Review

Scaling laws in cognitive sciences

Christopher T. Kello[1], Gordon D.A. Brown[2], Ramon Ferrer-i-Cancho[3], John G. Holden[4], Klaus Linkenkaer-Hansen[5], Theo Rhodes[1] and Guy C. Van Orden[4]

[1] Cognitive and Information Sciences University of California , Merced, 5200 North Lake Rd., Merced, CA 95343, USA
[2] Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom
[3] Department de Llenguatges i Sistemes Informatics, Universitat Politecnica de Catalunya, Campus Nord, Edifici Omega, Jordi Girona Salgado 1-3, 08034 Barcelona, Catalonia, Spain
[4] Center for Perception, Action and Cognition, Department of Psychology, University of Cincinnati, PO Box 210376, Cincinnati, OH 45221-0376, USA
[5] Department of Integrative Neurophysiology, VU University Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, the Netherlands

Scaling laws are ubiquitous in nature, and they pervade neural, behavioral and linguistic activities. A scaling law suggests the existence of processes or patterns that are repeated across scales of analysis. Although the variables that express a scaling law can vary from one type of activity to the next, the recurrence of scaling laws across so many different systems has prompted a search for unifying principles. In biological systems, scaling laws can reflect adaptive processes of various types and are often linked to complex systems poised near critical points. The same is true for perception, memory, language and other cognitive phenomena. Findings of scaling laws in cognitive science are indicative of scaling invariance in cognitive mechanisms and multiplicative interactions among interdependent components of cognition.

The scaling law debate

In the past, the ubiquity of the normal curve was observed throughout nature, but not satisfactorily explained. Then developments such as the central limit theorem showed how random, independent effects combine to produce the normal curve, thereby explaining its ubiquity. Today the normal curve is sometimes taken for granted, although still appreciated for the beauty and power with which it brings order to randomness.

The normal curve fails, however, to describe crucial facts about living systems and other complex systems – because such systems are more than collections of random, independent effects. Their complexity is defined by intricate regularities and dependencies that span multiple temporal and spatial scales of analysis. For instance, synchronization errors in a finger-tapping experiment follow the normal distribution, yet the temporal sequence of errors is highly non-random [1]. In other words, measurements of living systems often obey *scaling laws* rather than linear relations or Gaussian statistics. Bringing order to such regularities, which are inherent in nature's complexities, including the complexities of cognition, has proven to be as difficult as bringing order to randomness.

Most generally, scaling laws express one variable as a nonlinear function of another raised to a power, $f(x) \propto x^{\alpha}$, with $\alpha \neq 0$. Scaling laws are observed throughout the sciences, notwithstanding difficulties in determining whether measurements actually conform to scaling laws. Other functions (such as exponentials) can also provide good fits to data, and skeptics sometimes contend that scaling laws do not always provide the best fit [2,3]. However, improved statistical tests have provided strong evidence of pervasive scaling laws [4–6] over a substantial (although limited) range of scales [7–10].

Even among scientists who acknowledge the existence of scaling laws, some still see them as largely uninformative because there are many ways to produce scaling laws, and some of those ways are idiosyncratic or artifactual [11]. Thus, their unifying order could be more illusory than enlightening. However, as the extent of unexplained coincidence grows with each reported power law, coincidence becomes increasingly difficult to accept. We could instead seriously consider the hypothesis that scaling laws describe a fundamental order in living and complex systems. This working perspective motivates principles and theories to explain scaling laws in terms that can cross or integrate disciplines.

Although these debates over scaling laws have a long history in nearly every scientific discipline and domain, they have emerged only recently in cognitive science. Indeed, many cognitive scientists are yet unfamiliar with the debate, or the pervasiveness and meaning of scaling laws in other sciences. Here, we review evidence of scaling laws in cognitive science, at neural, behavioral and linguistic levels of description. The evidence indicates that cognitive phenomena occurring at relatively small temporal and spatial scales are intimately linked to those occurring at relatively large scales. This linkage can be explained by rooting cognitive functions in principles of statistical physics.

Scaling laws in perception, action and memory

Although the demonstration of the pervasiveness of scaling laws could be new to cognitive science, a few classic examples are well known in the field. Studies of psychophysics and motor control, in particular, have produced some of the most lawful phenomena of human behavior, including some allometric scaling laws (see Glossary). Stevens' law is one psychophysical example for which the physical magnitude of a stimulus (S) is proportional to its perceived intensity (I) raised to a power α, $S \propto I^{\alpha}$ [12].

Corresponding author: Kello, C.T. (ckello@ucmerced.edu).

1364-6613/$ – see front matter © 2010 Elsevier Ltd. All rights reserved. doi:10.1016/j.tics.2010.02.005 Available online 1 April 2010

Glossary

Allometric scaling laws: traditionally refer to relationships between different measures of anatomy and/or physiology that hold true across species or organs of different sizes. A classic example is Kleiber's law that relates organism mass to metabolic rate as $m{\sim}r^{0.75}$, and holds true across species ranging from bacteria to whales [75]. The mass of gray matter versus white matter in brains also obeys an allometric scaling law across species [76].

Criticality: refers to the state of dynamical systems poised near phase transitions, and it is characterized by scale invariance, for example power-law temporal fluctuations (1/f scaling) and power-law distributions with critical exponents. The Ising model of ferromagnetism is a classic example of a system with a critical point between ordered and disordered phases [77]. Self-organized criticality refers to models such as the ''sandpile'' model for which the critical point is an attractor [78].

Heavy-tailed distributions: have tails that decay more slowly than exponentially. All power-law distributions are heavy-tailed, but not all heavy-tailed distributions are power laws (e.g. the lognormal distribution is heavy-tailed but is not a power-law distribution).

Lévy flights: are random walks (i.e. flights) for which each step is drawn from a power-law distribution (the direction of each step is typically random but can instead be determined by some rule or algorithm). Points visited by steps in Lévy flights tend to be clustered in space, where clusters are separated by very large steps occasionally drawn from the heavy tail of the distribution.

Lognormal distributions: are heavy-tailed and have probability density functions that are normally distributed under a logarithmic transformation:

$$P(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \quad \text{for } x > 0.$$

For certain parameter values, lognormal distributions can be difficult to distinguish from power law distributions [79].

Metastability: is a delicate type of stability that is a property of systems poised near their critical points. It stems from the fact that small (microscopic) perturbations to near-critical systems can result in system-wide (macroscopic) changes in their states. Thus, states are only tenuously stable, resulting in many (nearly) equally potential states near critical points.

Pareto distributions: are one type of power-law distribution used to model phenomena from a diversity of fields, including economics, physics, anthropology, computer science, geology and biology. Its probability density function is

$$P(x) = \alpha\left(\frac{x_{\min}{}^{\alpha}}{x^{\alpha+1}}\right) \quad \text{for } x > x_{\min},$$

where $x_{\min}$ is used to express lower bounds that often exist on physical quantities (e.g. volumes and masses of particles must be >0).

Power-law distributions: have heavy-tailed probability functions of the form $P(x){\sim}x^{-\alpha}$, where typically $0<\alpha<3$. These distributions have properties of self-similarity and scale invariance.

''Rich get richer'': refers to a growth process (i.e. Yule process) whereby the probability of incrementing some quantity associated with a given unit (e.g. population of a city, frequency of a word) is proportional to its current value. Such quantities grow to be power-law distributed, for example networks that grow by *preferential attachment* have power-law distributed links (i.e. scale-free [80]).

Scale-free networks: are those with power-law distributions of links per node (i.e. node degree). The heavy tail means that some nodes act as hubs because they are linked to a substantial proportion of all nodes in the network.

1/f Scaling: (also known as 1/f noise, pink noise or flicker noise) occurs in time series with Long-range temporal correlations. A time series can be correlated with itself (i.e. autocorrelated) at varying temporal lags k, and autocorrelations, $C(k)$, are typically long-range if they decay slowly as an inverse power of lag, $C(k){\sim}k^{-\alpha}$. Expressing this power law in the frequency domain yields $S(f){\sim}f^{-\alpha}$, where f is frequency, $S(f)$ is spectral power and $\alpha{\approx}1$ for 1/f scaling.

