..., 200 consecutive terms of the time series, starting with each term in the series. The values obtained for each embedding are averaged for the entire series, and these averages are plotted as a function of the number of embeddings (Figs. 10 and 11). Deviation increases with the number of embeddings for many empirical series and for mathematical bios, while the correlation between deviations and embedding is zero or negative for chaos and uniform random data. Thus, both measures of diversification differentiate creative processes from conservative processes (equilibrium, periodic or random) that maintain their initial degree of diversity, and from processes that converge to equilibrium, periodic or chaotic attractors. It is surprising how such different measurements as global and local diversification generate the same results in most (but not all) cases.

Local diversification must be distinguished from relative dispersion (West & West, 2000), which computes changes in the coefficient of variation with embedding for a low number of embeddings (2 to 10). Relative dispersion portrays divergence as measured by the Lyapunov exponent. The measurement of local diversification measures the S.D. (or the A.D.) at high number of embeddings (10 to 200). Local diversification does not correlate with divergence. The largest positive Lyapunov exponent is more positive in chaotic than in biotic series generated by the process equation or in empirical data.

(Fig. 7 continued on next page)
Fig. 7. From left to right: Global diversification. Local diversification. Global and local diversification and diffusion in time series of the proportion of guanine in sequences of 100 consecutive bases (top), biotic (middle), and chaotic (bottom) series generated by the process equation $A_{r+1}=A_r + g\times\sin(A_r)$. Percentage of guanine in the yeast (Sacharomycetes cerevisiae) chromosome 1 (5000 bases (10000-15000) as in Fig. 6. In each of the series of graphs, the mean squared displacement (M.S.D.), the standard deviation (S.D.), and their ratio, are plotted as a function of the number of terms in the series (top row) and of embedding (bottom row).
DIFFUSION AND THE DIVERSIFICATION/DIFFUSION RATIO

Many creative processes (biological populations, languages, use of inventions) expand, a process that has been compared to diffusion. Diversification may result from diffusion. Diversification is most significant when it does result from nonstationarity of the mean. Random walks, for instance, have nonstationary means and therefore their S.D. increases without bound as N increases; this is described as "infinite variance". Heartbeat intervals, in contrast, show diversification in spite of a stationary mean (e.g., average heart rate is the same at waking up time). Bios is related to the phenomenon of deterministic diffusion described by Arnold and others (Chirikov et al., 1985; Geisel & Nierwetberg, 1982). Certainly the biotic series generated by the equations described above have a nonstationary mean, but biotic series with a stationary mean ("homeobios" can be constructed (Sabelli, 2005). In this, they resemble heartbeat series that are biotic and homeostatic, showing novelty and diversification but also a stationary mean for appropriate intervals of time.

Statistically, diffusion is measured by the mean squared displacement (M.S.D.), which is the average of the square of the deviations of each term in the series from the origin (Fig. 7). As in the case of diversification, we measure the M.S.D. globally, by epochs, as a function of the duration of the sample, and by embedding. Diffusion is commonly observed in empirical series.

Time series of heartbeats, bios, and random walks often (but not always) show both diversification and diffusion. Diversification without diffusion is observed in some heartbeat series and in mathematically generated homeobiotic series (bounded bios). For instance, the biotic time series generated by

$$A_{t+1} = A_t + \sin(5^* A_t) + \sin(5^* A_t) - 0.005^* (A_t - A_1)$$

do not diffuse globally or locally, and diversify only locally. This indicates that the generation of complexity is not reducible to diffusion. In these cases, the S.D./M.S.D. ratio increases with embedding and decreases with duration of the sample. In stochastic series, the S.D./M.S.D. ratio increases with embedding, and it varies in many different ways with duration of the sample. An increase in the S.D./M.S.D. indicates complexification greater than expected from diffusion, while a decrease in this ratio indicates simplification.

ASYMMETRY AND SYMMETRIZATION

Asymmetry is a fundamental property of natural processes (Pasteur [see Haldane, 1960; Sabelli, 1989]). Measures of Skewness
show that empirical data, biotic series, random walk and some chaotic series are highly asymmetric, while periodic data and most random and chaotic series are not. Moreover, in evolving mathematical recursions, there is a net increase in asymmetry from periodicity to chaos and from chaos to bios.

