RESEARCH ARTICLE

Past the Power Law: Complex Systems and the Limiting Law of Restricted Diversity

BRIAN CASTELLANI[1] AND RAJEEV RAJARAM[2]

[1]Department of Sociology, Kent State University, Ashtabula, Ohio 44004; and [2]Department of Mathematics, Kent State University, Ashtabula, Ohio 44004

Received 12 December 2015; revised 5 April 2016; accepted 6 April 2016

Probability distributions have proven effective at modeling diversity in complex systems. The two most common are the Gaussian normal and skewed-right. While the mechanics of the former are well-known; the latter less so, given the significant limitations of the power-law. Moving past the power-law, we demonstrate that there exists, hidden-in-full-view, a limiting law governing the diversity of complexity in skewed-right systems; which can be measured using a case-based version C_c of Shannon entropy, resulting in a 60/40 rule. For our study, given the wide range of approaches to measuring complexity (i.e., descriptive, constructive, etc), we examined eight different systems, which varied significantly in scale and composition (from galaxies to genes). We found that skewed-right complex systems obey the law of restricted diversity; that is, when plotted for a variety of natural and human-made systems, as the diversity of complexity $\rightarrow \infty$ (primarily in terms of the number of types; but also, secondarily, in terms of the frequency of cases) a limiting law of restricted diversity emerges, constraining the majority of cases to simpler types. Even more compelling, this limiting law obeys a scale-free 60/40 rule: when measured using C_c, 60%(or more) of the cases in these systems reside within the first 40% (or less) of the lower bound of equiprobable diversity types—with or without long-tail and whether or not the distribution fits a power-law. Furthermore, as an extension of the Pareto Principle, this lower bound accounts for only a small percentage of the total diversity; that is, while the top 20% of cases constitute a sizable percentage of the total diversity in a system, the bottom 60% are highly constrained. In short, as the central limit theorem governs the diversity of complexity in normal distributions, restricted diversity seems to govern the diversity of complexity in skewed-right distributions. © 2016 Wiley Periodicals, Inc. Complexity 000: 00–00, 2016

Key Words: complex systems; probability distributions; diversity of complexity; Shannon entropy; limiting laws; case-based modeling

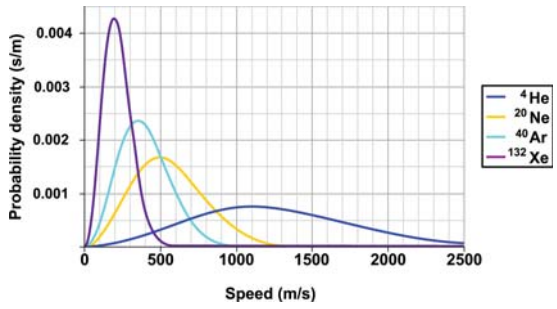
Correspondence to: Brian Castellani, E-mail: bcastel3@kent.edu

1. PROBABILITY DISTRIBUTIONS IN COMPLEX SYSTEMS

As identified by Sornette [1] and others [2,3], probability distributions are the "first quantitative characteristics of complex systems," [1] providing researchers

DOI 10.1002/cplx.21786
Published online 00 Month 2016 in Wiley Online Library (wileyonlinelibrary.com)

FIGURE 1



Maxwell-Boltzmann molecular speed distribution for noble gases figure from wikipedia creative commons (commons.wikimedia.org/wiki/File: MaxwellBoltzmann-en.svg).

C O L O R

an effective tool—albeit somewhat simplistic—for identifying and describing macroscopic regularities that, otherwise, would be difficult to detect. Perhaps one of the most classic examples, from statistical mechanics, is the Maxwell–Boltzmann distribution for kinetic energy of an ideal gas at thermodynamic equilibrium (See Figure 1).

Probability distributions are also useful because—as demonstrated by the central limit theorem, power-laws, etc—measurements on a wide range of physical, biological, psychological, sociological, and ecological systems are well approximated by their shape, particularly as the sample size (or number of samples or trials) $n \to \infty$.

The two most commonly studied distributions in the complexity sciences are the Gaussian bell-shaped distribution and the power-law distribution [1,3]. As Sornette explains, "The Gaussian distribution is the paradigm of the 'mild' family of distributions. In contrast, the power-law distribution is the representative of the 'wild' family"(p. 3). Mild distributions are those with short (thin) tails, as in the case of blood pressure, height or intelligence; while wild distributions are marked by fat (long) tails, as in the distribution of incomes, cities, or galaxies. Not all wild distributions, however, meet the assumptions of the power-law.

As Sornette [1] and others explain [3,4], the power-law, while significant, is somewhat finicky and particular. For example, it only works for those distributions that decrease monotonically; and, when it does work, it only does so" over a finite range of event sizes, either bounded between a lower and an upper cut-off, or above a lower threshold—that is, only in the tail of the observed distribution (p. 6)" [1]. In other words, while the power-law has important implications for understanding the distribution of complexity within skewed-right complex systems, as Clauset et al. (2009) state," "Unfortunately, the detection and characterization of power-laws is complicated by the large fluctuations that occur in the tail of the distribution—the part of the distribution representing large but rare events—and by the difficulty of identifying the range over which power-law behavior holds" (p. 1). In short, while important, the power-law itself is not powerful enough to model fully the distribution of the diversity of complexity within many skewed-right distributions.

The same "limitation of fit" is true of the larger class of wild to less-wild skewed-right models, such as stretched-exponential, Burr, Log-Pearson, Log-Logistic, Pert, and Dagum distributions. While each is useful in particular instances, none generalize across *all* skewed-right distributions sufficient to compare and contrast the diversity of complexity within and amongst different complex systems, and across different levels of scale.

The question therefore remains: is there a law *hidden in full view* within one of the most common distributions in complex systems, namely the skewed-right distribution? And, if so, can it be identified and measured, above and beyond the restrictions of the power-law?

1.1. Restricted Diversity and the 60/40 Rule

As we will spend the rest of this study attempting to demonstrate, we believe that the skewed-right distribution found in many complex systems is governed by the law of restricted diversity and its corresponding 60/40 rule.

As shown in Figure 2, an easy example of our argument is the phylogenetic tree: moving from bacteria and simple cells to primates and, more specifically, humans, it seems that life evolves toward increasing levels of complexity [5,6]. If one constructed a histogram of this evolution, grouping all of life according to some measure of complexity (i.e., cellular, morphological, cognitive, behavioral, or informational), the result would be (moving from the least to the most complex along the x-axis) an obviously rather diverse list of different species types. In other words, the probability distribution for life on planet earth, as a complex system, possesses a *natural distribution of the diversity of complexity.*