Self-similarity and scale invariance: both refer to objects or mathematical functions that exhibit similar shapes or relations among variables at different scales. A self-similar object is such that each portion can be considered a reduced-scale image of the whole. Mathematical fractals such as the Koch snowflake are examples of ideal self-similarity objects because their contours are recursively and identically repeated across all scales. A power-law distribution is scale invariant because multiplying x by a constant c only scales the function: $P(cx)=c^{\alpha}P(x)$, where $P(x)=x^{\alpha}$. Scale invariance in nature tends to be approximate and statistical, as illustrated by the coastline of Britain: it is not that a particular contour is repeated exactly across scales of the coastline. Instead, the statistical relation between measured length of the coastline and size of the measuring stick is invariant across different zoom levels [81].

Zipf's law: refers to a power-law distribution traditionally expressed in terms of frequencies of occurrence of a certain variable (e.g. word rank or population size of cities). When word frequencies are said to follow Zipf's law, their rank r (the most frequent word has rank 1, the second most frequent word has rank 2, and so on) is related to frequency as $f(r){\sim}r^{-\alpha}$.

With regard to motor control, Lacquaniti *et al.* [13] discovered a two-thirds power law in which angular velocity (A) is proportional to the curvature (C) of drawing movements, $A \propto C^{\sim .66}$.

Traditionally, such laws have been investigated independently of each other, especially when they are found in different research domains (e.g. perception versus action for the examples above). It is commonly assumed in cognitive science that different domains entail different mechanisms. However, common principles can underlie seemingly disparate mechanisms, and similar scaling laws are suggestive of such principles. Most broadly, the property of scale invariance inherent to the Stevens' and two-thirds laws (and all scaling laws) implies a property or principle that is adaptive at all scales (i.e. scaling laws can exist because natural selection or other mechanisms of adaptation select and repeat a pattern or process across scales). Consistent with this implication, Copelli *et al.* [14–16] hypothesized that Stevens' law reflects maximization of sensitivity and dynamic range in sensory systems, and Harris and Wolpert [17] hypothesized that the two-thirds law reflects minimization of movement errors caused by noise in motor systems.

A compelling fact about the Stevens' and two-thirds laws (which is also true of many other scaling law observations) is that data closely follow their power-law functions over more than three orders of magnitude. For instance, different muscles and muscle groupings are employed and coordinated for movements of very small versus very large curvature, yet all obey the two-thirds law. Scaling over multiple orders of magnitude is compelling because it ties together ostensibly different mechanisms at disparate scales. Given evidence for several other scaling laws in perception and action [18,19], one is led to principles that generally tie together perceptual and motor mechanisms across scales.

The purview of scaling laws broadens as we further consider their occurrence in other domains of cognitive function. Memory is a natural domain to consider after perception and action, and indeed scale invariance has been found in memory retrieval (Figure 1). Maylor *et al.* [20] instructed participants to recall what they did (or will do) in the previous (or next) day, week or year. Rate of item recall was generally invariant across target recall period: on average, participants recalled an invariant five items/min regardless of the span over which recall was bounded.

This dynamic scaling of memory retrieval is consistent with a scale-invariant temporal ratio model of memory [21,22], in which discriminability of memories depends on ratios of temporal intervals between encoding and retrieval events. For example, two different memory traces encoded 8 versus 10 min in the past (temporal ratio=8/10) will be as confusable with each other as two traces encoded 8 versus 10 h in the past. This model explicitly ties scales together because the same memory and retrieval processes are hypothesized to operate across all timescales and no qualitative distinction exists between short-term and long-term memory processes. Such an approach appears necessary to explain the many scale-similar effects in human memory and learning (for analogous effects in animal learning, see Ref. [23]).

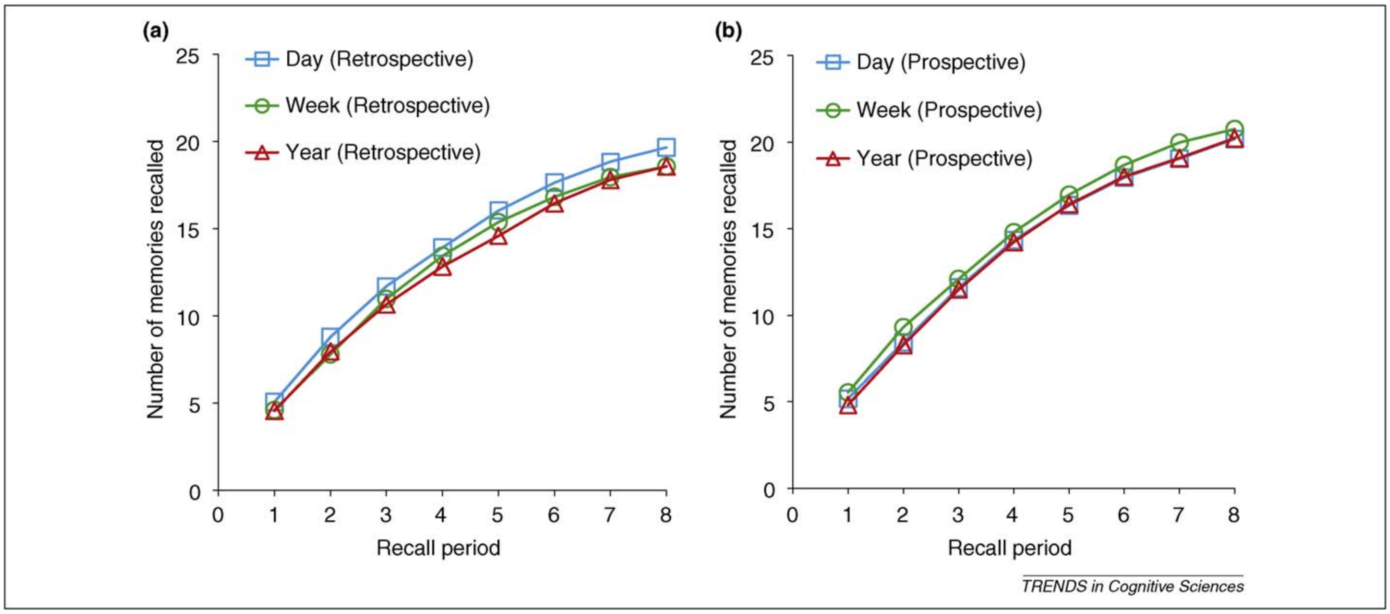

Figure 1. Scaling in retrospective and prospective memory. Recall data are plotted showing scaling in the retrieval of retrospective **(a)** and prospective **(b)** memories from periods varying from a day to a year [20]. Participants were given 4 min to recall ''. . .jobs, appointments, and things you have done yesterday/in the last week/in the last year'' (retrospective) or ''. . .jobs, appointments, and things you intend to do tomorrow/in the next week/in the next year'' (prospective) as one-word summaries. The figure shows that the cumulative number of items recalled (on the *y*-axis) by the end of each of eight 30-s recall periods (on the *x*-axis). These did not vary as a function of the time interval from which recall was permitted (day/week/year) – recall rate was timescale invariant.

Another type of scaling law in memory comes from a classic free recall paradigm, yet was only recently discovered by drawing an analogy to studies of animal foraging behaviors [24]. Birds, monkeys, fish and numerous other species have been reported to search for food in Lévy flight patterns [25], which have been hypothesized as effective search strategies because they cover more territory than, for example, a random walk with normally distributed steps [26]. Searching for items or events in memory is like foraging, particularly in tasks such as free recall of members of a given semantic category (e.g. animals) in a given time period [27]. Rhodes and Turvey [24] analyzed inter-response time intervals (IRIs) from this classic memory task, which are analogous to steps from one recalled item to the next. The authors found IRIs to be power-law distributed with exponents very similar to those found in animal foraging (Figure 2). These comparable results suggest that Lévy flights are generally adaptive across a variety of search ecologies. These results also illustrate how scaling laws can lurk unnoticed in data for decades, in the absence of theories and analytic techniques necessary to recognize them.