Differentiation, necessary for creativity, may be expected to increase asymmetry, while a tendency to equilibrium represents symmetrization. Symmetrization can be measured by quantifying its changes with duration of the sample and with embedding. Symmetrization is demonstrable in chaotic series (Fig. 8) but is not prominent in biotic series. Empirical series are heterogeneous in this regard.

Fig. 8. Symmetrization in three types of chaos but not in bios. The absolute values of skewness are here calculated prior to averaging.

**INFORMATION: REPETITION, RISE AND FALL**

In relatively static systems, information is conveyed by the difference between successive terms (Bateson, 1979). In the far more frequent case of active processes, information is foremost conveyed foremost by repetition (Sabelli, 2005). The BDA quantifies information in time series by counting rise, fall or repetition (with a 10% tolerance) between successive terms, and counting the frequency in which these changes occur as illustrated in Fig. 9. The distribution of sequences of repetition, rise and fall shows triadic patterns in logistic, Henon and tent map chaos, and tetradic patterns in process, Ikeda, Lorenz, and Rössler chaos and in biotic patterns found in empirical data or generated mathematically. This may be significant, as three attractors generated
chaos, and period 3 is prominent in the logistic recursion and is the endpoint of Sarkovskii’s series (Peitgen Jhrgens & Sauge, 1992), while period 4 is prominent in biotic equations, and at least four feedback values (two pairs of opposites) are required to generate bios mathematically (Sabelli, 2005). The relative frequency of rise and fall also allows one to distinguish three types of tetradic patterns.

Process chaos, $g=4.2$

![Process chaos diagram](image)

Logistic chaos

![Logistic chaos diagram](image)

**Fig. 9.** Repetition, rise and fall in logistic and process chaos. Each column represents the percentage of occurrence of each of nine possibilities: rise followed by fall, repetition or rise (white columns), etc.
TRIGONOMETRIC ANALYSIS OF OPPOSITION

The analysis of interacting opposites is of practical importance in many natural and human processes, but often only one time series is readily available. For instance, heart rate is largely determined by the opposing actions of the accelerating sympathetic and the decelerating parasympathetic nerves. Measuring their relative contribution is clinically relevant (Malik & Camm, 1995), but only one set of data, heartbeat intervals, is readily available. How can one study complementary subprocesses when only one time series is available? To study opposite subprocesses, we calculate the sine and cosine of each term in the series, thereby decomposing a single time series into two complementary surrogate series (sine and cosine transforms). Sine and cosine are paradigmatic of complementary opposites. They wax and wane out of phase but not independently. Sine and cosine are functions defined by the projection to the \( x \) and the \( y \) orthogonal axes of a point moving around a circle in the \( xy \) plane. Also, the sine and the cosine are orthogonal functions in the sense that the integral from 0 to \( 2\pi \) of the product \( \sin(x) \cos(x) \) is equal to 0. Sine and cosine can lie in orthogonal planes, just as the electrical and magnetic components of light waves. Sine and cosine transforms thus provide a two-dimensional framework in which to project a time series and examine the relation between its opposite components.

Complement plots (Fig. 10) are generated by plotting the cosine and sine transforms in the \( X \) and \( Y \) axes, and by drawing a straight line between successive points to represent transitions. In the complement plot, the value of each term in the times series (modulo \( 2\pi \)) corresponds to a point in a circumference. The vertical axis is the sine scale and the horizontal axis is the cosine scale, with zero in both scales corresponding to the center of circumference. The plot is circular for series of numbers with a range equal or larger than \( 2\pi \). Complement plots reveal intriguing Mandala patterns in heartbeat interval series (Sabelli, 2000); illness can obliterate this pattern. Similar Mandala patterns are detected in other empirical time series, including some economic and meteorological time series; many examples can be found (Sabelli, 2000a, 2005). Shuffling the data completely obliterates the Mandala pattern. Complement plots of non-integer random numbers and Brownian noise are uniform (Fig. 10), while biotic series generate patterned plots. Complement plots of integer random numbers generate simple forms. A Mandala pattern of concentric rings is present in integer biotic series generated by the process equation \( A_{t+1} = A_t + g \cdot \sin(A_t) \) and by random walks with integer steps (greater than 2 and less than 13). Shuffling destroys these patterns.
The fact that remarkable regularity can be obtained by graphing data in a complement plot when the same data appear irregular in return maps points to the vital dialectic of complementary opposites in natural processes. This is understandable in the case of heart rate variation that is generated by the opposing actions of the sympathetic and parasympathetic nerves. It is also readily explainable by interactions of consumption and production in economic processes.

