need for vaccines (*14*, *15*). This technology could accelerate the construction of live vaccine strains.

Many medically or industrially important microbes are difficult to manipulate genetically. This has severely limited our understanding of pathogenesis and our ability to exploit the knowledge of microbial biology on a practical level. We hope that the cycle presented here can be applied to other species, to help solve these problems.

References and Notes

1. A. J. DaMassa, D. L. Brooks, H. E. Adler, *Am. J. Vet. Res.* **44**, 322 (1983).
2. L. Manso-Silvan *et al.*, *Int. J. Syst. Evol. Microbiol.* **59**, 1353 (2009).
3. Manso-Silvan *et al*. (*2*) report the very recent renaming of *Mycoplasma mycoides* subspecies *mycoides* Large Colony to *Mycoplasma mycoides* subspecies *capri*.
4. C. Lartigue *et al.*, *Science* **317**, 632 (2007).
5. D. G. Gibson *et al.*, *Science* **319**, 1215 (2008).
6. D. G. Gibson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20404 (2008).
7. K. W. King, K. Dybvig, *Plasmid* **26**, 108 (1991).
8. J. I. Glass *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 425 (2006).
9. Materials and methods are available as supporting material on *Science* Online.
10. N. Kouprina, V. Larionov, *Nat. Protocols* **3**, 371 (2008).
11. V. Larionov, N. Kouprina, G. Solomon, J. C. Barrett, M. A. Resnick, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7384 (1997).
12. J. D. Boeke, F. LaCroute, G. R. Fink, *Mol. Gen. Genet.* **197**, 345 (1984).
13. C. Janis *et al.*, *Appl. Environ. Microbiol.* **71**, 2888 (2005).
14. L. Dedieu-Engelmann, *Comp. Immunol. Microbiol. Infect. Dis.* **31**, 227 (2008).
15. F. Thiaucourt *et al.*, *Dev. Biol. (Basel)* **114**, 147 (2003).

16. We thank R. Roberts for helpful information and the Synthetic Biology team at the J. Craig Venter Institute (JCVI) for critical discussions about the manuscript. This work was supported by Synthetic Genomics, Inc. (SGI). J.C.V. is Chief Executive Officer and Co-Chief Scientific Officer of SGI. H.O.S. is Co-Chief Scientific Officer and on the Board of Directors of SGI. C.A.H. is Chairman of the SGI Scientific Advisory Board. All three of these authors and JCVI hold SGI stock. JCVI has filed patent applications on some of the techniques described in this paper.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1173759/DC1
Materials and Methods
Figs. S1 to S7
Table S1
References

18 March 2009; accepted 13 August 2009
Published online 20 August 2009;
10.1126/science.1173759
Include this information when citing this paper.

On Universality in Human Correspondence Activity

R. Dean Malmgren,[1,2]* Daniel B. Stouffer,[1,3] Andriana S. L. O. Campanharo,[1,4] Luís A. Nunes Amaral[1,5,6]*

The identification and modeling of patterns of human activity have important ramifications for applications ranging from predicting disease spread to optimizing resource allocation. Because of its relevance and availability, written correspondence provides a powerful proxy for studying human activity. One school of thought is that human correspondence is driven by responses to received correspondence, a view that requires a distinct response mechanism to explain e-mail and letter correspondence observations. We demonstrate that, like e-mail correspondence, the letter correspondence patterns of 16 writers, performers, politicians, and scientists are well described by the circadian cycle, task repetition, and changing communication needs. We confirm the universality of these mechanisms by rescaling letter and e-mail correspondence statistics to reveal their underlying similarity.

Power law statistics are a hallmark of critical phenomena. A less obvious characteristic of criticality is the emergence of universality classes that capture the similarity of seemingly disparate systems. For example, despite the fact that water and carbon dioxide have different chemical properties, they were observed to behave in the same manner when close to their respective critical points (*1*). This is because idiosyncrasies, such as the existence of electric dipoles or the ability to form hydrogen bonds, become irrelevant near the liquid/gas critical point. For physical systems, renormalization group theory (*2*, *3*) has enabled researchers to understand the deep connection between the symmetries of a system and the mechanisms that underlie its behavior. The similarity of different fluids near their respective liquid/gas critical points is often demonstrated by rescaling their statistics so that they collapse onto the same universal curves (often power law curves),which have particular scaling exponents. By grouping different substances into the same universality class, as identified by its scaling exponents, one discovers that fluids are described by the same statistical laws near the liquid/gas critical point as uniaxial magnets are near their paramagnetic critical point (*1*). One can also differentiate the behavior of these systems from the behavior of polymers near the sol/gel transition, which belong to a different universality class (*1*).