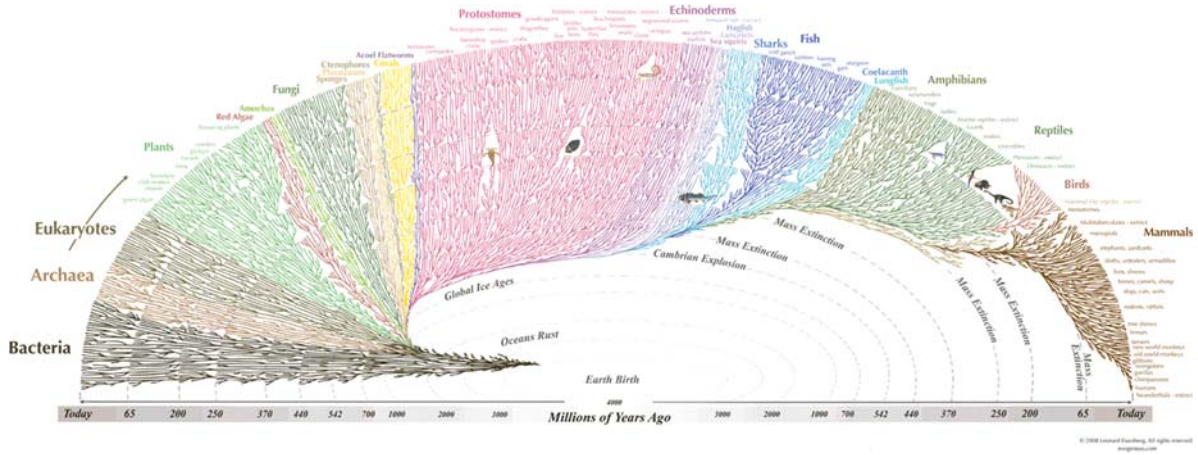
What is also interesting about the diversity of complexity for this system is that there appears to be a *majority of simplicity* phenomena. In other words, despite the drive toward complexity, most cases seem to maintain a simpler form, confining themselves to the least complex probability types in a system. The result is a probability distribution that takes the signature shape of a skewed-right (positive) distribution, with or without long-tail—See Figure 3.

For example, as creatively demonstrated in Figure 2, starting with single-cell eukaryote or prokaryote cells, the majority of living organisms (cases) on the phylogenetic tree are located near the lower bound of the diversity of complexity, while the remaining distribution of cases decreases asymptotically as complexity increases [5,6].

DOI 10.1002/cplx

C
O
L
O
R

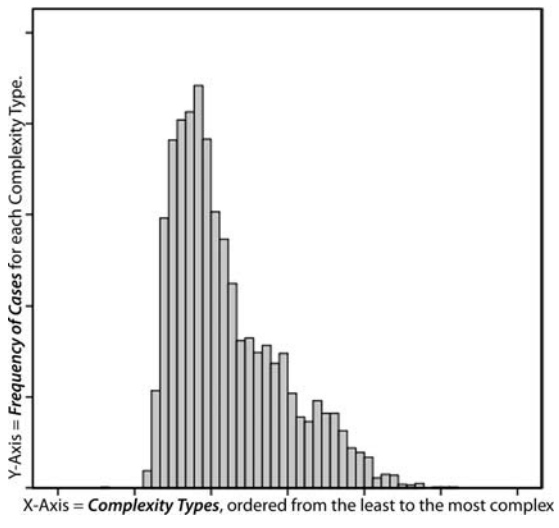
FIGURE 2



Phylogenetic tree of life. Figure by Leonard Eisenberg; http://www.evogeneao.com.

For many scholars, such an insight may seem trivial. And, perhaps, at first glance, it is. But, upon further exploration, it is not. While researchers have shown that the height of mammals is governed by the central limit theorem, mammalian body mass is not (See Figure 5); and, while the economic distribution of countries is often Gaussian (see e.g., [7]), the financial distribution of the largest 500 companies is not (See Figure 5). In other words, a different type of limiting law seems to be at play in skewed-right complex systems.

FIGURE 3



Example histogram for the diversity of complexity.

Let us explain. While the study of power-laws has proven insightful, its focus on the long-tail has blinded scholars to an important and hidden-in-plain-sight countervailing point: the majority of cases in many skewed-right complex systems do not obey the central limit theorem; instead, they obey the law of restricted diversity; that is, they maintain a simpler form despite the drive toward an increased diversity of complexity. In other words, when plotted for a variety of natural and human-made systems, the resulting distribution yields a limiting law, albeit different from the central limit theorem: as the diversity of complexity $\rightarrow \infty$ (primarily in terms of the number of types; but also, secondarily, in terms of the frequency of cases) a *limiting law of restricted diversity* emerges in complex systems. Said again, the universal drive toward the diversity of complexity has a countervailing law such that, for any given system, as the level of complexity $\rightarrow \infty$, restricted diversity emerges to counteract it. And, it is not confined to power-law distributions.

And, if this is not compelling enough, as we will show, this *restriction in diversity* appears stringent and measurable. In fact, when assessed using a case-based version C_c of Shannon entropy H (which measures the true diversity of a system [8]) restricted diversity obeys a universal *60/40 rule*.

The notion of measurability is, by the way, the same reason the power-law is so compelling. As Newman [3] and others explain [1,4], while the pervasiveness of skewed-right distributions may not impress, it is significant that many of these distributions, when replotted logarithmically, take the straight-line form of a power law. "Now, 'argues Newman [3]', a remarkable pattern emerges (p. 27)." The same is true of the 60/40 rule: as shown in

DOI 10.1002/cplx

Figure 6, a remarkable pattern emerges. To make quick sense of this pattern, we need to explain the denominator in the 60/40 rule.

As we outline later, the denominator in the 60/40 rule is based on the total diversity of complexity in a system. To determine this total diversity we use C_c, which, for a given value of cumulative relative frequency c, calculates the percentage of the total number of equiprobable types needed to account for the total information in a system. In the case of the Phylogenetic tree, for example, if we were studying mammalian body mass, C_c would be equal to the percentage of the total number of equiprobable species-types needed to account for the entire diversity of information in this system, up to a cumulative relative frequency of c.

Based on C_c, given our focus on skewed-right complex systems, the 60/40 rule is interested in the lower bound of complexity; that is, the first 40% of equiprobable diversity types. And, what do we find when looking at this region of the graph? As shown in Figure 6, we find that (as in the case of mammalian body mass) 60% (or more) of the cases in skewed-right complex systems consistently reside within the first 40% (or less) of its lower bound of equiprobable diversity types—with or without long-tail and whether or not the distribution fits a power-law. In other words, as the central limit theorem seems to govern the diversity of complexity in normal distributions, the law of restricted diversity and its 60/40 rule seem to govern the diversity of complexity in skewed-right distributions. But, we are getting ahead of ourselves, as all of this remains to be demonstrated. And so we need to proceed to the purpose of our study.

1.2. Purpose of Current Study

This study is second in a series addressing the empirical/statistical distribution of the diversity of complexity within and amongst complex systems. In our first paper [8], we introduced and tested a measure called *case-based entropy* C_c. A modification of the Shannon-Wiener entropy measure H, [9] *case-based entropy* C_c is an effective tool for statistically calculating (a) the distribution of the diversity of complexity within and across multiple natural and human-made systems; as well as (b) the diversity contribution of complexity of any part of a system. What is more, C_c is not limited to only those distributions fitted by a power-law. In fact, C_c can be used to measure the distribution of the diversity of complexity in any system, as it is grounded, theoretically, in the framework of the true diversity of a uniform distribution and its corresponding equiprobable types [8]—more on this point below.

Using C_c as our measure, for the current study we seek to outline a series of scientific observations about the distribution of the diversity of complexity within different natural and human-made systems. For our study, we will examine eight different complex systems: (1) a segment of the World-Wide-Web, (2) household income in the United States for 2013, (3) body mass of Late Quaternary mammals, (4) the human diseasome map, (5) Hubble's classic data on the velocity of galaxies, (6) U.S. cities by population size for 2011, (7) Financial Times 2014 biggest 500 companies, and (8) the 12 month prevalence of mental disorders in the United States for a given year (See Figure 3).