Scaling laws in reaction times and word frequencies

Another "lurking" scaling law was recently discovered in the distributions of word-naming latencies of individual readers [28]. Cognitive psychologists have known for decades that reaction time (RT) distributions tend to be positively skewed, but usually this skew has been treated as mere deviation from normality; indeed, very long RTs are typically considered outliers and hence are removed or truncated. Extreme values are expected, however, if RTs are drawn from heavy-tailed distributions, rather than Gaussian distributions. Lognormal and power-law distributions are heavy-tailed, and naming latencies (as well as other response times) appear to be best modeled as mixtures of lognormal and power-law distributions (Figure 3). These heavy-tailed distributions remain contentious nevertheless because they are difficult to reconcile with traditional theories of RTs based on additive interactions among component processes.

This debate has only just begun for RTs, but for a different power-law distribution of linguistic behavior, it has been ongoing for over 50 years. In his pioneering work, G. K. Zipf [29] studied the inverse power law of word usage that bears his name (Figure 3). Zipf's law as originally formulated states that the frequency of a word (f) in a given corpus is proportional to the inverse of its frequency rank (r), $f \propto \frac{1}{r}$. Zipf's law is apparently a universal property of human language, yet its origins remain controversial. Power laws such as Zipf's law are found not just in word usage but in many aspects of language, such as syntactic dependency networks [30], and letter sequences in lexicons [31].

Zipf originally explained his law in terms of a principle of least effort, which states that language structure and language use minimize both speakers' and listeners' efforts. Speakers prefer high-frequency words for ease of memory recall, and listeners prefer low-frequency words with unambiguous meanings. Zipf hypothesized that his law reflects a compromise between these competing constraints on communication. The same basic principle can also be applied at other linguistic scales, which would explain Zipf's law as an adaptive property of communication.

Some researchers, however, claim that Zipf's law is inevitable (and therefore uninteresting) because randomly generated letter sequences can also exhibit scaling [32]. Numerous and recent analyses have refuted this claim by showing that random letter sequences do not mimic closely

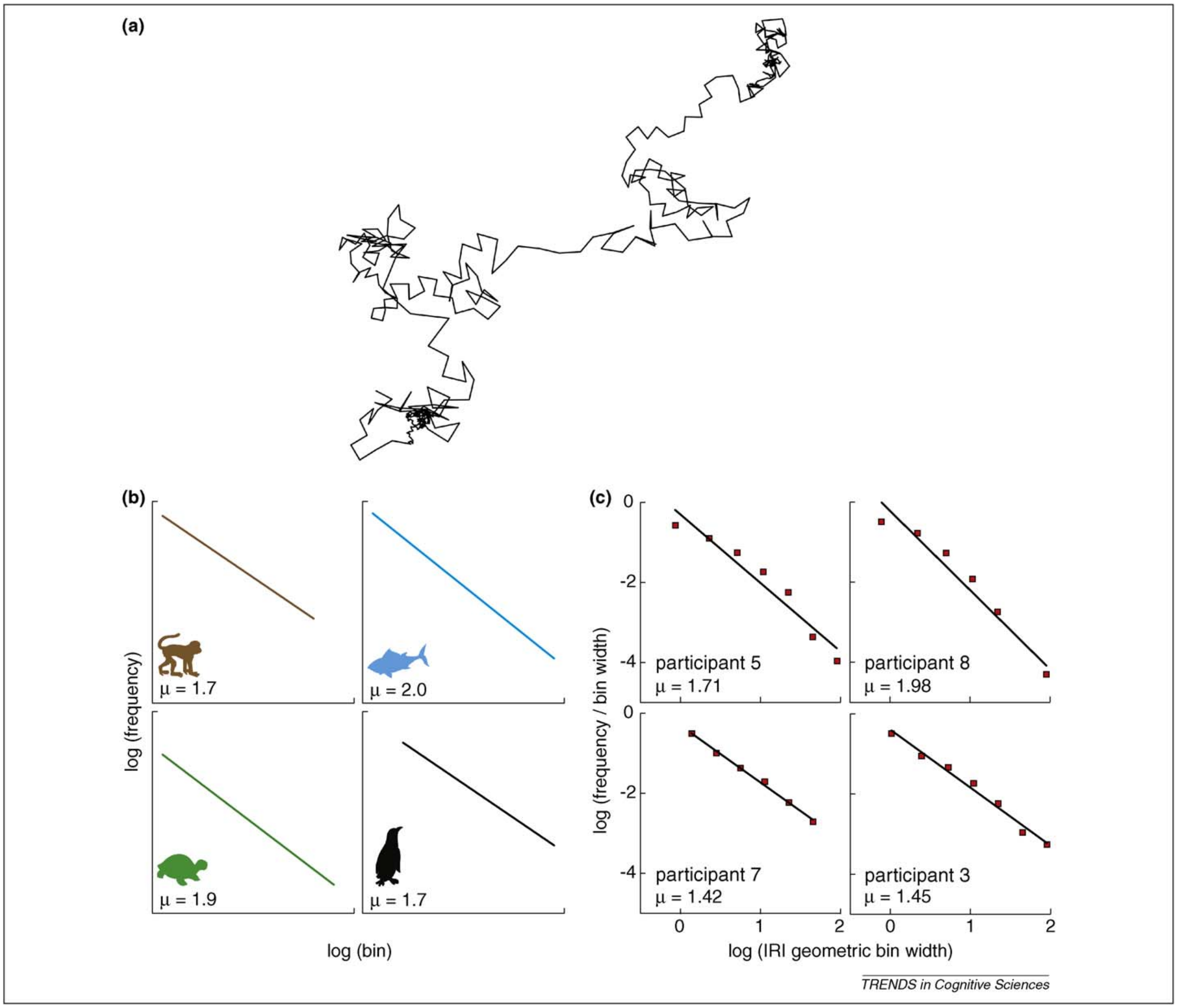

Figure 2. Lévy flights in animal and memory foraging. An artificially generated Lévy flight path is shown in two dimensions in **(a)** (note the clusters and clusters within clusters). In **(b)**, estimated Lévy flight power-law exponents [9,83] are graphed as straight lines in log–log coordinates for four different species: **(i)** spider monkey; **(ii)** bigeye tuna; **(iii)** leatherback turtle; and **(iv)** Magellanic penguin. Analogous histograms are shown in **(c)** for four representative participants in a category member free recall task [24]. The histograms are of inter-response intervals (IRIs) between successive category member recalls.

enough the actual shape of Zipf's law in real languages; Zipf's law is universally and nontrivially descriptive of language use [33–36]. Although this important law remains unaddressed in most linguistic theories, it has been hypothesized to underlie two fundamental properties of human language: syntax and symbolic reference [37]. Once a communication system organizes the patterning of word frequency according to Zipf's law, a rudimentary form of language emerges for free as a side effect of competing constraints of communication [38].

It has been argued that statistical laws of language might be interconnected [39], and these interconnections appear to include scaling laws. For instance, the power law of word connectivity in syntactic dependency networks could be a natural consequence of Zipf's law for word frequencies [30,37]. Traditional research on typology of linguistic universals has focused on sentence level phenomena such as word order [40]. In contrast, scaling laws offer a new source of linguistic universals at the level of the large scale organization of language and also offer the possibility to integrate linguistics and cognitive science.