In addition to describing critical phenomena, power law scaling has also been widely reported in biology, economics, and sociology (*4*–*10*). Renormalization group theory therefore offers a tantalizing hypothesis for the prevalence of particular power law scaling exponents in social systems: Social systems, in analogy with physical systems, may operate near critical points and can therefore be classified into a small number of distinct universality classes. A heated debate has consequently ensued in the literature concerning the "universality of human systems" (in the statistical physics meaning of the phrase). Is there enough statistical evidence for the asymptotic power law description of the heavy-tailed distributions reported in human systems (*11*–*14*)? Is it reasonable to postulate that social systems, like their physical counterparts (*2*, *3*, *15*), can be classified into universality classes according to scaling exponents (*16*)?

Human correspondence is a paradigmatic area where the matters of power law scaling and universality are contentious issues. One view that has recently received considerable attention in the literature (*17*, *18*) posits that correspondence patterns are driven primarily by the need to respond to other individuals. This is formalized by a priority queuing model (*19*), which, under certain limiting conditions, reproduces the asymptotic scaling of empirically observed heavy-tailed correspondence statistics. In particular, the heavy-tailed statistical properties of e-mail correspondence are reportedly reproduced by a fixed-length queue with a single task type (*19*, *20*), whereas the heavy-tailed statistical properties of letter correspondence are reportedly reproduced by either a variable-length queue with a single task type (*21*, *20*) or by a fixed-length queue with multiple task types (*22*). The fact that there are different exponents for the two modes of correspondence has been taken as evidence that human correspondence falls into one of two universality classes (*20*). When interpreted in the statistical mechanics sense of universality, one would conclude that e-mail and letter correspondence are fundamentally different activities.

In contrast, we hypothesize that human correspondence patterns are not driven by responses to others but by more prosaic mechanisms: the circadian cycle, task repetition, and changing communication needs. We formalize these mechanisms with a cascading, nonhomogeneous Poisson process that we have previously shown to be statistically consistent with e-mail communication patterns (*14*). We hypothesize that the same model is capable of describing letter correspondence and that the heavy-tailed correspondence statistics primarily arise from the variation in an individual's communication needs over the course of his or her lifetime.

---

[1]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA. [2]Datascope Analytics, Evanston, IL 60201, USA. [3]Integrative Ecology Group, Estación Biológica de Doñana, CSIC, 41092 Sevilla, Spain. [4]Instituto Nacional de Pesquisas Espaciais, 12227-010 São José dos Campos, São Paulo, Brazil. [5]Howard Hughes Medical Institute, Northwestern University, Evanston, IL 60208, USA. [6]Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL 60208, USA.

*To whom correspondence should be addressed. E-mail: dean.malmgren@u.northwestern.edu (R.D.M.); amaral@northwestern.edu (L.A.N.A.)

We obtained the letter correspondence records for 16 writers, performers, politicians, and scientists. Each data set consists of a list of letters that were sent by each of these individuals, and each record comprises the name of the sender, the name of the recipient, and the date when the letter was written [see supporting online material (SOM) text S1 for details]. The nature of the data raises two issues to consider during analysis. First, the precise authorship date of some letters is unknown, so we restricted our analysis to only those letters that have precise authorship dates. Second, it is highly unlikely that all of the letters written by a particular individual are present in the database. We have confirmed that our results are insensitive to sampling effects from this method of data collection (SOM text S2).

An important consideration in studying the letter correspondence patterns of these individuals is that the data cover their entire lifetimes. As a result, it is conceivable that changing communication needs might affect letter correspondence patterns. For example, before Einstein became widely known, the bulk of his recorded communication was to friends and relatives. After the confirmation of his theory of relativity in 1919, Einstein's need to communicate with other individuals substantially increased. By that time, his stepdaughter Ilse Einstein was helping him with secretarial tasks, resulting in greatly improved coverage of his recorded correspondence (*23*). Because of this secretarial assistance and his increased fame, we expect that the average time between consecutively sent letters, the average interevent time $\langle\tau\rangle$, is significantly larger during the beginning of Einstein's life than during the latter part of his life. Our expectations are verified in Fig. 1, A and B, demonstrating that these time series are nonstationary; that is, the heavy-tailed τ distribution results from a mixture of time scales (*24*).

Because these time series are nonstationary, we partitioned each complete time series into smaller time segments so that we could approximate stationary behavior within each time segment. We accomplished this by splitting the time series into segments lasting 364 days (52 weeks), unless fewer than 10 events fell within that time period, in which case consecutive segments were merged until this criterion was met.

Assuming that the correspondence patterns within each time segment are stationary, we can then model the behavior within each time segment with standard techniques. As a first approximation, one might naively expect that letters are sent at a constant rate ρ and that the time at which every letter is sent is independent of all others. Such a process is referred to as a homogeneous Poisson process, which gives rise to an exponential τ distribution $p(\tau) = \rho e^{-\rho\tau}$. Whereas the tail of the τ distribution within these time segments is approximately exponential, the best-estimate predictions of a homogeneous Poisson process do not produce the correct decay rate (Fig. 1C). This suggests that only a few changes to the homogeneous Poisson process are needed to reproduce the observed τ distribution. We hypothesize that, as for e-mail correspondence, two additional ingredients must also be considered for modeling letter correspondence (*14*).