Our resulting observations revolve around the four major themes highlighted above, which we will use the remainder of this study to further introduce, define, and empirically validate: (a) the natural diversity of complexity, (b) the limiting law of restricted diversity, (c) the "majority of simplicity" in complex systems, and (d) the 60/40 rule. But first we need to review how we chose the eight systems we studied, based on the challenges surrounding the definition and ranking of complexity, as well as provide a bit more detail about our new measure of complexity, case-based entropy C_c, including how it differs from the power-law and other such complexity measures.

2. DEFINING AND MEASURING COMPLEXITY IN SYSTEMS

2.1. Defining Complexity

As scholars in the complexity sciences know, given the academic breadth of the field (which ranges from sociology and economics to physics and biology to computer science and mathematics) there is considerable debate about what *complexity* is—that is, while a reasonable degree of consensus exists about such core concepts as self-organization, emergence and nonlinearity, scholars nonetheless differ somewhat widely in terms of (a) what, exactly, they think makes something complex; and (b) how, in turn, they think this complexity is best measured [10–13]. The result has been a proliferation of a wide array of approaches [10,11,14–18].

For example, a simple survey by Lloyd [19] counted over 40 different measures of complexity; which he is able to organize into one of three types. *Descriptive Complexity*: The first type defines and orders the complexity of a system based on its respective difficulty of description. Examples of such metrics include Shannon-entropy, algorithmic complexity and fractal complexity. It is worth nothing that, for this first type, complexity applies to the difficulty of description and not to the system itself.

2.1.1. Constructive Complexity

The second type is based on the challenge of creating a respective complex system, which can be measured in terms of such things as time, energy, dollars, etc. From this perspective, the more complex a system is, the harder it is to construct. But, "harder" in this instance is not restricted to the number of steps necessary to make something; instead, as in the case of evolutionary theory, it is

DOI 10.1002/cplx

based on the number and sequence of events necessary for something to come into existence or maintain. In this way—as Lloyd [19] and Mitchell explain [11]—humans are more complex than bacteria, given their phylogenetic order; the same it true of larger cities, companies or incomes: in terms of their evolution, they are all harder to create than (and often come from) smaller versions of themselves. Examples of such metrics include thermodynamic depth, computational complexity theory, and logical depth.

2.1.2. Organizational Complexity

The third type defines and orders the complexity of systems based on two inter-related quantities. The first quantity concerns the difficulty of describing a complex system's organizational structure, be that "structure" ecological, social, psychological, biological, chemical, physical, or mathematical in form. The second quantity concerns "the amount of information shared between the parts of a system" as the result of its particular structure. From this perspective, the more hierarchical, near-decomposable subsystems that a complex system is comprised of, the more complex that system is—as argued, for example, in Herbert Simon's famous 1962 paper, *The Architecture of Complexity*. In this way, the most complex networks, diseases, galaxies, organisms, or incomes are those comprised of the largest hierarchies of near-decomposable subsystems.

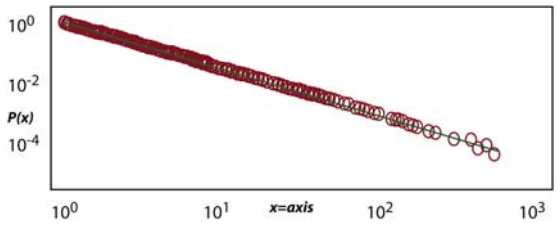
2.1.3. Additional Measures of Complexity

In addition to the three most highly recognized types summarized by Lloyd [19] there exists an additional set, including such metrics as cognitive, behavioral, cellular, morphological, or sociological complexity [5,6,11]. For example, one can measure disease in terms of genetic makeup or degree of psychological severity; or, one could measure the complexity of humans in comparison to other species in terms of the capacity for sociality or behavioral coordination, etc. And, that is not the end of it, as we also have to include the power-law and its *complexity of scale* approach.

2.1.4. Complexity of Scale

Here, the diversity of complexity is measured and ranked according to exponential differences in size, severity or magnitude [1–4]—which, when plotted as a log-log graph, takes the form of a power law (See Figure 4). In the case of networks, for example, scale is often defined as degree of connectivity; in terms of cities, galaxies or earthquakes it is often defined as magnitude; in terms of thermodynamics (as in the case of the Stefan–Boltzmann law) it is often defined as energy or temperature; and in terms of species or financial wealth, it is often defined as frequency.

FIGURE 4



Classic example of log-log power law as straight line.

Still, as these different approaches to defining and ranking the diversity of complexity suggest, no single one of them is sufficient. Instead, given different circumstances, they each prove useful in different ways. There is also overlap amongst the definitions. For example, the distribution of income, galaxies or cities can be seen as one of scale, organization, construction, or description.

Given this diversity of definition, the current study was careful to select from the complex systems literature a variety of different examples to study. And we chose widely, knowing that not everyone will agree with some of our choices. The result is the eight systems examined here, which varied significantly in scale and composition, as well as the definition and ranking of their corresponding complexity.

2.2. Eight Complex Systems

Here is a brief overview of each system, including how those who study them typically define and order the distribution of the diversity of complexity for each. Please note that, in some systems (as already discussed above), diversity constitutes an explicit ordering of cases from less to more complex forms, given the three major ways in which complexity is regularly defined (descriptive, constructive, organizational); while, in others, complexity is a function of scale (be it based on size, severity, magnitude, etc). For complete access to these data, go to: http://www.personal.kent.edu/~bcastel3/datawebsite.html.

1. *The World Wide Web* segment is a directed network comprised of 325,729 vertices and 1,497,135 arcs (27,455 loops; See http://vlado.fmf.uni-lj.si/pub/networks/data/ND/NDnets.htm). Following the literature, complexity was defined and ranked as the number of links for each document, with one link (arc) being the least complex and the hubs (the most densely connected nodes in the network) being the most complex. Of the examples chosen, this system (along with household income and cities) is one of the most widely studied using the power-law.