Scaling laws and criticality

Widely reported evidence of scaling laws calls for cognitive and linguistic theories that explain their ubiquity [41,42]. As a starting point for providing an explanation of the ubiquitous presence of scaling laws, a key alternative to additive summations of components is multiplicative interactions which produce heavy-tailed distributions [28,43]. Multiplicative interactions in cognition can be expressed when the operation of one component depends on the state of another, which is often expressed empirically as interaction effects. The preponderance of such effects in

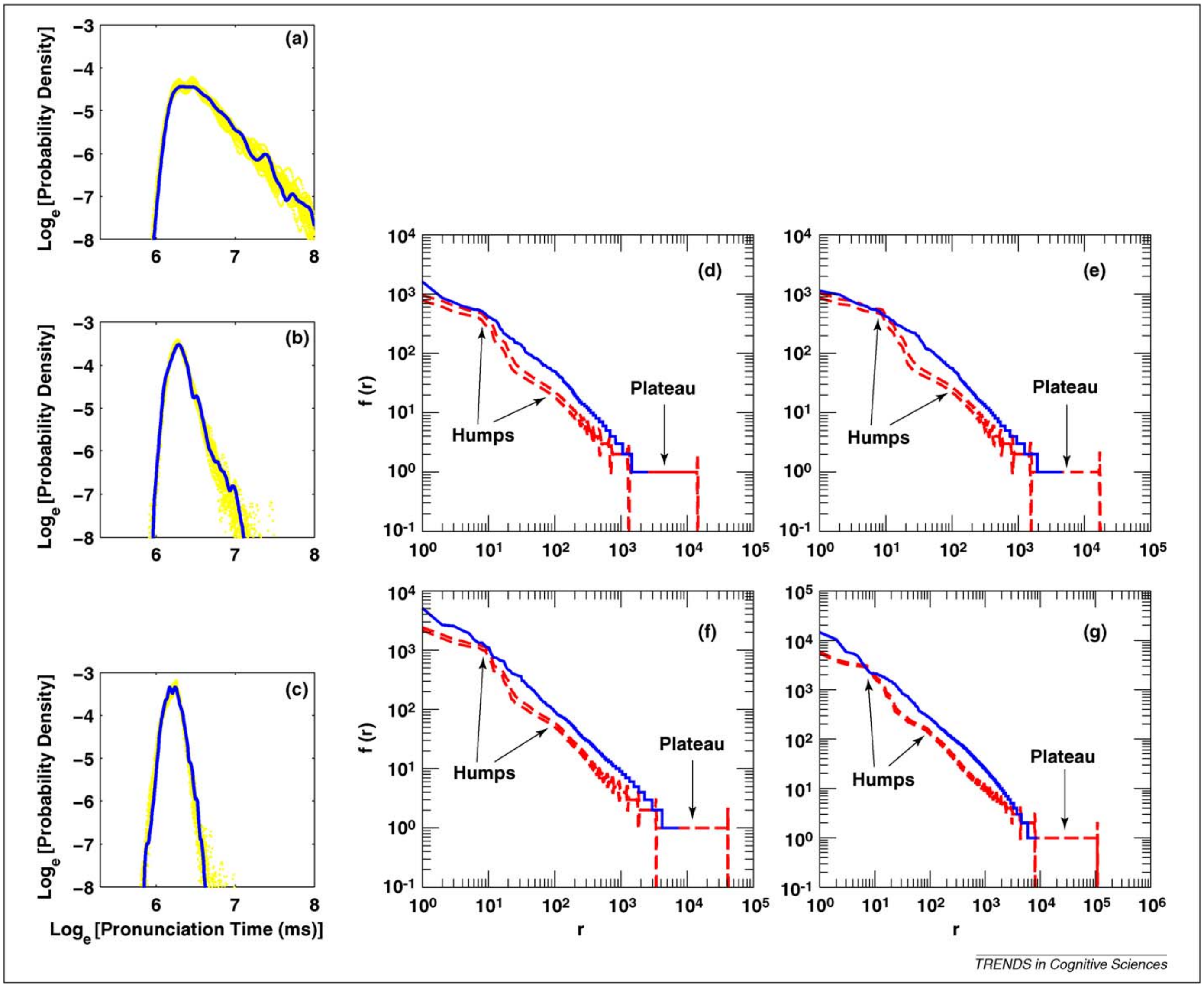

Figure 3. Power-law distributions of word naming latencies and word frequencies. Distributions of speeded word naming latencies (milliseconds, ms) for three representative readers are shown in the left column, in log–log coordinates (reproduced, with permission from [28]). Heavy blue lines are observed distributions and yellow lines are mixtures of ideal lognormal and inverse power-law distributions falling with 90% confidence intervals of observed distributions. Some readers exhibited heavy tails and hence greater power-law proportions **(a)**, others were more balanced **(b)**, and still others were predominantly lognormal **(c)**. Plots on the right show inverse power law distributions of word ranks (reproduced from Ref. [33]). Words were counted and ranked by frequency count in four different texts (black lines): Alice's Adventures in Wonderland **(d)**; Hamlet **(e)**; David Crockett **(f)**; and The Origin of Species **(g)**. Rank distributions (blue lines) are compared with those generated by a random text model in which letters and spaces were sequentially sampled according to their probabilities in real texts (red dashed lines). The random text model does not match observations of Zipf's law because **(i)** observations fall outside ± 3 standard deviations of the random text model; **(ii)** random texts have relative humps in the higher frequencies and wider plateaus in the lower frequencies; and (iii) the rank histogram of random texts extends well beyond that of real texts.

cognitive phenomena suggests a system in which multiplicative interaction is the rule [41], and the exception is linear combinations of component effects amenable to linear decomposition (e.g. additive and subtractive logic). Systems dominated by multiplicative interactions are known to produce heavy-tailed distributions (Box 1).

As a rule, multiplicative interactions also create interdependencies among component activities over time. These interdependencies can lead to long-range correlations when component effects travel through feedback loops across scales of component interactions, thereby changing the dynamics of interactions [44]. Interdependence has been shown in model systems to generate self-similar structures and fluctuations [45] and thereby generate spatial and temporal long-range correlations, as well as power-law distributions. Box 2 uses the Ising model as an illustrative example, one that stands as a pillar of statistical physics, and a good starting point for cognitive scientists interested in investigating scaling laws as emergent from multiplicative, interdependent components of cognition.

In statistical physics, scaling laws have been studied for decades in the context of phase transitions [46,47]. When systems are poised near order–disorder phase transitions (i.e. critical points), microscopic changes can propagate through spatial correlations across many scales to become macroscopic effects that evolve on many time scales. Thus, criticality yields multiple-scale dynamics expressed as

Box 1. Additive versus multiplicative effects

When measurements are independent and measured values are essentially sums of independent effects, the central limit theorem leads one to expect a normal distribution of values (i.e. a Gaussian probability function; Figure I blue). Illustrative examples are distributions of organism size in a population, such as height or weight, and distributions of scores on various tests of cognitive ability, such as the IQ test. Each observation of size or IQ is independent of other observations, and although factors affecting these measures are myriad and poorly understood, they are assumed to make largely independent and additive contributions to each individual's size or IQ.

Normal distributions are not expected when measured values reflect multiplicative combinations of effects. An illustrative example is distributions of city population sizes. Cities appear to grow multiplicatively (i.e. bigger cities are more likely to have larger growth rates than smaller cities [82]). The consequence is that city populations appear to be power-law distributed over a wide range of sizes [10,80]. Multiplicative effects can also lead to lognormal distributions (Figure I red), and simple multiplicative models have been shown to generate either lognormal or power-law distributions depending on small parametric changes [79]. Lognormal and power-law distributions are both heavy-tailed, and hence heavy-tailed distributions are often interpreted as evidence for multiplicative processes. An important difference between heavy-tailed and normal distributions is that moments of the former (e.g. mean and variance) poorly characterize the distribution (in fact, they are undefined for certain power-law distributions).