First, circadian and weekly cycles of activity may influence when individuals communicate. Previously, we accounted for these cycles of activity in e-mail communication with a nonhomogeneous Poisson process whose rate $\rho(t)$ changes periodically on daily and weekly time scales. For letter correspondence, however, the resolution of the data does not permit us to identify activity patterns within a day, and day-to-day changes in activity provide no additional insight (SOM text S3). We therefore approximate the nonhomogeneous Poisson process $\rho(t)$ by a homogeneous Poisson process with constant rate ρ_i during time segment i; that is, we model the rate of activity $\rho(t)$ throughout each individual's life by a piecewise constant function of time.

Second, individuals are much more likely to continue writing letters once they have written one letter, in order to use their time more effectively. We account for this behavior by hypothesizing that, once an individual finishes writing a letter, there is a probability ξ_i that they will write another letter. This process repeats itself until this cascade of additional letters concludes with probability $1 - \xi_i$, at which point the individual's behavior is again governed by a homogeneous Poisson process with rate ρ_i (*25*). We refer to the resulting model as a cascading Poisson process.

To compare the predictions of the cascading Poisson process (*26*) to the empirical data, we

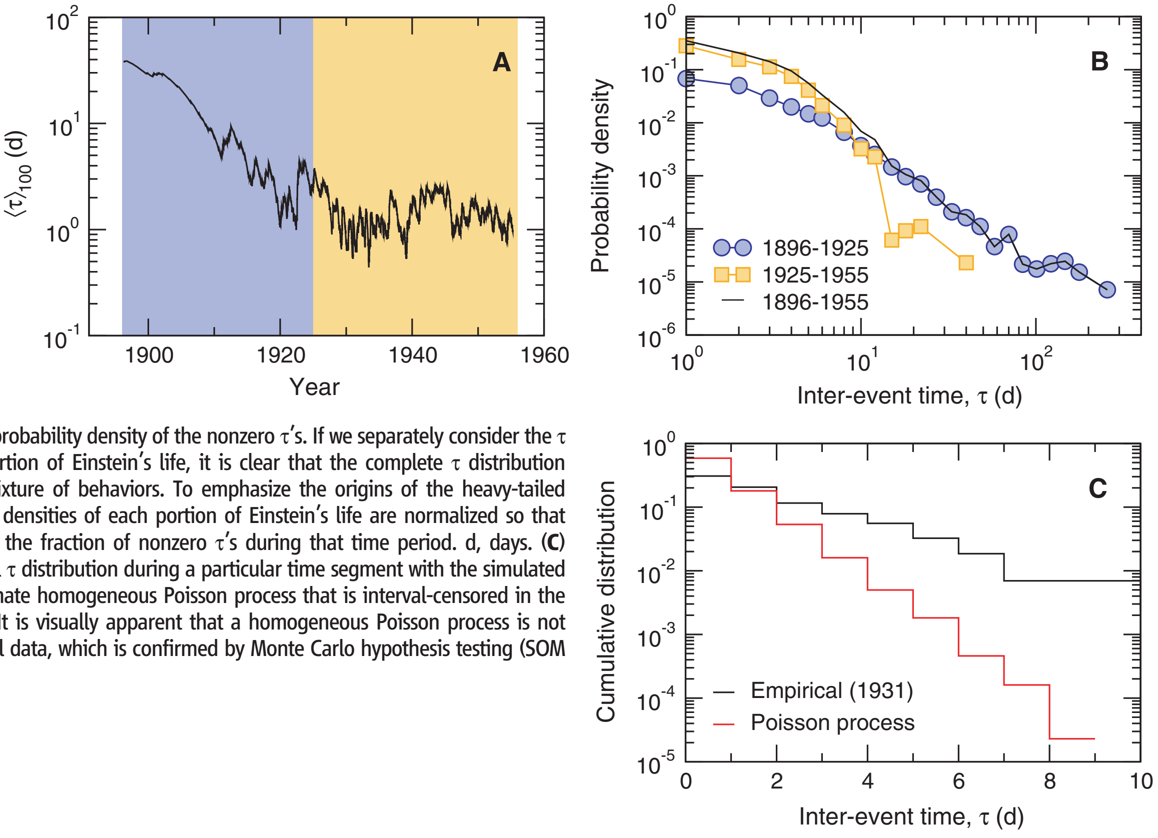


Fig. 1. Nonstationarity of Albert Einstein's letter correspondence activity. We selected Einstein as an example, but nonstationarities are present for all 16 writers, performers, politicians, and scientists studied here. (**A**) Average of τ over 100 consecutive τ's. During the beginning of Einstein's life (blue-shaded region), $\langle\tau\rangle_{100}$ is significantly larger than during the end of his life (orange-shaded region). (**B**) Logarithmically binned probability density of the nonzero τ's. If we separately consider the τ distribution during each portion of Einstein's life, it is clear that the complete τ distribution (black line) is actually a mixture of behaviors. To emphasize the origins of the heavy-tailed distribution, the probability densities of each portion of Einstein's life are normalized so that their integrals are equal to the fraction of nonzero τ's during that time period. d, days. (**C**) Comparison of the empirical τ distribution during a particular time segment with the simulated predictions of the best-estimate homogeneous Poisson process that is interval-censored in the same manner as the data. It is visually apparent that a homogeneous Poisson process is not consistent with the empirical data, which is confirmed by Monte Carlo hypothesis testing (SOM text S3).