DOI 10.1002/cplx

2. *The 2013 USA Household Income* database is a sample of $N = 122{,}952$ households; and is part of the U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplement. For this sample, household income was binned in increments of \$4999 dollars, starting at $X \leq \$5000$ dollars and ending at $X \geq \$200,000$ dollars. Following the literature on the Pareto Law, 80/20 rule and Zipf's Law, complexity was defined and ranked as a function of income, moving from the smallest to the largest. (See http://www.census.gov/hhes/www/cpstables/032014/hhinc/hinc01_000.htm.)
3. *The Body Mass of Late Quaternary Mammals* (MOM v4.1) is part of a taxonomic list of all known mammals of the world (N=4629 species) to which the researchers added status, distribution, and body mass estimates compiled from the literature (See http://biology.unm.edu/fasmith/Datasets/) [20,21]. Here—following the measurement literature on organizational and constructive complexity, as well as allometry and scaling-laws [22]—complexity was defined and ranked as a function of body mass, moving from the smallest and least complex to the largest and most complex.
4. *The Human Disease Network* (HDN) is a list of 1284 distinct human diseases based on the number of genes for each disease [23]. Following the complex network literature and the literature on organizational and constructive complexity, complexity was defined and ranked according to the number of genes in the HDN associated with a disorder, ranging from diseases with one gene (N= 870) to diseases with a complex network of genes. (See http://www.barabasilab.com/pubs/CCNR-ALB_Publications/200705-14_PNAS-HumanDisease/Suppl.)
5. *Velocity Dispersion of Galaxies* comes from Hubble's original 1929 data. Following the literature on organizational, thermodynamic depth and constructive complexity, for this system complexity was defined and ranked according to galaxy mass, with larger and more complex galaxies being defined as spinning faster. (See http://darknet.fnal.gov/eeu/hubblediagram/).
6. *USA City Populations* is a 2011 U.S. Census Bureau database of all incorporated cities with populations $\geq 50,000$. For this system, following the current literature—which combines scale, organizational and constructive complexity (See [24–26])—complexity was defined and ranked as a function of city size, moving from the smallest and least complex to the largest and most complex (See http://www.census.gov/popest/data/cities/totals/2011/index.html.)
7. *The Financial Times 2014 Biggest 500 Companies* is an annual rating. Following the organizational complexity literature (specifically, managerial studies, econo-physics and economic complexity) [27–30] complexity was defined and ranked as a function of market value (also known as market capitalization), moving from the smallest and least complex to the largest and most complex companies.
8. *The U.S. National Comorbidity Survey Replication* (NCS-R) is a nationally representative household survey of English speaking adults (-year old) [31]. It was used by Kessler *et al* [31] to assess the 12-month *prevalence and severity of mental disorders in the U.S. population*. For the current study, we focused on the NCS-R Part 2, which clustered the sample according to differences in their comorbidity profile. Based on these criteria, the NCS-R Part [2] clustered the sample (N=3199) into seven major classes, creating a seven-class index of the prevalence and severity of mental disorders in the United States, ranked from the least complex (Class 1) to the most complex (Class 7). For our study, we used this 7-class ranking and its corresponding frequency distribution, based on a combination of constructive and organizational complexity, as well as psychological and cognitive complexity.

2.3. Measuring Complexity

An effective way to define and measure the diversity of complexity D in a probability distribution (as show in Figure 3 above) is through a specific form of Shannon entropy H [9]. With such a modified measure (i.e., complexity index), researchers can determine the overall information (complexity) contributed by the range of diversity types in a system, relative to their different frequencies of occurrence.

Our impetus for using the Shannon entropy index H comes from evolutionary biology and ecology, where it is employed to measure the true diversity of species (types) in a given ecological system of study—here, by way of illustration, we refer the reader back to our earlier discussion of the phylogenetic tree [32–35]. More specifically, our impetus comes from Jost's paper on *Entropy and Diversity*, [33], in which he states that: "In physics, economics, information theory, and other sciences, the distinction between the entropy of a system and the effective number of elements of a system is fundamental. It is this latter number, not the entropy, that is at the core of the concept of diversity in biology" (p. 363). Following Jost, we approach the definition of the diversity of complexity as follows (for complete details of our approach, see [8])

Definition

Given an ordered set of diversity types numbered as $i \in \mathbf{N}$ and their corresponding probabilities p_i, with the

DOI 10.1002/cplx

understanding that larger values of i denote a higher degree of complexity, the diversity of complexity of the entire distribution (or a part of it) is defined as the number of equi-probable types of complexity needed to yield the same value of Shannon entropy H.

Here, following our above review of the measurement of complexity, a *complexity type* (which would be located on the *y*-axis) constitutes some important quantitative or qualitative distinction amongst cases, relative to their differences in complexity. Such differences (as outlined in our review of measuring complexity) can be descriptive, constructive or organizational in form, as well as differences in scale or other such metrics (cognitive, behavioral, cultural, etc.). Also, given the metric used, higher degrees of complexity vary in form, ranging, again, based on matters of description, construction, organization, scale, etc.

Following this definition, given the probability p_i of the occurrence of a particular type of complexity *i*, the Shannon–Weiner entropy index *H* is given by:

$$H=-\sum_{i=1}^{N} p_i \ln(p_i) \tag{1}$$

The problem with the Shannon entropy index *H*, however, is that, while it is useful for studying the diversity of a single system, it cannot be used to compare the diversity of complexity between or across systems. In other words, *H* is not multiplicative, that is, a doubling of value for *H* does not mean that the actual diversity of complexity has doubled. For example, one could not compare the *H* for the world-wide-web in relation to the *H* for the human genome map or U.S. cities by population size.

In response, we draw upon the *true diversity measure D* [32,36,37]. The utility of this measure is that, given the value of *H*, it can compute the number of diversity types that have the same probability of occurrence and gives the same value of *H*, as given by the following formula:

$$D=e^{H}=\Pi_{i=1}^{N}\frac{1}{p_i^{p_i}} \tag{2}$$

In other words, the utility of *D* for comparing the diversity of complexity across systems is that in *D* a doubling of the value means that the number of equi-probable types of complexity in a complex system of study has doubled as well, a property that is not true of *H*. *D* calculates the number of such equi-probable types of complexity that will give the same value of Shannon entropy *H* as observed in the complex system.

But that is not where our develop of *H* ends. To define and measure the distribution of the diversity of complexity, we also need to be able to compute the percentage contribution of any part of a complex system—say, for example the first *K* set of types—relative to the overall diversity. In other words, we need to be able to compute the diversity contribution D_c up to a certain cumulative probability *c* of the first *K* set of types.

To do so, we still employ *D*, but we replace *H* with H_c—which is the conditional entropy, given that only the first *K* types are observed with conditional probability of occurrence $\hat{p}_i$ of the first *K* set of types within the given cumulative probability *c*, as given below:

$$\hat{p}_i=\frac{p_i}{c} \tag{3}$$

$$H_c=-\sum_{i=1}^{N}\hat{p}_i\ln(\hat{p}_i) \tag{4}$$

$$D_c=e^{H_c}=\Pi_{i=1}^{K}\frac{1}{\hat{p}_i^{\hat{p}_i}}=\frac{c}{\Pi_{i=1}^{K}p_i^{\left(\frac{p_i}{c}\right)}} \tag{5}$$

$$C_c=\frac{D_c*100}{D}. \tag{6}$$

In other words, if, for each complex system studied, *D* stands for the true diversity of the entire dataset, and D_c stands for the true diversity of the dataset up to a cumulative probability *c*, then we can plot the percentage diversity contribution (given by $\frac{D_c*100}{D}$) versus the cumulative probability *c*. For example, we could plot the percentage diversity contribution for some *K* set of nodes for some segment of the World Wide Web versus their cumulative probability *c*. The same could be done for the velocity spin of galaxies, income distributions, and so forth, across all the examples listed earlier. And, because 100% of the diversity contribution comes from the entire database ($c=1$ or a cumulative frequency of a 100%), $(100, 100)$ will be the end point of the graph. We call this novel measure *case-based entropy* C_c.

2.4. Case-Based Entropy versus the Power-Law

As a point of clarification—and also to build on our previous discussion about measuring complexity—it is necessary to understand that case-based entropy C_c far exceeds the limitations of r-law for modeling skewed-right distributions, wild, otherwise.

Summarizing the current literature [1,3,4], a power-law is a description of the nature of decay of the distribution of the diversity of a complex system at its tail-end. In most cases, only the tails of distributions fit power-laws. While a larger exponent in the power-law signifies faster decay of the tail, it does not characterize the beginning or the middle part of a distribution—which is the focus of restricted diversity. Furthermore, the power-law cannot be used as a tool to characterize the distribution of diversity of complexity, as the cumulative frequency of cases varies from 0 to 100. And, it cannot be used to make a comparison of the distribution of diversity across multiple

DOI 10.1002/cplx

distributions. Two reasons: (1) such a comparison requires considering the entire distribution and not just the tail; and (2) such a measurement and comparison requires a standardization of scales so to speak, which is exactly what is achieved by our case-based entropy measure—which standardizes the distributions by computing the number (or extent) of types required to maintain the same entropy as the uniform distribution, where all types are equi-probable.