**Fig. 10.** Complement plots rounded-off time series display distinct patterns not readily observable in the original data. Mandala patterns are observed for biotic series and for random walks but not for random data. N=500.

We also generate two surrogate time series by adding the successive terms of the series of sines and of the series of cosines:

\[ S_{t+1} = S_{t+1} + \sin A_t, \quad S_1 = 0 \]

and

\[ C_{t+1} = C_t + \cos A_t, \quad C_1 = 0. \]

We refer to these series as sine and cosine walks. These trigonometric walks clearly reveal the degree of symmetry of opposite components of the process. Series generated by symmetric bipolar feedback such as \( A_{t+1} = A_t + g \ast \sin(A_t) + g \ast \cos(A_t) \) generate quantitatively similar sine and cosine walks, while series generated by asymmetric bipolar feedback, either \( A_{t+1} = A_t + g \ast \sin(A_t) \) or \( A_{t+1} = A_t + g \ast \cos(A_t) \), show
marked difference between them. One may thus distinguish symmetric from asymmetric opposition. In almost all empirical time series examined, the sine and cosine walks are of markedly different magnitude. This indicates that natural opposites do not wax and wane together in a quantitatively related fashion.

![Graphs of various chaotic and empirical series]

**Fig. 11.** Trigonometric plots of chaotic series and of empirical series. Note the differences in the horizontal and vertical scales. Logistic, Henon, and Ikeda's chaos show simple linear trajectories. Trigonometric plots of empirical series show both curvilinear and superficial trajectories.

Heartbeat intervals of all healthy persons show superficial trajectories such as observed with mathematical bios. In contrast, curvilinear trajectories appear in some heartbeat series of patients with
severe cardiac illness. Curvilinear trajectories are also found in profoundly psychotic individuals who are cardiologically healthy. Linear, curvilinear and superficial trajectories obtain in physical, meteorological, biological, and economic time series. Thus, in spite of their simplicity, trigonometric plots differentiate illness from health, distinguish different types of chaos, different types of bios, and apparently similar empirical time series.

Trigonometric plots are generated by plotting the cosine and sine walks in the X and Y axes. Dekking and Mendès-France (1981) use such plots, which they called "curlicues" by analogy with architecture, to study mathematical curves, measuring their properties in terms of entropy. We interpret these plots as portraits of opposites in empirical and mathematical series. Linear plots indicate linear opposition; this is observed in Henon, Ikeda and logistic chaos (Fig. 11). In contrast, process chaos produces a uniform trajectory that covers the entire surface of the plot, indicating orthogonal uncorrelated opposites. Consideration of these simple mathematical series clarifies how sine and cosine walks provide information regarding opposites that generate a given process. The logistic recursion involves the multiplication of linear opposites $A_t$ and $(1-A_t)$; this generates a quadratic chaotic process, which we may regard as two-dimensional in the same sense as the product of two lengths generates a plane. The cobweb plot of logistic chaos represents the properties of this recursion in terms of a two-dimensional process in the graph of the parabola $y = g (A_t \cdot (1-A_t))$ in the plane whose axes are $A_t$ and $y$. The trigonometric plot of logistic chaos represents the linearity of the opposition by the linearity of its trajectory, and the bidimensionality of the process by its diagonal orientation. In contrast, process chaos is generated by trigonometric functions that involve a circle of opposites—i.e., nonlinear and bidimensional opposites. Correspondingly, the trajectory in the trigonometric plot is a rectangular area; there is no correlation between sine and cosine walks.