must first estimate the parameters $\boldsymbol{\theta}_i = \{\rho_i, \xi_i\}$ from the data during each time segment. The nature of the data, however, raises an important concern for parameter estimation: Because each event is only known to occur within a particular day, not at a precise time of the day, the data are interval-censored (*27*). We account for the interval-censored data and calculate the best-estimate parameters $\hat{\boldsymbol{\theta}}_i$ by numerically maximizing the censored likelihood function (see SOM text S4 for the derivation).

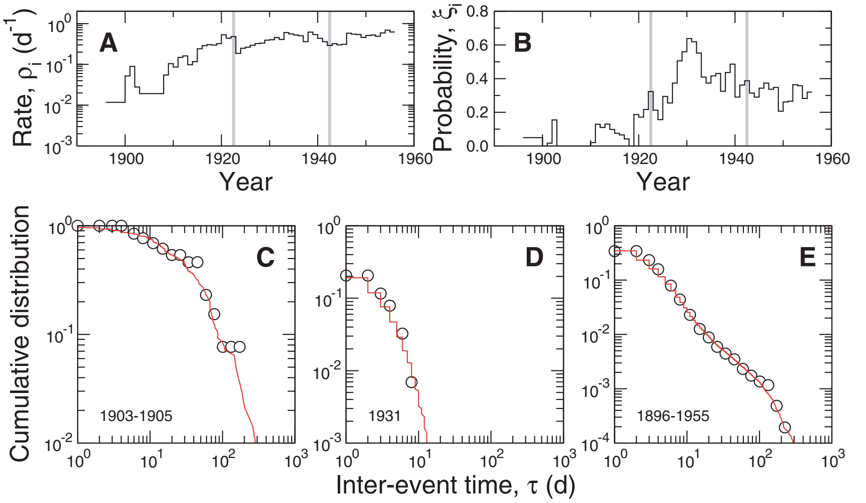
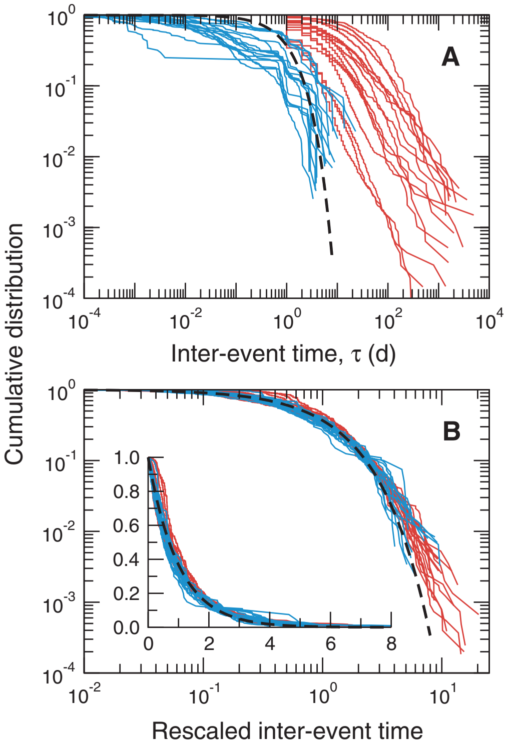


Fig. 2. Origin of heavy-tailed τ distribution for Albert Einstein. We selected Einstein as an example, but the same explanation is relevant for all 16 writers, performers, politicians, and scientists studied here. (**A** and **B**) We estimate the parameters $\boldsymbol{\theta}_i = \{\rho_i, \xi_i\}$ by maximizing the censored likelihood function for each time segment (SOM text S4). Gray-shaded regions denote time segments during which the cascading Poisson process was rejected by Monte Carlo hypothesis testing. Parameter estimates for all individuals studied here are shown in fig. S4. There is a 50-fold changes in the rate ρ_i and dramatic changes in ξ_i for Einstein. (**C** and **D**) Cumulative distribution of τ's (circles) for Einstein during 1903–1905 and 1931. The data agree with the predictions of a nonstationary cascading Poisson process (red line) with the best-estimate parameters shown in (A) and (B). The model predictions are generated numerically by running the model defined by $\boldsymbol{\theta}(t)$ 10 times and interval-censoring the resulting synthetic time series in the same manner as the empirical data. (**E**) The cumulative distribution of τ's (circles) for Einstein over his entire life. The data agree with the predictions of a nonstationary cascading Poisson process (red line) with the best-estimate parameters shown in (A) and (B). The τ distributions for all 16 letter correspondents studied here are shown in fig. S5. These results clarify the origin of the heavy-tailed τ distribution.