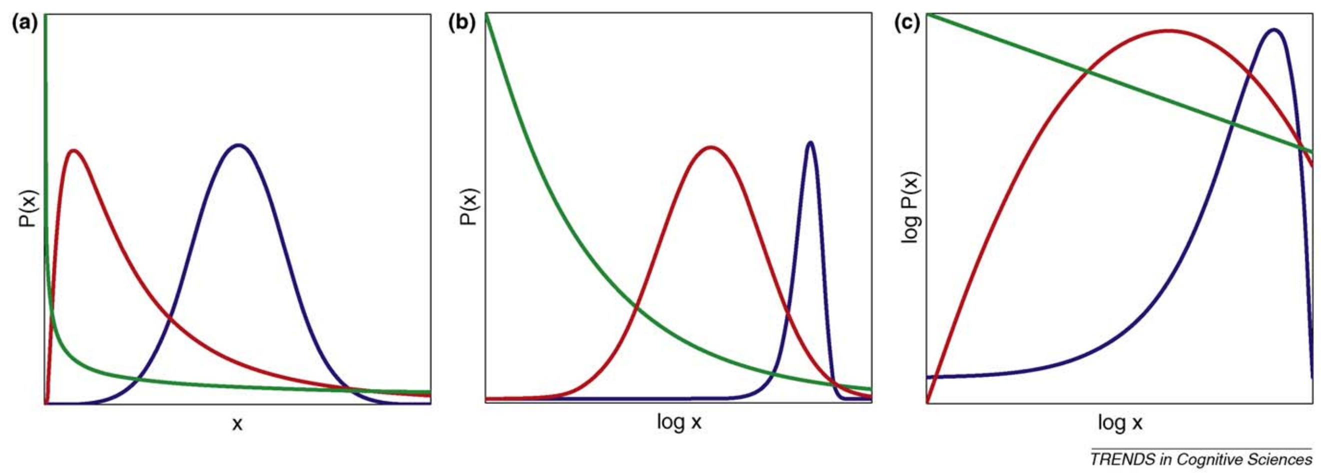

Figure I. Idealized normal (blue), lognormal (red), and power law (green) probability functions are plotted in raw (left), semi-log (middle) and log–log (right) coordinates.

spatial and temporal long-range correlations [48]. Evidence for criticality has been investigated in a wide variety of physical, biological, computational and social systems [49].

The possible role of criticality in cognitive science can be illustrated through neural networks [50–53]. A fundamental requirement of any neural network is to transmit and process information via activities of its neuronal components. On timescales of milliseconds to seconds and even minutes, information is transmitted in neural networks via action potentials (i.e. spikes). Regardless of how information is coded in spikes, neurons must be able to affect each others' spiking dynamics to transmit and process information. Thus, if neurons are too independent of each other, information cannot be transmitted. But if neurons are too interdependent, their spiking dynamics will be slaved to each other, and hence too uniform and unchangeable to code information. Criticality strikes a balance between independence versus interdependence among component activities, and when applied to neural spiking dynamics, it could support information transmission and processing. More generally, evolution can favor critical states because their associated metastability (i.e. the delicate stability that characterizes systems poised near their critical points) strikes an optimal compromise between change (flexibility and adaptation) and stability (memory and continuity) necessary for information transmission and computation [47,53–58].

A connection between criticality, metastability and computation was first proposed for cellular automata [59,60] and has since been demonstrated in the dynamics of neural networks [61]. Beggs and Plenz [62] found that cortical slice preparations exhibit critical branching dynamics (i.e. "neural avalanches"), and probabilistic spiking models were shown to optimize information transmission near their critical points. Neurons can also be modeled by thresholding sums of incoming weights to be above or below zero (+1, –1), and networks of threshold neurons have similarly been shown to optimize memory and representational capacity near their critical points [56], and psychophysical models have linked Stevens' law with criticality in neural network dynamics [14–16]. At a different scale, Zipf's law was recently derived from a critical point between speaker and listener efforts quantified in information theoretic terms [63–65]. Taken together, these models realize testable connections between criticality and cognition as expressed in neural and behavioral activity.

Evidence for criticality in cognitive science has also come in the form of temporal long-range correlations (i.e. 1/f scaling), which can be seen in fluctuation time series as undulations at many timescales. 1/f scaling has been observed in many aspects of neural and behavioral activity [6]. For instance, 1/f scaling has been observed in acoustic energy fluctuations across word repetitions and in fluctuations of the amplitude envelope of ongoing neuronal oscillations in healthy subjects (Figure 4). The amplitude

Box 2. Short-range versus long-range correlations

In physical systems, events occurring nearby in time or space are often similar to each other, and such similarities typically fall off as distance increases. Physicists use the *correlation function* to express the effect of distance on similarity, and the observed shape of this function constitutes evidence about the type of system being observed.

To illustrate we use a characterization of the Ising model [77]. Imagine a 2D grid of lights of varying brightness (from off to maximum), where brightness is a function of two variables. One is a random noise factor (individual to each light) and the other is a neighbor conformity factor whereby each light tends towards the brightness of its four nearest neighbors on the grid. These two variables are weighted together to determine the brightness of each light. In this illustration, the correlation function measures the degree to which lights have equal brightness levels as a function of their distance apart on the grid. If noise is heavily weighted, then brightness levels are independent across lights and the correlation function will be near zero for all distances >0. If instead neighbor conformity is heavily weighted, then brightness levels will be interdependent and approach uniformity, with a correlation function near one across a wide range of distances.

Neither extreme is typical of physical systems. Instead, component interactions are somewhere between independent and interdependent. Weak interactions can result in short-range correlations (Figure I green) that decay exponentially with distance. Stronger interactions can result in long-range correlations that decay more slowly (Figure I pink) (i.e. as an inverse power of distance). The correlation function can also be defined for distances in time, with an analogous comparison between weak (short-range) versus strong (long-range) interactions. No interactions can result in uncorrelated noise (Figure I grey), and integrating over uncorrelated noise results in a random walk (Figure I brown).

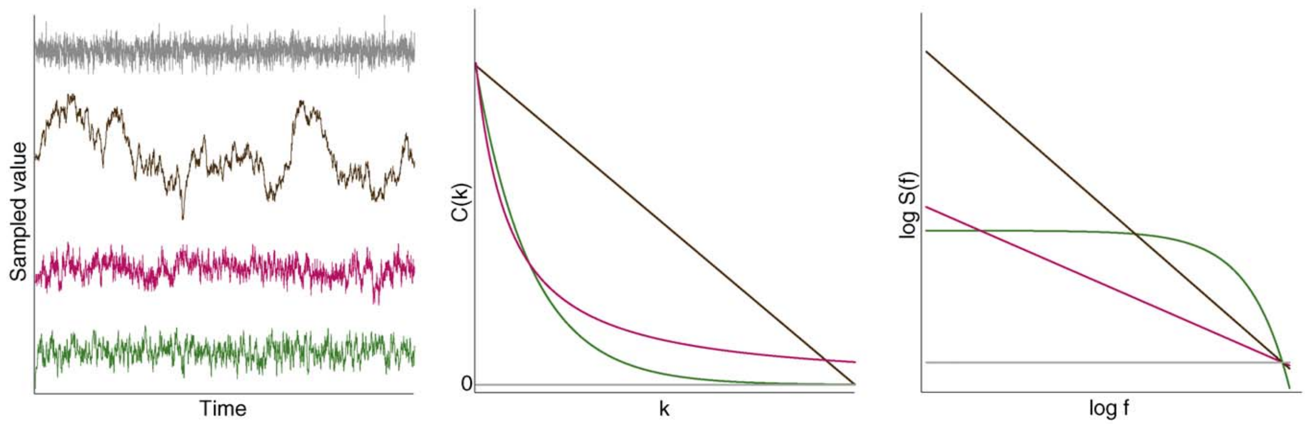

Figure I. Four example time series are plotted in the left-hand panel: random samples from a normal distribution with zero mean and unit variance (i.e. white noise, in grey), a running sum of white noise (i.e. brown noise, also known as a random walk, in brown), 1/f noise (i.e. pink noise, in pink) and an autoregressive moving average (ARMA, in green), where each sampled value is a weighted sum of a noise sample, plus the previous noise value, plus the previous sampled value. Idealized autocorrelation functions are shown in the middle panel for each of the time series, where *k* is distance in time. Note that white noise (i.e. pure independence) has no correlations, ARMA has short-range correlations that decay exponentially with *k*, 1/f noise has long-range correlations that decay as an inverse power of *k* and brown noise has correlations that decrease linearly with *k*. Idealized spectral density functions (where *f* is frequency and *S*(*f*) is spectral power) are shown in the right-hand panel in log–log coordinates. White, pink and brown noises correspond to straight lines with slopes of 0, −1 and −2, whereas ARMA plateaus in the lower frequencies.

envelope can be illustrated by tracing a line from peak to peak along a given waveform (i.e. the convex hull). Interestingly, the temporal scaling of amplitude fluctuations in ongoing oscillations has recently been associated with cognitive impairments such as depression [66] and dementia [67]. Also, criticality has been supported by multifractal patterns in the same data previously supporting 1/f scaling [44]. Multifractal patterns occur when scaling relations (i.e. their exponents) vary over time or space, thereby adding a further dimension of complexity to data.