In summary then, our notion of case-based entropy goes above and beyond the power-law in terms of modeling the distribution of diversity of complexity both within and across distributions, irrespective of whether or not the distribution follows a power-law at its tail-end. And, this is precisely what we demonstrated in our initial exposition of case-based entropy in Rajaram and Castellani (2016) [8].

And, it is what we also seek to demonstrate in the current study. While some of the distributions in our examples might fit a power-law for their tails, like we have explained above, that is beside the point we want to make in this paper. Our goal is to: (a) provide a standardized measure of the distribution of diversity of complexity of the entire distribution (as opposed to the tail or any other part) as the cumulative relative frequency of cases varies from 0 to 100; (b) create this standardized measure based on the number of diversity types required to provide the same amount of entropy or information; (c) use this standardized measure to compare the distribution of diversity as a function of cumulative frequency for various distributions of trace variables of complexity; and (d) use this standardized measure as a means of detecting the presence of a restriction of diversity in complexity, by checking for the 60–40 rule.

3. OBSERVATIONS ON THE DIVERSITY OF COMPLEXITY

The following constitute our preliminary observations on the diversity of complexity in systems. They are based on our analysis of the phylogenetic tree of life and, as shown in Figures 5 and 6, eight additional systems—see section 2 for details on these systems and their databases. For complete access to all these data, go to http://www.personal.kent.edu/~bcastel3/datawebsite.html.

Our results revolve around four major themes: (1) the natural diversity of complexity, (2) the "majority of simplicity" in complex systems, (3) the limiting law of restricted diversity, and (4) the 60/40 rule.

3.1. The Natural Diversity of Complexity

Observation 1

There is a natural distribution to the diversity of complexity within most systems. As shown in Figure 5, all eight systems we examined, from the velocity spin of galaxies to the body mass of Late Quaternary mammals to the human diseasome map, demonstrated a natural distribution of the diversity of complexity. By the word "natural" we mean two things:

First, we mean that, while complexity can often be assessed macroscopically [10–12,14,19], for many systems, the distribution of cases, relative to this global measure, will vary, and often to a considerable degree. As one example, while it is certainly the case that the world-wide-web is far more complex than a local friendship network, it is not the case that every website (node) in the world-wide-web (in terms of degree connectedness, e.g.) is equally complex. In other words, there is a *diversity of complexity* within the world-wide-web, ranging from its most densely connected websites (hubs) to the overwhelming majority of websites, which have an extremely low level of connectedness. And, this diversity of complexity is distinct from (albeit it relative to) its overall complexity.

Second, we mean that these "distributions of complexity" are easily ordered along the x-axis of their respective histograms, going from the least to the most complex—be the definition of complexity descriptive, constructive or organizational in form or based on scale or another approach to measuring complexity. In other words, while the metric for each system obviously has to be defined (which is true of any measure, be it physical, biological, social, etc), it is not necessary to impose or force the order of complexity on the distributions for these systems, which is why we use the term "natural diversity" to describe them. In other words, the order of the diversity of complexity in these systems appears to be spontaneous and self-organizing.

Observation 2

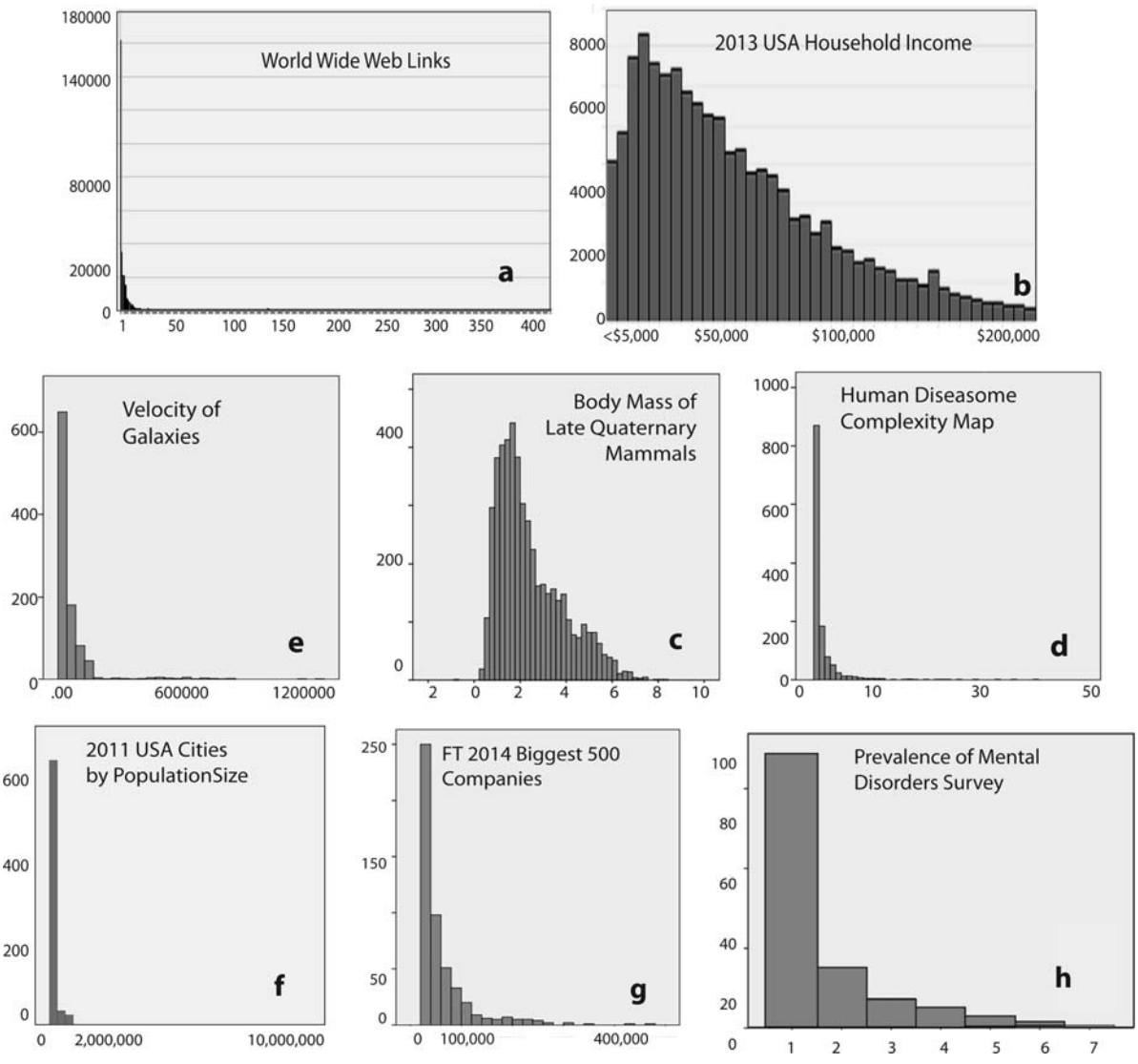
The natural distribution of the diversity of complexity takes the form of multiple types. At its most basic, a *type* constitutes a category of similarity, based on some shared set of characteristics. In terms of the current study, as explained in the section 2, a *type* is a form of complexity, based on a common set of attributes or features, which allows these types to be ordered, relative to their respective level or degree of complexity. Consider, for example, the phylogenetic tree: the distribution of its diversity is roughly 8.7 million species and counting. Each species is a type; each species is a particular form of complexity. In other words, depending upon the measure used (i.e., cellular, morphological, cognitive, behavioral, or informational), each species on earth, from prokaryote bacteria to modern humans, can be identified, grouped, and ordered according to its degree of complexity [5,6] —which we actually did in the case of the body mass of Late Quaternary mammals, resulting in the histogram shown in Figure 5.