Fig. 3. Collapse of τ distributions for letter and e-mail correspondence. (**A**) Cumulative distribution of τ's for all 16 letter correspondents (red lines) and 16 randomly selected e-mail correspondents (blue lines). (**B**) Cumulative distribution of rescaled τ's on logarithmic and linear (inset) axes. The interevent time $\tau_k = t_{k+1} - t_k$ is rescaled by the average τ expected during the interval $[t_k, t_{k+1}]$, $\langle\tau_k\rangle = (t_{k+1} - t_k)/\int_{t_k}^{t_{k+1}}\rho(s)ds$. By the time-rescaling theorem (*30*), the resulting rescaled τ distribution is given by the expected τ distribution for a homogeneous Poisson process with unit rate $P(\tau) = e^{-\tau}$ (black dashed line). We only consider inter-event times $\tau > 0$ for letter correspondence.



The resulting best-estimate parameters $\hat{\boldsymbol{\theta}}_i$ provide insight into the correspondence patterns of each individual (Fig. 2, A and B, and fig. S4). For example, whereas both Schoenberg and Einstein have a 50-fold increase in the rate at which they send letters, presumably due to their increasing correspondence obligations and a more complete sampling of their overall letter correspondence, their use of cascades of activity is markedly different. Schoenberg, for instance, sent about 21% of his letters during cascades of multiple letters throughout his life. In contrast, Einstein rarely used cascades of activity as a young man (before 1910), whereas in later years (after 1933), he sent approximately 34% of his letters during cascades of multiple letters.

In the period 1928–1933, Einstein sent over 50% of his letters during cascades of multiple letters. The start of this period coincides with the hiring of Einstein's long-time secretary Helen Dukas, who more systematically retained copies of his outgoing correspondence. After the Nazis took power in January 1933, his correspondence patterns change markedly; this possibly reflects changes in his correspondence obligations at the Institute for Advanced Study at Princeton University after immigrating to the United States in late 1933 (*23*).

Of course, inferring how an individual's behavior changes based on a model's parameter estimates is contingent on the model being consistent with the data. We tested the statistical consistency of our model with the data by Monte Carlo hypothesis testing (SOM text S5). We reject the model during a particular time segment if the *P* value obtained from the Monte Carlo hypothesis testing procedure is less than a threshold of 0.05. Because this threshold is greater than zero, it means that there is a finite chance that we will reject the hypothesis that the model is consistent with the data even if the data were generated from the model.

If we assume that each time segment is independent, then we would expect to reject each of the time segments with a 5% chance, and the total number of rejections is expected to be distributed according to a binomial model (*28*). Out of the 54 independent time segments for Einstein, for example, we would expect to reject the model 2.7 times, with 0 to 6 defining the bounds of the 95% confidence interval (CI) of the corresponding binomial model. For Einstein, our procedure rejects the cascading Poisson process for 2 out of 54 time segments, indicating that we cannot reject the hypothesis that the model is able to explain his correspondence patterns. Indeed, our hypothesis testing confirms that the cascading Poisson process cannot be rejected as an explanatory model for the letter correspondence of any of the individuals under consideration (Table 1). These results

Table 1. Summary of the letter correspondence records and hypothesis testing results for the 16 individuals under consideration, ordered chronologically. For each individual, we note the time period and duration of the letter correspondence records, the total number of letters sent, the number of time segments with at least 10 letters per time segment, the 95% CI bounds of the corresponding binomial model with $P = 0.05$, and the number of rejections of the cascading Poisson process based on our Monte Carlo hypothesis testing procedure. The number of Monte Carlo hypothesis testing rejections is within the 95% CI bounds for all 16 individuals, indicating that this model cannot be rejected for any individual's letter correspondence patterns. We have conducted the same analysis for three alternative models; we find that a cascading Poisson process provides the most parsimonious and statistically consistent explanation of the data [SOM text S3].