1/f scaling characterizes the central tendency of multifractal human performance, and thus the intrinsic fluctuations in neural and behavioral activity, be they from ion channels or brain images or text sequences [55]. 1/f scaling suggests that criticality underlies cognitive function at multiple scales and levels of analysis. Although observations of 1/f scaling in isolation do not constitute conclusive evidence for criticality (for other explanations, see Refs [3,68–70]), multifractal 1/f scaling greatly strengthens the case [44]. Additionally, criticality predicts power-law distributions and pervasive temporal and spatial long-range correlations in collective measures of component activities. These predictions are supported by the evidence reviewed here for neural avalanches [62], power-law distributions in word frequencies [33] and reaction times [28], and analyses showing pervasive 1/f scaling in neural [54] and behavioral activity [5] fluctuations. Adding multifractality to the mounting evidence means that metastability near critical points is the only candidate hypothesis that could explain the existing data.

Concluding remarks

In this brief review, a variety of scaling laws in cognitive science were discussed that plausibly express adaptive properties of perception, action, memory, language and computation. The working hypothesis of criticality can provide a general framework for understanding scaling laws and has motivated the application of new analytical tools to understand variability in cognitive systems. Much work lies ahead, however, to further test these new hypotheses and also to bring more scientists into the debate (Box 3).

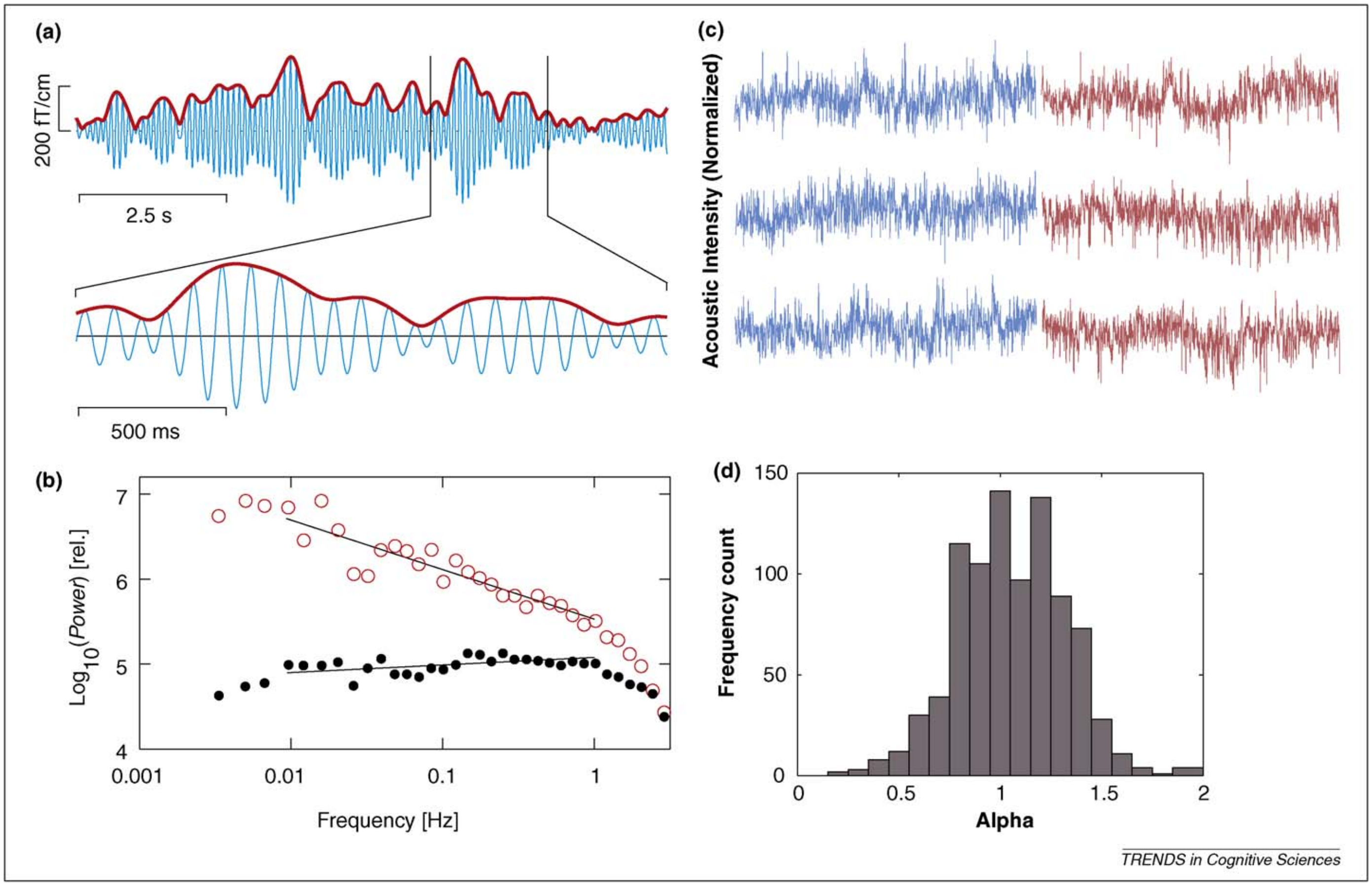

Figure 4. 1/f scaling in neural and behavioral activity. A band-pass filtered signal (6.7–13.3 Hz, thin blue lines) from a single channel (0.1–100 Hz) of magnetoencephalography (MEG) recording is shown in **(a)** at two time scales, filtered through a Morlet wavelet with a passband from ~6.7 to 13.3 Hz (reproduced, with permission from [84]). The log–log power spectrum of the resulting amplitude envelope of the oscillations (**a**, thick red lines) is shown in **(b)**. Evidence for 1/f scaling is seen in the negatively sloped line for MEG data (open red circles), and evidence against an artifactual explanation is seen in the contrasting flat line for reference channel control data (filled black circles). 1/f scaling indicates that ongoing neural oscillations carry a long-range memory of their own dynamics across hundreds or even thousands of cycles. The same type of memory is also found in acoustic power intensity fluctuations in spoken word repetitions, shown in **(c)** for one speaker's 1024 repetitions of the word ''bucket'' (reproduced with permission from [5]). Intensity fluctuations are shown separately for each acoustic syllable, at three different passbands (center frequencies of ~150 Hz, 6 kHz and 13 kHz). In total, 90 fluctuation series were observed for each of 10 speakers, and the $1/f^{\alpha}$ exponent was estimated for each series. The resulting distribution **(d)** was centered around $\alpha \sim 1$.

Some of this work will need to address difficulties in distinguishing alternative accounts of data (e.g. long-range versus short-range correlations and exponential versus lognormal versus power-law distributions). Recent advances in model identification methods have strengthened conclusions, but evidence is still more compelling when scaling laws are observed to span many orders of magnitude. Such observations require large amounts of data to be collected, which can be prohibitive, but technological advances are making large datasets more viable (e.g. in brain imaging and electronic corpora).