Trigonometric plots of many empirical series (Fig. 11), mathematical bios, some chaotic series (Lorenz, Rössler), random data, and stochastic noise display curvilinear trajectories (actually resembling curlicues) or mottled superficial plots. These patterns indicate partial opposition—i.e., the opposite components are neither linear (as in logistic chaos) nor orthogonal (as process chaos), but instead they display an intermediate degree of nonlinearity. Trigonometric transformations thus demonstrate bipolar, partial, uncorrelated and asymmetric opposites in empirical processes and bios, supporting the hypothesis that bios is generated by bipolar feedback in cardiac
(sympathetic vs. parasympathetic innervation), economic (supply vs. demand), and mathematical equations, and at variance with standard concepts of polar opposites and complementary opposites.

**PROCESS ENTROPY**

To portray both simple and complex levels of organization, process entropy is computed in multiple dimensions. Such measurements distinguish creative from mechanical and random processes, and show that entropy measures symmetry, variety, and organization rather than disorder (a drastic departure from standard presentations of thermodynamics).

Measures of entropy have been found useful in many fields. A large number of entropy measures have been introduced, some of them intended to quantify the complexity of organization. The BDA measures the entropy of time series with the de Moivre equation, used as the definition of entropy in statistics, statistical mechanics and information theory (Wicken, 1987):

\[ H = -\sum P_i \log (P_i), \]

where \( P_i \) is interpreted as the relative frequency of a given value (not necessarily a probability). This equation measures the regularity of a distribution in a histogram: when there is an equal number of data points in each bin, \( H \) is maximal. \( H \) has no intrinsic dimension. Whenever one bin is empty, \( P_i \log_2 (P_i) \) is computed as 0.

The BDA measures entropy with different number of bins (e.g. 2 to 100 bins). The value of entropy increases with the number of bins used for its calculation (Fig. 12), just as the amount of information received depends on the discriminatory power of the receiver. Varying the number of bins allows one to deconstruct entropy into two components: symmetry \( s \) and diversity \( d \) (Fig. 3). Entropy \( H \) is, within limits, approximated by a linear function of the logarithm of the number of bins \( n \) is the number of bins used to calculate entropy:

\[ H = s + d^* \log_2 n \]

where both \( s \) and \( d \) vary from 0 to 1. The slope \( d \) of the entropy/bin curve measures diversity, i.e., the number of different values in the data. The slope is 0 for numerical series with two equally probable values, regardless of their temporal arrangement (random or periodic), and it is 1 for random, sinusoidal, chaotic, biotic series stochastic series noise
(random walks, pink noise), and many empirical time series (Fig. 12). The intercept s depends on the symmetry of the data. It is 1 for all symmetric distributions (random, sinusoidal, chaotic, etc), and is lower for economic and physiological data, biotic series generated by the process equation, and in random walks. Thus, in evolving equations, entropy decreases in the transition from chaos to bios. This is in line with the notion that complexity is associated with lower entropy. To consider creative phenomena, the BDA considers simple and complex components by measuring the entropy of successive epochs of a time series; the entropy of the differences between consecutive terms of the series; and the entropy of consecutive recurrence of isometric vectors of N successive terms at various embeddings (recurrence entropy). Comparison of these various measures of entropy for multiple types of series, however, indicate a more complex relation that lie beyond the scope of this article (see Sabelli, 1999 and in press 2005).

![Fig. 12. Entropy-bin plot shows the calculation of symmetry (intercept at 2 bins, log2 2 = 1) and diversity (slope). Biotic series have a certain degree of asymmetry. Chaos, random, ordered series, and period 2 are totally symmetric. Note: Lines for chaos, random, and ordered series practically overlap.](image-url)
DISCUSSION

In summary, the BDA presents an integrated set of new methods for time series analysis that defines (1) operationally creative processes and (2) distinguishes causal chaos and bios from stochastic processes. The embedding plot is particularly useful in both respects because it analyzes both simple (causal) and complex (created) components of the time series. It thus distinguishes causal creative processes that include both simple and complex components of variation from stochastic processes that include only complex components, and from chaotic processes that decay from order to randomness as the number of dimensions is increased.