Individual	Time period	Duration (years)	Number of letters	Number of segments	95% CI	Number of rejections
Francis Bacon	1574–1626	53	443	19	[0,3]	3
James H. Leigh Hunt	1790–1859	70	408	25	[0,3]	1
Charles Darwin	1822–1882	61	6785	52	[0,5]	4
Anna Brownell Jameson	1833–1860	28	119	8	[0,2]	1
Friedrich Engels	1833–1895	63	369	24	[0,3]	1
Robert E. Lee	1835–1870	36	282	10	[0,2]	0
Karl Marx	1837–1882	46	469	25	[0,3]	1
Henry Irving	1852–1905	54	1205	35	[0,4]	0
Sigmund Freud	1872–1939	68	3130	49	[0,5]	2
Marcel Proust	1879–1922	44	668	25	[0,3]	2
H. G. Wells	1895–1946	52	422	16	[0,2]	0
Albert Einstein	1896–1955	60	10,319	54	[0,6]	2
Carl Sandburg	1898–1966	69	1894	37	[0,4]	2
Arnold Schoenberg	1902–1951	50	6899	47	[0,5]	3
Ernest Hemingway	1909–1961	53	1934	42	[0,5]	5
Stan Laurel	1924–1964	41	685	17	[0,3]	1

demonstrate that the origin of the heavy-tailed τ distribution is a mixture of distributions with different time scales (Fig. 2, C to E).

Our findings enable us to address a crucial question: Do e-mail and letter correspondence belong to different universality classes (*20*)? Because the same mechanistic model is capable of describing both e-mail and letter correspondence, we can answer this question in the negative. We demonstrate the underlying similarity of both correspondence activities by rescaling and collapsing the τ distributions for 16 randomly selected e-mail correspondents (*29*) for which we have model parameter estimates (*14*) and the 16 letter correspondents studied here (Fig. 3). The rescaled τ distributions agree with theoretical expectations (*30*), demonstrating that the same exponential statistical law is indeed capable of describing both correspondence patterns.

Only by understanding and validating the underlying mechanisms can we appropriately rescale e-mail and letter correspondence to reveal their underlying similarity. Unlike critical phenomena, the universality here does not arise from the irrelevance of idiosyncrasies but rather from the fact that these two different modes of communication are governed by the same mechanisms. This insight is not apparent just from studying the asymptotic scaling of an empirical distribution obtained from an individual; one cannot simply infer that different scaling exponents necessarily imply different mechanisms.

Our results therefore raise important questions about the nature of universality in complex phenomena in general, and in human correspondence in particular. Perhaps the most common universal statistical law is due to the central limit theorem: Sums of variates with finite fluctuations converge to a Gaussian distribution. When confronted with statistical patterns that are non-Gaussian, one is tempted to surmise that the system's fluctuations are not finite. In analogy to physical systems, the recurrence of power law dependencies with similar exponent values in biological or social systems is frequently hypothesized to arise from the operation of these systems near critical points, where particular details of the system become irrelevant.

A less explored hypothesis, as exemplified here, is that heavy-tailed distributions emerge as a result of nonstationarities in the absence of criticality (*14*, *31*). Our study demonstrates that human correspondence can be accurately modeled as a cascading nonhomogeneous Poisson process: a noncritical process. This process gives rise to heavy-tailed statistics but not to power law statistics characterized by critical exponents. Instead, the correspondence patterns of each individual are uniquely characterized by the parameters of our model (*32*); the process is universal, but the parameters are not.

Indeed, we hypothesize that the cascading Poisson process, which formally incorporates the effects of the circadian cycle, task repetition, and changing needs, may accurately describe many other aspects of human activity. The circadian cycle has physiologic impact and may thus affect numerous human behaviors, from eating habits to commuting routines. Task repetition is similarly important because of the increased efficiency it enables; once an individual makes one purchase at a mall, it is easier to make other purchases within that mall during the same trip than it is to return to the mall the following day. As one ages and changes roles, it is not hard to imagine that the extent to which the circadian cycle and task repetition influence an individual's activity might change over time. It is therefore plausible that the cascading Poisson processes presented here could be generalized to account for different types of activities, each with its own evolving parameters.