Observation 3

In a complex system, each case constitutes an empirical instance of a complexity type. In the systems we

DOI 10.1002/cplx

FIGURE 5

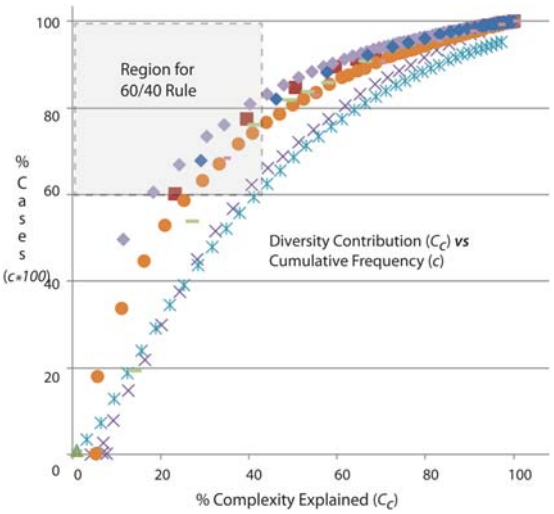


Restricted diversity histograms for eight systems. The data for the eight histograms came from the following databases: (a) a directed network segment (325,729 vertices and 1,497,135 arcs) of the World Wide Web; (b) The U.S. Census Bureau's 2013 database of U.S. household income; (c) the body mass of all known ($N = 4629$) species of mammals; (d) a list of 1284 human diseases by gene count; (e) Hubble's original 1929 galaxy velocity data; (f) the U.S. Census Bureau's 2011 database of all incorporated U.S. cities with populations over 50,000; (g) the Financial Times (FT) 2014 market value ratings for the largest 500 companies; and (h) a prevalence study ($N = 3199$) of mental disorders (during a 12-month period) in the United States. Reading the Histograms. The first seven systems are quantitative measures of complexity: in other words, the range of complexity types found on the *x*-axis for these systemsÅmoving from left (less complex) to right (more complex)Åwas a continuous increase. Moreover, the *y*-axis was an empirically derived frequency measure of actual cases, sampled from the population at large for each particular complex system. In turn, the last systemÅthe Prevalence of Mental Disorders SurveyÅwas qualitative: while the *y*-axis was the same as the other eight systems, the *x*-axis contained a nominal ordering of complexity types, arranged (roughly speaking) from the least to the most complex, moving from left to right.

observed, a case could be just about anything, from an atom or gas particle to a gene or biological cell to an ant or human being to an ecosystem or economy to a planet or galaxy. In fact, a case could even be a "complexity type." Case in point: mammalian body mass is interesting because, for this system, the metric is body mass, but the cases are also, in a sense, types, as they constitute the range of all known mammals. In other words, in this example we have a distribution of diversity types (mammalian species)—which we treat as our cases—relative to

DOI 10.1002/cplx

FIGURE 6



Graph of 60/40 rule for eight complex systems.

C
O
L
O
R

another type of complexity (body mass). And, even with such a different approach to defining cases and their complexity, we found that all the scientific observations made in our study held true.

Observation 4

Each complexity type in a system constitutes a probability state: one form of complexity possible for the cases in a complex system. As a category of similarity, *diversity types* function as probability states, around which the cases in a system naturally cluster, given some shared set of characteristics—descriptive, constructive, organizational, scale, etc. This is not, by the way, a new idea: in the case of the classic Maxwell–Boltzmann distribution, for example, mass and temperature determine the probability state (type) for each particle in an ideal gas, thereby determining the *natural distribution of the diversity of particle speeds* at thermodynamic equilibrium. The same capacity to predict probability seems to be true for the distribution of the diversity of complexity. Be it the population size of a city, the income of a household or the market value of a business, the diversity types for a system constitute its empirically defined range of probability states, along which its cases will distribute themselves, based on similarities in attributes.

Observation 5

The range of complex probability types can be discrete or continuous. Across all eight systems we examined, the natural organization of types was possible for both continuous and discrete distributions. We therefore stress the point, so that no confusion arises, that the distribution of the diversity of types along the x-axis for a complex system is possible with both continuous and discrete measures. In our study, for example, all of the systems examined, except for the phylogenetic tree and the prevalence of mental disorders, were based on a continuous measure of complexity. Either way—discrete or continuous—the cases in all eight systems were naturally grouped and ordered according to their complexity type.

Observation 6

The range of diversity types in a system can vary significantly. While it is true that all the systems we examined had a range of diversity types greater than or equal to $N = 7$ (the number of types for our prevalence of mental disorders example), the range of types varied considerably. For example, while the range of household incomes was significant, the genetic count for human diseases only ranged from $N = 1$ genes (which accounted for roughly $N = 870$ disease types) to roughly $N = 41$ genes (which accounted for only $N = 1$ disease type).

3.2. The Limiting Law of Restricted Diversity

Observation 7

The distribution of the diversity of complexity in skewed-right distributions appears to be governed by a limiting law, which results in a restriction of diversity. Across all eight systems studied, as well as the phylogenetic tree, we found that the distribution of the diversity of complexity was not uniform; that is, not all of the diversity types identified for each system were equiprobable. Nor did we find any Gaussian or skewed-left distributions. Instead, across all eight systems, the distribution of cases along the range of complex probability types was heavily skewed to the right, with or without long-tail, suggesting the possibility of a limiting law (See Figure 5).

Stated more formally, we found that, as the diversity of complexity $\rightarrow \infty$ (primarily in terms of the number of diversity types; but also, secondarily, in terms of the frequency of cases), a limiting law of restricted diversity emerges, which results in a skewed-right distribution, which decreases asymptotically as complexity increases (with or without long-tail). These results support the findings of Sornette [1] that the mechanics for skewed-right distributions may differ significantly from the mechanics for Gaussian normal distributions. In other words, a different limiting law (other than the central limit theorem) seems to be at play.

Observation 8

In terms of mechanics, *the limiting law of restricted diversity may be a countervailing force to the universal drive toward complexity.* In other words, in terms of mechanics, restricted diversity appears to be a naturally-

DOI 10.1002/cplx

occurring corrective to the drive toward increasing complexity; which limits or slows the growth of the diversity of complexity by constraining the majority of cases to the simplest complex probability types in a system.

As we discussed in the beginning of this study, across nature and various human-made systems, scientists have found that, despite the scale considered, there seems to be a drive toward increasing levels of complexity [11,38]. The historical mark of this drive is found in a variety of scientific taxonomies, as in the case of the phylogenetic tree (which we have repeatedly discussed), and also the numerous probability distributions that complexity scientists have fitted with the power-law [2,3],—hence Sornette's important point that probability distributions are the "first quantitative characteristics of complex systems." [1]

In response to these wild (skewed-right) probability distributions, the majority of scholars have focused on their long-tails and the upper bounds of complexity. [2–4] While the insights achieved through this focus (as in the case of power-laws and scale-free behavior and fractal self-similarity) have proved profoundly significant, they have masked, we believe, a universal limiting law, which resides (hidden) in the lower bound of these skewed-right distributions.