A process is creative when its time series reveals the coexistence of simple and complex components of variation, and displays intermittent pattern, diversification, novelty, nonrandom complexity, and irreversibility.

1. Coexistence of simple and complex components of variation can be shown by embedding plots and by power spectrum. Creative evolution is the growth of complexity (high dimensional pattern) from a simple origin. In contrast, high dimensional random processes generate stochastic series. Dimensionality and pattern are determined and unchangeable in mechanistic processes.

2. Episodic patterning: When a process generates new patterns, by necessity each pattern must have a limited lifetime, so the time series consists of a sequence of episodes separated by transitions. These time-limited patterns can be detected as clusters of recurrences that we call complexes. In contrast, the time series of noncreative processes must have a uniform pattern, no matter how variable the individual terms.

3. Diversification: In a creative process, the patterns are not only new but also unique in regards to each other. As a result, variety increases with time; this is diversification; in noncreative processes, the variance, no matter how large, is stable.

4. Novelty: Creative processes generate innovation; novelty represents a different type of innovation than that measured by the Hurst exponent (Mandelbrot, 1977).

5. Nonrandom Complexity: Arrangement measures nonrandom complexity, in contrast to algorithmic complexity (Chaitin, 2001) that is highest for random data.

6. Irreversibility: Irreversibility, characteristic of natural processes, can be demonstrated in biotic series generated by mathematical recursions (Sabelli, 2003), but not in chaos, stochastic series, or idealized mechanical models.
A more detailed description of these and related methods, and a discussion of their applications and limitations is presented in Bios. *A Study of Creation*. Bios analysis complements chaos methods. Bios is a chaotic process, so in the analysis of a time series is necessary first to demonstrate chaoticity and then examine for creative features. For instance, when we studied the time series of oil prices and volume of trade with chaos methods, we found complexity (Levy-Carciente et al., 2004), but only further analysis with bios methods revealed that this complexity involves bipolar feedback processes (Sabelli, 2004).

Bios analysis is a component of a generic theory of creative processes (Creation Theory, Sabelli, 1989; 1999; 2005) according to which fundamental processes at every level of organization—physical, biological and human—are causal and creative, rather than determined or stochastic. More specifically, the co-creation hypothesis postulates the generation of bios by bipolar feedback as one of the generic mechanisms creating novelty and complexity at all levels of organization (Sabelli, 2001c). This concept provides reasons and tools for human intervention.

**ACKNOWLEDGMENTS**

This research was supported by the Society for the Advancement of Clinical Philosophy. We are thankful to Maria McCormick for her support, to Dr. J. Messer for electrocardiographic data, and to K. Walthall, C. Tom and O. Zdanovics for technical assistance.

**APPENDIX**

Heartbeat interval series one hour or longer are obtained from the CCCD-Rush Psychocardiology Data Base that consists of 24 hour electrocardiographic recordings obtained by Holter monitoring from adult subjects, both women and men as described in previous publications (Carlson-Sabelli et al., 1994, 1995, 1997; Sabelli et al., 1994, 1995a, 1995b, 1997; Sabelli & Carlson-Sabelli, 1999) The data collection was authorized by the Human Investigation Committee of Rush-Presbyterian-St. Luke's Medical Center. These recording are scanned to eliminate artifacts, and, when desired, extrasystoles. Economic data includes economic indexes (Dow-Jones Industrial Average, Standard & Poor's Composite Index), prices of commodities such as corn, oil, and gold; and currency exchange rates in the public domain as described (Sabelli & Kauffman, 1999; Levy-Carciente et al., 2004). Data for the distribution of galaxies have been collected by researchers at the Las Campanas Observatory in Chile (Shectman et al., 1996) and the Anglo-Australian Observatory (Colless, 1999) and the processing of these data is described in detail in Sabelli and Kovacevic
(2004). Meteorological data are obtained from PICES Technical Committee for Data Exchange and other sources in the public domain. Paleoclimate indicators are obtained from the NOAA/NGDC Paleoclimatology Program, Boulder CO, USA.

REFERENCES