References and Notes

1. H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford Univ. Press, Oxford, 1971).
2. S.-K. Ma, *Modern Theory of Critical Phenomena* (Addison-Wesley, Redwood City, CA, 1976).
3. N. D. Goldenfeld, *Lectures on Phase Transitions and the Renormalisation Group* (Addison-Wesley, Reading, MA, 1992).
4. L. Bachelier, *Ann. Sci.Éc. Norm. Supér.* **3**, 21 (1900).
5. V. Pareto, *Manuale di Economia Politica* (Societa Editrice, Milan, 1906).
6. F. Auerbach, *Petermanns Geogr. Mitt.* **59**, 74 (1913).
7. R. Gibrat, *Les Inégalités Économiques* (Librairie du Recueil Sirey, Paris, 1931).
8. G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Addison-Wesley, Cambridge, MA, 1949).
9. Y. Ijiri, H. A. Simon, *Skew Distributions and the Sizes of Business Firms*, vol. 24 of *Studies in Mathematical and Managerial Economics* (North-Holland, Amsterdam, 1977).
10. V. Plerou *et al.*, *Nature* **400**, 433 (1999).
11. D. Avnir, O. Biham, D. Lidar, *Science* **279**, 39 (1998).
12. L. A. N. Amaral, A. Scala, M. Barthélémy, H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11149 (2000).
13. A. Clauset, C. R. Shalizi, M. E. J. Newman, preprint available at http://arxiv.org/abs/0706.1062 (2007).
14. R. D. Malmgren, D. B. Stouffer, A. E. Motter, L. A. N. Amaral, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18153 (2008).
15. H. E. Stanley, L. A. N. Amaral, P. Gopikrishnan, V. Plerou, *Phys. A* **283**, 31 (2000).
16. H. E. Stanley *et al.*, *Physica A* **224**, 302 (1996).
17. C. Castellano, S. Fortunato, V. Loreto, preprint available at http://arxiv.org/abs/0710.3256 (2007).
18. T. Zhou, X.-P. Han, B.-H. Wang, preprint available at http://arxiv.org/abs/0801.1389v1.
19. A.-L. Barabási, *Nature* **435**, 207 (2005).
20. A. Vázquez *et al.*, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **73**, 036127 (2006).
21. J. G. Oliveira, A.-L. Barabási, *Nature* **437**, 1251 (2005).
22. G. Grinstein, R. Linsker, *Phys. Rev. Lett.* **97**, 130201 (2006).
23. T. Sauer, personal communication.
24. J. M. Hausdorff, C.-K. Peng, *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **54**, 2154 (1996).
25. We assume that $\xi_i >> \rho_i dt$. This model formulation is equivalent to having N_a additional events, where N_a is drawn from an exponential distribution with mean $\xi_i/(1 - \xi_i)$.
26. Materials and methods are available as supporting material on *Science* Online.
27. R. B. D'Agostino, M. A. Stephens, *Goodness-of-Fit Techniques* (Marcel Kekker, New York, 1986).
28. R. G. Miller Jr., *Simultaneous Statistical Inference* (Springer-Verlag, New York, 1991).
29. J.-P. Eckmann, E. Moses, D. Sergi, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14333 (2004).
30. H. M. Taylor, S. Karlin, *An Introduction to Stochastic Modeling* (Academic Press, San Diego, CA, 1994).
31. L. A. N. Amaral, S. V. Buldyrev, S. Havlin, M. A. Salinger, H. E. Stanley, *Phys. Rev. Lett.* **80**, 1385 (1998).
32. R. D. Malmgren, J. M. Hofman, L. A. N. Amaral, D. J. Watts, in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, New York, 2009), pp. 607–615.

33. We thank T. Sauer, R. Guimerà, M. Sales-Pardo, M. J. Stringer, E. N. Sawardecker, J. Duch, and P. McMullen for insightful comments and suggestions. D.B.S. acknowledges the support of a Consejo Superior de Investigaciones Científicas–Junta para la Ampliación de Estudios (JAE) postdoctoral fellowship. A.S.L.O.C. acknowledges the support of a Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Brazil) doctoral fellowship. L.A.N.A. gratefully acknowledges the support of NSF award SBE 0624318. All figures were generated with PyGrace (http://pygrace.sourceforge.net) with color schemes from http://colorbrewer.org.

Supporting Online Material

www.sciencemag.org/cgi/content/full/325/5948/1696/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
Tables S1 to S3
References

6 April 2009; accepted 5 August 2009
10.1126/science.1174562

Antennal Circadian Clocks Coordinate Sun Compass Orientation in Migratory Monarch Butterflies

Christine Merlin, Robert J. Gegear, Steven M. Reppert*

During their fall migration, Eastern North American monarch butterflies (*Danaus plexippus*) use a time-compensated Sun compass to aid navigation to their overwintering grounds in central Mexico. It has been assumed that the circadian clock that provides time compensation resides in the brain, although this assumption has never been examined directly. Here, we show that the antennae are necessary for proper time-compensated Sun compass orientation in migratory monarch butterflies, that antennal clocks exist in monarchs, and that they likely provide the primary timing mechanism for Sun compass orientation. These unexpected findings pose a novel function for the antennae and open a new line of investigation into clock-compass connections that may extend widely to other insects that use this orientation mechanism.

Eastern North American monarch butterflies, *Danaus plexippus*, undergo one of the most magnificent long-distance migrations observed in animals. Each fall in the northern United States and southern Canada, migratory monarchs travel distances up to 4000 km to arrive at their overwintering grounds in central Mexico (*1*, *2*). The navigational abilities of the migrants include the use of a time-compensated Sun compass (*3*–*5*). Previous studies show that a circadian clock provides the internal timing device that allows the butterflies to correct their flight orientation, relative to skylight parameters, and to maintain a southerly flight bearing as the Sun moves across the sky during the day (*3*–*5*).