Restricted diversity reveals that, despite the drive toward an increasing diversity of complexity, there is a constraint on this drive, which appears near universal; and which does not have to do with the central limit theorem. We also conjecture that the law of restricted diversity may serve some type of optimization function on the diversity of complexity, be it evolutionary or thermodynamic in nature, for certain types of complex systems.

Observation 9

The drive toward complexity does not always evolve from less complex to more complex forms. Across all the systems we examined, there was nothing to suggest that the evolution of complexity (across time/space) always went from less to more complex probability types. For example, in terms of mental illness, there is no reason to assume that less complex forms of mental disorder preceded more complex types, or that, in terms of human health, less genetically complex diseases necessarily preceded more complex diseases. The same is true of income: wealthy households do not all start poor. Also, not all nodes in the world-wide-web necessarily evolved from simple to more complex forms of degree-connectedness. In other cases, however, as in the growth of cities, one does find "constructive complexity" at play (see section 2 for more on this term).

Observation 10

Restricted diversity appears to be scale-free. By the term "scale-free" we mean that restricted diversity is scale-invariant and self-similar across a wide variety of complex systems—from galaxies to the human diseasome map—even when its signature skewed-right distribution (a) is non-monotonic, or (b) is long-tailed or short-tailed.

Observation 11

The limiting law of restricted diversity produces a "majority of simplicity" phenomenon. Across all eight systems studied, we found that, like the phylogenetic tree of life, the majority of cases in a complex system cluster around the lower bound of the diversity of complexity; suggesting that, across a wide variety of natural and human-made systems, the simplest types of complexity are in the majority; and, overwhelmingly so—which takes us to our final theme.

3.3. The 60/40 Rule of Restricted Diversity

Observation 12

We found that, using C_c to measure the distribution of the diversity of complexity, it appears that restricted diversity obeys a 60/40 rule. In other words, it appears that, for those complex systems with skewed-right distributions, the clustering of the cases around the lower bound of the diversity of complexity is not arbitrary. Instead, similar to the central limit theorem for mild distributions, there is actually a consistently-followed rule *hidden in full view* within these distributions.

To make sense of the 60/40 rule, we created Figure 6, which is read as follows. First, there is the graph. Its purpose is to visualize the diversity contribution C_c of some set of complex probability types relative to the cumulative frequency (c) of cases, for each of the eight systems studied. The diversity contribution for the complex probability types C_c is shown on the x-axis; while the y-axis shows the cumulative percentage of cases (c*100) relative to the x-axis. When examining the curves, we noted that they all passed through the same shaded region on the graph, demonstrating that, for the systems we studied, 60% (or more) of cases consistently resided within the first 40% (or less) of a system's lower bound of complexity.

To explore further the specifics of these curves, we created the table in Figure 7, which provides the coordinates for a typical point in the shaded region, for each of the systems we examined. The reader can see that, for some systems—as in the case of the world-wide-web, the velocity spin of galaxies, and the city size by population—the cumulative percentage of cases was as high as 80/40. Still, while this insight is useful, the more important point is that all eight systems passed through the same region on the graph, suggesting that the limiting law of restricted diversity follows the 60/40 rule. F7

Observation 13

The 60/40 rule demonstrates that, for a variety of natural and human-made systems, the majority of cases

DOI 10.1002/cplx

FIGURE 7

Complex System	Coordinates for Each System in the 60/40 Region		
World Wide Web	80.86	40.16	325,729
2013 USA Household Income	59.32	41.03	122,953
Body Mass of Late Quaternary Mammals	62.30	42.56	4,916
Human Diseasome Complexity Map	67.80	29.10	1,248
Velocity of Galaxies in km/s	76.10	41.18	1,000
2011 USA Cities by Population Size	77.50	39.44	715
Financial Times 2014 Biggest 500 Companies	74.20	40.77	500
2012 Prevalence of Mental Disorders	68.50	34.32	3,199
	% Cases	% Complexity	N= total cases

Table of 60/40 rule for eight complex systems.

C O L O R

account for a small percentage of the total diversity of complexity. Such an insight, however, is not simply the inverse of the Pareto Principle, otherwise known as the oft misused and misunderstood 80/20 rule.

As discussed in section 2 of our study, the distribution of the diversity of complexity, as measured by C_c, is not the same as measuring the decay of a system using the power-law. The Pareto principle—which has been extensively documented (See [39])—only applies to the tail and not to the total distribution of diversity of complexity types in a system. The 60/40 rule, by contrast, concerns the distribution of the lower bound of equiprobable types relative to the cumulative distribution of complexity in the entire system—which is what Figure 6 shows. As such, when the 60/40 rule says above that "the majority of cases account for a small percentage of the total diversity of complexity," it means that, regardless of the distribution studied, 60% of the cases account for only 40% of the equiprobable types necessary to explain the total information in a system.

In the case of 2013 household income in the United States, for example, this translates as follows: 60% of all households constitute only 40% of all equiprobable income types; as such, the majority of households accounted for a small percent of the total diversity of economic complexity. In turn, the top 20% of U.S. households in 2013—which is the same group that the Pareto Principle focuses on; and which amounts to roughly $105,000 dollars or higher—accounted for 34% of the total diversity of equiprobable types. In other words, not only are the top 20% of U.S. household incomes the richest, they also constitute (as a group) a comparably higher percentage of the equiprobable diversity types necessary to explain the U.S. income system; which, to come full circle to the beginning of our study, means that not only do the least affluent possess a smallest percentage of the total wealth, they also constitute the greatest restriction of diversity. And, looking at Figure 7, a similar conclusion can be drawn about all the systems examined in our study: while the top 20% of cases constitute a sizable percentage of the total diversity of equiprobable types necessary to explain a system, the bottom 60% constitute a form of restricted diversity.

4. CONCLUSIONS

4.1. Implications

The limiting law of restricted diversity significantly advances the study of complex systems in several key ways:

1. First, its offers new empirical insights into the macroscopic regularities of skewed-right complex systems, as concerns the distribution of the diversity of their complexity, that otherwise would be difficult to detect with only the power-law or other such curve-fitting estimates.
2. More specifically, like the central limit theorem, restricted diversity appears to function as a universal limiting law for skewed-right complex systems, most likely operating as a countervailing response to the universal drive toward increasing complexity.
3. As a limiting law, restricted diversity suggests there is a natural order to the *distribution of the diversity*

DOI 10.1002/cplx

of complexity in a wide variety of natural and human-made skewed-right complex systems.

4. It also suggests that, even in highly complex skewed-right systems, this law results in a *majority of simplicity phenomenon*, as if there is a natural constraint or cap on the amount of diversity of complexity that can be managed or tolerated.
5. In terms of measuring this phenomenon, it appears that the 60/40 rule constitutes a region through which the distribution of the diversity of complexity must pass.
6. Finally, in terms of the distribution of equiprobable types, the 60/40 rule suggests that the majority of cases account for a small percentage of the total complexity in a system; that is, the greatest percentage of restricted diversity.