Other work will need to advance models of cognitive processes, because most of them are currently not designed to account for scaling laws, yet scaling laws appear widespread in cognitive science. Research is needed to determine whether current models could explain scaling laws within their purview, perhaps with small modifications or extensions, or whether new models and theories are needed to explain them. It is likely that common principles will be needed to fully explain some observations of scaling laws. Preferential attachment and self-organized criticality are two examples proposed to explain a wide range of scaling law observations throughout nature, including cognitive science [5,71]. Although it is unlikely that all observations of scaling laws in cognitive science have a common explanation, they can deepen our understanding of scaling laws and their meaning for cognitive function (e.g. if some are logical or mathematical consequences of other more fundamental laws of nature).

Box 3. Outstanding questions

- How can scaling laws be robustly and reliably detected in cognitive science data?
- How can models of cognitive processes explain observations of scaling laws?
- How many scaling laws in cognitive science can be explained by fundamental principles such as those found in statistical physics?
- How can variability in scaling law exponents be explained as it is found across different individuals, different measures of cognitive performance and different measurement conditions?
- How are scaling law exponents empirically and theoretically related across different measures and types of scaling laws in cognitive science?

Any fundamental approach to scaling laws in cognitive science will need to explain variability observed in scaling law parameters estimated from data. Scaling laws are parameterized by exponents, and exponents are observed to vary across individuals [28,72], across tasks [6] and across time [44]. It should not be surprising that cognitive science data are the most complex in this regard throughout nature, because exponents of scaling laws in other empirical domains are often observed to be relatively constant. It is an open question how variability in scaling laws reflects the flexibility and contextuality of cognition [73,74].

Acknowledgments
This article originated from a symposium entitled "Scaling Laws in Cognitive Science" that the authors presented at the 31st Annual Meeting of the Cognitive Science Society in Amsterdam, the Netherlands. The authors would like to thank the reviewers and A. Corral for helpful comments and suggestions. GDAB was supported by grant RES 062 23 0545 from the Economic and Social Research Council (UK). RFiC was supported by the SESAAME-BAR [Secuencias Simbólicas: Análisis, Aprendizaje, Minería y Evolución-Bacelona (in English "Symbolic Sequences: Analysis, Learning, Mining and Evolution-Barcelona")] project of the Spanish Ministry of Science and Innovation (TIN2008-06582-C03-01). KL-H was supported by the Innovative Research Incentive Schemes of the Netherlands Organization for Scientific Research. JGH was supported by NSF BCS-0446813, and JGH and GVO were supported by BCS-0642718. CTK was supported by NSF BCS-0842784 and the Keck Futures Initiative.

References
1 Chen, Y. *et al.* (1997) Long memory processes (1/f alpha type) in human coordination. *Phys. Rev. Lett.* 79, 4501
2 Wagenmakers, E-J. *et al.* (2004) Estimation and interpretation of l/f alpha noise in human cognition. *Psychon. Bull. Rev.* 11, 579–615
3 Farrell, S. *et al.* (2006) 1/f noise in human cognition: is it ubiquitous, and what does it mean? *Psychon. Bull. Rev.* 13, 737–741
4 Gisiger, T. (2001) Scale invariance in biology: coincidence or footprint of a universal mechanism? *Biol. Rev.* 76, 161–209
5 Kello, C.T. *et al.* (2008) The pervasiveness of 1/f scaling in speech reflects the metastable basis of cognition. *Cogn. Sci.* 32, 1217–1231
6 Kello, C.T. *et al.* (2007) The emergent coordination of cognitive function. *J. Exp. Psychol. Gen.* 136, 551–568
7 Thornton, T.L. and Gilden, D.L. (2005) Provenance of correlations in psychophysical data. *Psychon. Bull. Rev.* 12, 409–441
8 Gilden, D. (2009) Global model analysis of cognitive variability. *Cogn. Sci.* 33, 1441–1467
9 Sims, D.W. *et al.* (2008) Scaling laws of marine predator search behaviour. *Nature* 451, 1098–1102
10 Clauset, A. *et al.* (2009) Newman, Power-law distributions in empirical data. *SIAM Rev.* 51, 661–703
11 Rapoport, A. (1982) Zipf's law revisited. In *Studies on Zipf's Law* (Guiter, H. and Arapov, M.V., eds), pp. 1–28, Studienverlag Brockmeyer
12 Stevens, S.S. (1957) On the psychophysical law. *Psychol. Rev.* 64, 153–181
13 Lacquaniti, F. *et al.* (1983) The law relating the kinematic and figural aspects of drawing movements. *Acta Psychol.* 54, 115–130
14 Copelli, M. and Campos, P.R.A. (2007) Excitable scale free networks. *Eur. Phys. J. B Cond. Matter Complex Syst.* 56, 273–278
15 Copelli, M. *et al.* (2005) Signal compression in the sensory periphery. *Neurocomputing* 65–66, 691–696
16 Kinouchi, O. and Copelli, M. (2006) Optimal dynamical range of excitable networks at criticality. *Nat. Phys.* 2, 348–351
17 Harris, C.M. and Wolpert, D.M. (1998) Signal-dependent noise determines motor planning. *Nature* 394, 780–784
18 Chater, N. and Brown, G.D.A. (1999) Scale-invariance as a unifying psychological principle. *Cognition* 69, 17–24
19 Chater, N. and Brown, G.D.A. (2008) From universal laws of cognition to specific cognitive models. *Cogn. Sci.* 32, 36–67
20 Maylor, E.A. *et al.* (2001) Scale invariance in the retrieval of retrospective and prospective memories. *Psychon. Bull. Rev.* 8, 162–167
21 Brown, G.D.A. *et al.* (2007) A temporal ratio model of memory. *Psychol. Rev.* 114, 539–576
22 Brown, G.D.A. *et al.* (2008) Serial and free recall: common effects and common mechanisms? A reply to Murdock (2008). *Psychol. Rev.* 115, 781–785
23 Gallistel, C.R. and Gibbon, J. (2000) Time, rate, and conditioning. *Psychol. Rev.* 107, 289–344
24 Rhodes, T. and Turvey, M.T. (2007) Human memory retrieval as Lévy foraging. *Physica A* 385, 255–260
25 Reynolds, A.M. and Rhodes, C.J. (2009) The Lévy flight paradigm: random search patterns and mechanisms. *Ecology* 90, 877–887
26 Lomholt, M.A. *et al.* (2008) Lévy strategies in intermittent search processes are advantageous. *Proc. Natl. Acad. Sci. U. S. A.* 105, 11055–11059
27 Bousfield, W.A. and Sedgewick, C.H.W. (1944) An analysis of sequences of restricted associative responses. *J. Gen. Psychol.* 30, 149–165
28 Holden, J.G. *et al.* (2009) Dispersion of response times reveals cognitive dynamics. *Psychol. Rev.* 116, 318–342
29 Zipf, G.K. (1949) *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*, Addison-Wesley
30 Ferrer i Cancho, R. *et al.* (2004) Patterns in syntactic dependency networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 69, 051915
31 Kello, C.T. and Beltz, B.C. (2009) Scale-free networks in phonological and orthographic wordform lexicons. In *Approaches to Phonological Complexity* (Chitoran, I. *et al.*, eds), Mouton de Gruyter
32 Miller, G.A. and Chomsky, N. (1963) Finitary models of language users. In *Handbook of Mathematical Psychology* (Luce, R.D. *et al.*, eds), pp. 419–491, Wiley
33 Ferrer-i-Cancho, R. and Elvevåg, B. (2010) Random texts do not exhibit the real Zipf's law-like rank distribution. *PLoS ONE* 5 (3), e9411
34 Ferrer-i-Cancho, R. and Gavaldà, R. (2009) The frequency spectrum of finite samples from the intermittent silence process. *J. Am. Soc. Inf. Sci. Technol.* 60, 837–843
35 Ferrer i Cancho, R. and Solé, R.V. (2002) Zipf's law and random texts. *Adv. Complex Syst.* 5, 1–6
36 Cohen, A. *et al.* (1997) Numerical analysis of word frequencies in artificial and natural language texts. *Fractals* 5, 95–104
37 Ferrer-i-Cancho, R. *et al.* (2005) The consequences of Zipf's law for syntax and symbolic reference. *Proc. R. Soc. Lond. B Biol. Sci.* 272, 561–565
38 Ferrer-i-Cancho, R. (2006) When language breaks into pieces. A conflict between communication through isolated signals and language. *Biosystems* 84, 242–253
39 Köhler, R. (1986) *Zur linguistischen Synergetik: Struktur und Dynamik der Lexik*, Brockmeyer
40 Haspelmath, M. *et al.* (2005) *World Atlas of Language Structures*, Oxford University Press
41 Van Orden, G.C. *et al.* (2003) Self-organization of cognitive performance. *J. Exp. Psychol. Gen.* 132, 331–350
42 Van Orden, G.C. *et al.* (2005) Human cognition and 1/f scaling. *J. Exp. Psychol. Gen.* 134, 117–123
43 Van Orden, G.C. *et al.* (2009) Living in the pink: intentionality, wellness, and complexity. In *Philosophy of Complex Systems: Handbook of the Philosophy of Science* (Hooker, C., ed.), Elsevier
44 Inhen, E.A.F. and Vereijken, B. Beyond 1/f fluctuations in cognitive performance. *J. Exp. Psychol. Gen.* (in press).
45 Turcotte, D.L. and Rundle, J.B. (2002) Self-organized complexity in the physical, biological, and social sciences. *Proc. Natl. Acad. Sci. U. S. A.* 99, 2463–2465
46 Christensen, K. and Moloney, N.R. (2005) *Complexity and Criticality*, Imperial College Press
47 Chialvo, D.R. (2008) Emergent complexity: what uphill analysis or downhill invention cannot do. *New Ideas Psychol.* 26, 158–173
48 Bak, P. and Paczuski, M. (1995) Complexity, contingency, and criticality. *Proc. Natl. Acad. Sci. U. S. A.* 92, 6689–6696
49 Sornette, D. (2004) *Critical Phenomena in Natural Sciences: Chaos, Fractals, Self-organization, and Disorder: Concepts and Tools*, (2nd edn), Springer
50 Levina, A. *et al.* (2007) Dynamical synapses causing self-organized criticality in neural networks. *Nat. Phys.* 3, 857–860
51 Chialvo, D.R. *et al.* (2008) The brain: What is critical about it? *AIP Conf. Proc.* 1028, 28–45