A distinctive circadian clock mechanism has been recently elucidated in the monarch butterfly (*6*). It relies on a negative transcriptional feedback loop that involves the transcription factors CLOCK (CLK) and CYCLE (CYC), which drive the expression of *period* (*per*), *timeless* (*tim*), and a vertebrate-like *cryptochrome* designated *cry2*. The translated PER, TIM, and CRY2 proteins

Department of Neurobiology, University of Massachusetts Medical School, Worcester, MA 01605, USA.

*To whom correspondence should be addressed. E-mail: Steven.Reppert@umassmed.edu

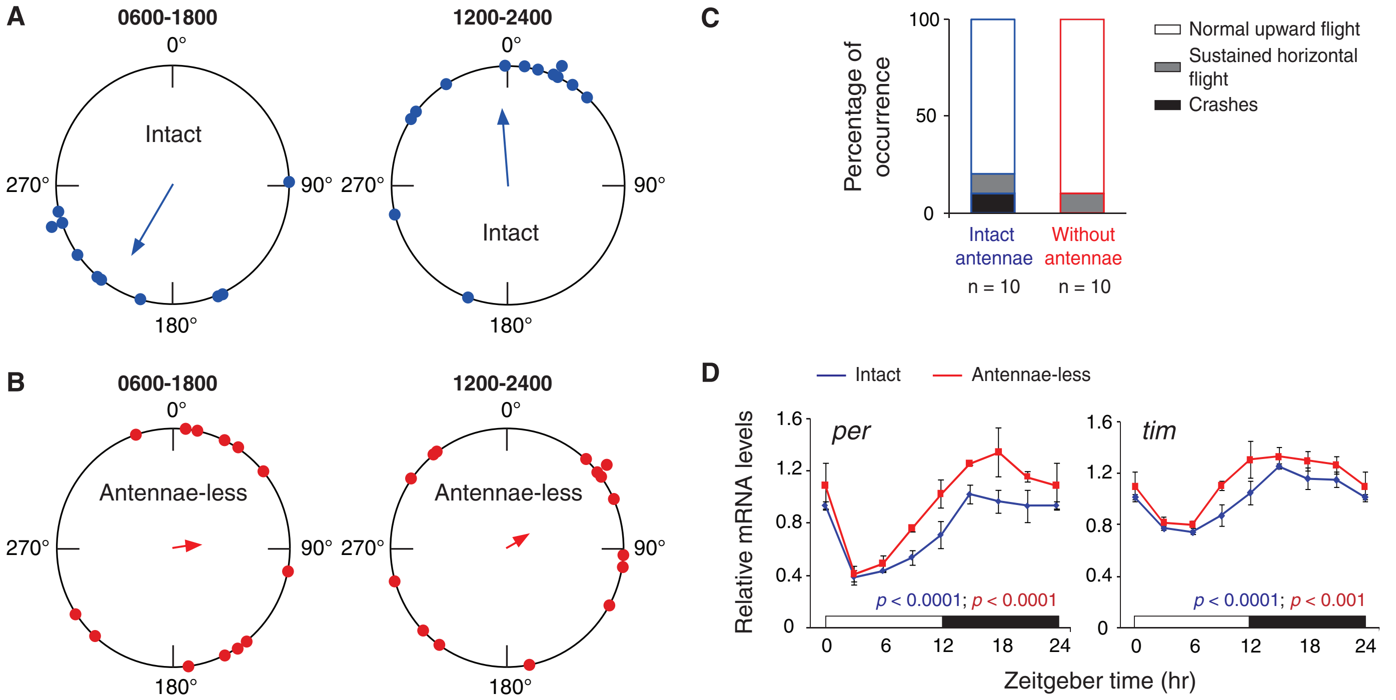


Fig. 1. Antennae are necessary for time-compensated sun compass orientation. (**A**) Flight orientation of intact migrants under different lighting conditions. Butterflies were flown between 1100 and 1500 hours from 24 September to 18 October 2008. The large circle represents the 360° of possible directions (0° is north); small solid circles on the perimeter represent the flight orientation of individual butterflies. The arrow indicates the mean vector; arrow length, *r* value. Left, orientation data of butterflies in LD. Right, orientation data of butterflies in 6-hour-delayed LD. (**B**) Orientation of antennaeless migrants under the different lighting conditions. (**C**) Free-flight behavior of intact migrants (left bar) and those without antennae (right bar). (**D**) Temporal profiles of *per* and *tim* mRNA levels in brains of monarchs with antennae (blue) and without antennae (red). Values are mean ± SEM of three brains. Points at CT0 are replotted at CT24 to show 24-hour trend. Horizontal bars: open, light; black, darkness. *P* values, one-way analysis of variance (ANOVA).