4.2. Limitations of Current Study and Future Directions

There are several limitations to the current study. First, it does not explore if the limiting law of restricted diversity and its corresponding 60/40 rule regulate the changing microscopic dynamics of cases. In other words, it needs to be determined if restricted diversity constrains the movement of cases up and down the continuum of diversity types, across time/space; as well as, if such a constraint obeys the 60/40 rule. Second, while the eight systems explored here have proved useful in suggesting the possibility of the universality of restricted diversity, it is necessary to examine further if and how restricted diversity works for a larger set of systems, as well as a variety of different probability distributions. For example, does restricted diversity apply to something as fundamental as the Maxwell–Boltzmann distribution for kinetic energy of an ideal gas at thermodynamic equilibrium? Or, how about the Bose–Einstein distribution or the Fermi–Dirac distribution? Finally, the current study is entirely descriptive, offering no explanation as to why restricted diversity, the majority of simplicity phenomenon, or the 60/40 rule take place. While we conjecture that they may all serve an optimization function, such a thing may turn out to be "system specific," due to the mechanics of different systems studied. As such, future research is needed.

ACKNOWLEDGMENTS

Acknowledgments: The authors want to thank the following colleagues at Kent State University: (1) Dean Susan Stocker, (2) Kevin Acierno and Michael Ball (Computer Services), and (3) the Complexity in Health and Infrastructure Group for their support.

REFERENCES

1. Sornette, D. Probability distributions in complex systems. In: Encyclopedia of Complexity and Systems Science; Meyers, R., Ed.; Springer, 2009; pp 7009–7024.
2. Barabasi, A.L. Scale-free networks: A decade and beyond. Science 2009, 325, 412.
3. Newman, M. Power laws, pareto distributions and zipf's law. Contemporary Phys 2005, 46, 323–351.
4. Clauset, A.; Shalizi, C.R.; Newman, M.E. Power-law distributions in empirical data. SIAM Rev 2009, 51, 661–703.
5. Yaeger, L.S. How evolution guides complexity. HFSP J 2009, 3, 328–339.
6. Carroll, S.B. Chance and necessity: The evolution of morphological complexity and diversity. Nature 2001, 409, 1102–1109.
7. Hidalgo, C.A.; Klinger, B.; Barabási, A.L.; Hausmann, R. The product space conditions the development of nations. Science 2007, 317, 482–487.
8. Rajaram, R.; Castellani, B. An entropy based measure for comparing distributions complexity. Phys A 2016, 453, 35–43.
9. Shannon, C. A mathematical theory of communication. Bell Syst Tech J 1948, 27
10. Contreras-Reyes, J.E. Rényi entropy and complexity measure for skew-gaussian distributions and related families. Phys A 2015, 433, 84–91.
11. Mitchell, M. Complexity: A Guided Tour; Oxford University Press, 2009.
12. Lui, L.T.; Terrazas, G.; Zenil, H.; Alexander, C.; Krasnogor, N. Complexity measurement based on information theory and kolmogorov complexity. Artif Life 2015, 21, 205–224.
13. Shalizi, C.R. Methods and techniques of complex systems science: An overview. In: Complex Systems Science in Biomedicine; Springer, 2006; , pp 33–114.
14. Adami, C. What is complexity? BioEssays 200 1085–1094.
15. Lopez-Ruiz, R.; Mancini, H.; Calbet, X. A statistical measure of co xity. arXiv 2002, preprint nlin/0205033
16. McCabe, T.J. A complexity measure. IEEE Trans Software Eng 1976 308–320.
17. Rojdestvenski, I.; Cottam, M.; Oquist, G.; Huner, N. Thermodynamics of complexity. Phys A 2003, 320, 318–328.
18. Yamano, T. A statistical measure of complexity with nonextensive entropy. Phys A 2004, 340, 131–137.
19. Lloyd, S. Measures of complexity: A nonexhaustive list. IEEE Control Syst Mag 2001, 21, 7–8.
20. Smith, F.A.; Brown, J.H.; Haskell, J.P.; Alroy, J.; Charnov, E.L Dayan, T., et. al. B.J.E.: Similarity of mammalian body size across the taxonomic hierarchy and across space and time. Am Nat 2004, 163, 672–691.
21. Smith, F.A.; Lyons, S.; Ernest, S.; Jones, K.; Kaufman, D.; Dayan, T.; Marquet, P.; Brown, J.; Haskell, J. Body mass of late quaternary mammals (updated version of data obtained from senior author). Ecology 2003, 84, 3402.
22. West, G.B.H.B.J.; Enquist, B.J. A general model for the origin of allometric scaling laws in biology. Science 1997, 276, 122–126.

DOI 10.1002/cplx

23. Goh, K.I.; Cusick, M.E.; Valle, D.; Childs, B.; Vidal, M.; Barabási, A.L. The human disease network. Proc Natl Acad Sci USA 2007, 104, 8685–8690.
24. Batty, M. Cities, prosperity, and the importance of being large. Environ Plann B Plann Des 2011, 38, 385–387.
25. Eeckhout, J. Gibrat's law for (all) cities. Am Economic Rev 2004, 94, 1429–1451.
26. Benguigui, L.; Blumenfeld-Lieberthal, E. A dynamic model for city size distribution beyond zipf's law. Phys A 2007, 384, 613–627.
27. Haynes, P: Managing Complexity in the Public Services; Routledge, 2015.
28. Berry, T.K.; Bizjak, J.M.; Lemmon, M.L.; Naveen, L. Organizational complexity and ceo labor markets: Evidence from diversified firms. J Corporate Finance 2006, 12, 797–817.
29. Hidalgo, C.A.; Hausmann, R. The building blocks of economic complexity. Proc Natl Acad Sci USA 2009, 106, 10570–10575.
30. Rosser, J.B. Econophysics and economic complexity. Adv Complex Syst 2008, 11, 745–760.
31. Kessler, R.; Demler, W.T.C.O.; Walters, E. Prevalence, severity, and comorbidity of 12-month dsm-iv disorders in the national comorbidity survey replication. Arch Gen Psychiatry 2005, 62, 617–627.
32. Hill, M. Diversity and evenness: A unifying notation and its consequences. Ecology 1973, 54, 427–432. DOI doi:10.2307/1934352
33. Jost, L. Entropy and diversity. Oikos 2006, 113, 363–375.
34. Leinster, T.; Cobbold, C.A. Measuring diversity: The importance of species similarity. Ecology 2012, 93, 477–489.
35. Beck, J.; Schwanghart, W. Comparing measures of species diversity from incomplete inventories: An update. Methods Ecol Evol 2010, 1, 38–44.
36. MacArthur, R. Patterns of species diversity. Biol Rev 1965, 40, 510–533.
37. Peet, R. The measurement of species diversity. Annu Rev Ecol Syst 1974, 5, 285–307. DOI doi:10.1146/annurev.es.05.110174.001441
38. Capra, F., Luisi, P.L. The Systems View of Life: A Unifying Vision; Cambridge University Press, 2014.
39. Reed, W.J. The pareto law of incomesan explanation and an extension. Phys A 2003, 319, 469–486.

DOI 10.1002/cplx

AQ1: Please confirm whether affiliation is OK as typeset.

AQ2: Please confirm whether corresponding author information is OK as typeset.

AQ3: Please note that Refs. [9 and 10] and [17 and 18] are duplicate references, hence deleted and renumbered accordingly. Please check.

AQ4: Please provide page range for Ref. 9.

AQ5: Please provide place of publication for ALL book-type references.

AQ6: Please provide complete details for Ref. 15.

AQ7: Please provide volume number for Ref. 16.

AQ8: Please provide complete list of authors for ALL et al.-type references.

AQ9: Please confirm that given names (red) and surnames/family names (green) have been identified correctly.