52 de Arcangelis, L. *et al.* (2006) Self-organized criticality model for brain plasticity. *Phys. Rev. Lett.* 96, 028107
53 Poil, S-S. *et al.* (2008) Avalanche dynamics of human brain oscillations: relation to critical branching processes and temporal correlations. *Hum. Brain Mapp.* 29, 770–777
54 Linkenkaer-Hansen, K. *et al.* (2001) Long-range temporal correlations and scaling behavior in human brain oscillations. *J. Neurosci.* 21, 1370–1377
55 Kello, C.T. and Van Orden, G.C. (2009) Soft-assembly of sensorimotor function. *Nonlinear Dynamics Psychol. Life Sci.* 13, 57–78
56 Bertschinger, N. and Natschlager, T. (2004) Real-time computation at the edge of chaos in recurrent neural networks. *Neural Comput.* 16, 1413–1436
57 Tognoli, E. and Kelso, J.A.S. (2009) Brain coordination dynamics: true and false faces of phase synchrony and metastability. *Prog. Neurobiol.* 87, 31–40
58 Kelso, J. and Tognoli, E. (2009) Toward a complementary neuroscience: metastable coordination dynamics of the brain. In *Downward Causation and the Neurobiology of Free Will* (Heidelberg, S.B., ed.), pp. 103–124, Springer
59 Packard, N. (1988) Adaptation towards the edge of chaos. In *Dynamic Patterns in Complex Systems* (Kelso, J.A.S. *et al.*, eds), pp. 293–301, World Scientific
60 Langton, C.G. (1990) Computation at the edge of chaos – phase-transitions and emergent computation. *Physica. D* 42, 12–37
61 Beggs, J.M. (2008) The criticality hypothesis: how local cortical networks might optimize information processing. *Philos. Transact. A Math. Phys. Eng. Sci.* 366, 329–344
62 Beggs, J.M. and Plenz, D. (2003) Neuronal avalanches in neocortical circuits. *J. Neurosci.* 23, 11167–11177
63 Ferrer i Cancho, R. and Solé, R.V. (2003) Least effort and the origins of scaling in human language. *Proc. Natl. Acad. Sci. U. S. A.* 100, 788–791
64 Ferrer-i-Cancho, R. (2005) Zipf's law from a communicative phase transition. *Eur. Phys. J. B* 47, 449–457
65 Ferrer-i-Cancho, R. and Diaz-Guilera, A. (2007) The global minima of the communicative energy of natural communication systems. *J. Stat. Mech.* P06009
66 Linkenkaer-Hansen, K. *et al.* (2005) Breakdown of long-range temporal correlations in theta oscillations in patients with major depressive disorder. *J. Neurosci.* 25, 10131–10137
67 Montez, T. *et al.* (2009) Altered temporal correlations in parietal alpha and prefrontal theta oscillations in early-stage Alzheimer disease. *Proc. Natl. Acad. Sci. U. S. A.* 106, 1614–1619
68 Grigolini, P. *et al.* (2009) A theory of 1/f noise in human cognition. *Phys. A Stat. Mech. Appl.* 388, 4192–4204
69 Medina, J.M. (2009) 1/f alpha noise in reaction times: a proposed model based on Pieron's law and information processing. *Phys. Rev. E* 79, 011902
70 Torre, K. and Wagenmakers, E-J. (2009) Theories and models for 1/f[beta] noise in human movement science. *Hum. Mov. Sci.* 28, 297–318
71 Steyvers, M. and Tenenbaum, J.B. (2005) The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cogn. Sci.* 29, 41–78
72 Gilden, D.L. and Hancock, H. (2007) Response variability in attention-deficit disorders. *Psychol. Sci.* 18, 796–802
73 Riley, M.A. and Turvey, M. (2002) Variability and determinism in motor behavior. *J. Mot. Behav.* 34, 99–125
74 Zbilut, J.P. (2004) *Unstable Singularities and Randomness: Their Importance in the Complexity of Physical, Biological and Social Sciences*, Elsevier
75 Spence, A.J. (2009) Scaling in biology. *Curr. Biol.* 19, R57–R61
76 Zhang, K. and Sejnowski, T.J. (2000) A universal scaling law between gray matter and white matter of cerebral cortex. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5621–5626
77 Onsager, L. (1944) A two-dimensional model with an order-disorder transition. *Phys. Rev.* 65, 117–149
78 Bak, P. *et al.* (1988) Self-organized criticality. *Phys. Rev. A* 38, 364
79 Mitzenmacher, M. (2004) A brief history of generative models for power law and lognormal distributions. *Internet Math.* 1, 226–251
80 Newman, M.E.J. (2005) Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* 46, 323–351
81 Mandelbrot, B. (1967) How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science* 156, 636–638
82 Gabaix, X. (2006) Zipf's law for cities: an explanation. *Q. J. Econ.* 114, 739–767
83 Ramos, F. *et al.* (2004) Levy walk patterns in the foraging movements of spider monkeys (*Ateles geoffroyi*). *Behav. Ecol. Sociobiol.* 55, 223–230
84 Linkenkaer-Hansen, K. *et al.* (2004) Stimulus-induced change in long-range temporal correlations and scaling behaviour of sensorimotor oscillations. *Eur. J. Neurosci.* 19, 203–211