Supplementary Information:
Human Dynamics: The Correspondence Patterns of Darwin and Einstein

Datasets

The databases used in our study were provided by the Darwin Correspondence Project (http://www.lib.cam.ac.uk/Departments/Darwin/) and the Einstein Papers Project (http://www.einstein.caltech.edu/). Each dataset contains the information about each sent/received letter in the following format:

SENDER RECIPIENT DATE

where either the sender or the recipient is Einstein or Darwin. The Darwin dataset contained a record of a total of 7,591 letters sent and 6,530 letters received by Darwin (a total of 14,121 letters). Similarly, the Einstein database contained 14,512 letters sent and 16,289 letters received (total of 30,801). Note that 1,541 letters in the Darwin database and 1,861 letters in the Einstein database were not dated (or were assigned only potential time intervals spanning days or months). We discarded these letters from the dataset. Furthermore, the dataset is naturally incomplete, as not all letters written or received by Darwin and Einstein were kept. Yet, assuming that letters are lost at a uniform rate, they should not affect our main findings.

To analyze Einstein's and Darwin's response time we have followed the following procedure: if individual A sent a letter to Einstein on DATE1, we search for the next letter from Einstein to individual A, sent on DATE2, the response time representing the time difference τ=DATE2-DATE1, expressed in days. If there are multiple letters from Einstein or Darwin to the recipient, we always consider the first letter as the response, and discard the later ones. Missing letters could increase the response time, the magnitude of this effect depending on the overall frequency of communication between the respective correspondence partners. Yet, if the response time does not follow a power law distribution, but rather a distribution with an exponential tail, then randomly distributed missing letters would not generate a power law waiting time—they would only shift the exponential waiting times to longer average values. Thus the observed power law cannot be attributed to the data incompleteness. Finally, note that while in most cases the identified reply is indeed a response to a received letter, there are exceptions as well: many of the much delayed replies represent the renewal of a long lost relationship.

Modeling the observed correspondence pattern

To explain the observed scaling in Darwin and Einstein's communication patterns we use a simple model of mail based communication, which assumes that letters arrive at rate λ following a Poisson process with exponential arrival time distribution. The responses are written at rate μ, reflecting the overall time Einstein or Darwin devoted to their correspondence. Each letter is assigned a discrete priority parameter $x = 1,2,...,r$, such that always the highest priority unanswered letter will be chosen for a reply. The lowest priority task will have to wait the longest before executed, and therefore it dominates the waiting time probability density for large waiting times. This model was introduced in 1964 by Cobham [1] as a queuing model of some manufacturing processes. Since most manufacturing processes focus on the low arrival rate limit, where significant queuing is absent, most work has focused on the average waiting time of the tasks in the queue (i.e. the average time it takes for a task to leave the queue, corresponding to the average time it would take to respond to a letter in our case). Recently the waiting time distribution was also calculated, finding that it follows [2]

$$P(\tau) \sim A\tau^{-3/2} \exp(-\tau/\tau_0) \ , \tag{1}$$

where A and τ_0 are functions of the model parameters. In particular, the characteristic waiting time τ_0 is given by

$$\tau_0 = 1/\left[\mu\left(1-\sqrt{\rho}\right)^2\right] , \tag{2}$$

where $\rho = \lambda/\mu$ is the traffic intensity. Therefore, the waiting time distribution generated by the model is characterized by a power law decay with exponent $\alpha = 3/2$ combined with an exponential cutoff.

The model can be extended to the case where the priorities are not discrete, but take up continuous values from an arbitrary $\eta(x)$ distribution. The Laplace transform of the waiting time distribution for this case has been calculated in Ref. [3], but the resulting equation cannot be inverted, forcing us to study the model numerically (Fig. S1). The natural control parameter is $\rho = \lambda/\mu$, allowing us to distinguish three qualitatively different regimes:

Subcritical regime, $\rho < 1$: Given that the arrival rate of the letters is smaller than the response rate, the queue will be often empty, i.e. there are no letters waiting for a response. This significantly limits the waiting time, most letters being responded soon after their arrival. The simulations indicate that the waiting time distribution exhibits an asymptotic scaling behavior consistent with (1) (Fig. S1), dominated by the exponential decay. As ρ increases to $\rho = 1$ a power law regime with exponent $\alpha = 3/2$ emerges, combined with the exponential cutoff.

Critical regime, $\rho = 1$: When the arrival and the response rate of the letters are equal, according to (1) and (2) we should observe a power law waiting time distribution with $\alpha = 3/2$ for any τ (Fig. S1). This regime would imply that Darwin and Einstein respond to all letters they receive, which is not the case, given that their response rate is 0.32 (D) and 0.24 (E). In this case it is easy to show that the queue length $\ell(t)$ performs a one-dimensional random walk bounded at $\ell = 0$. These fluctuations in the queue length will limit the waiting time distribution, as the tasks will wait at most as long as it takes for the queue length to return to $\ell = 0$. Therefore, the waiting time distribution will have as upper bound the return time distribution of a one-dimensional random walk. It is known, however, that the return time distribution of a random walker follows $P(\tau_{ret}) \sim \tau_{ret}^{-3/2}$ [4], which is the origin of the 3/2 exponent in Eq. (1). This argument indicates that $\alpha = 3/2$ is related to the fluctuations in the length of the priority list.

Supercritical regime, $\rho > 1$: Given that in this regime the arrival rate exceeds the response rate, the average queue length grows linearly as $\langle \ell \rangle = (\lambda - \mu)t$. Therefore, a $1 - 1/\rho$ fraction of the letters is never responded to, waiting indefinitely in the queue. Given Darwin and Einstein's small response rate, this regime captures best their correspondence pattern. We can measure the waiting time for each letter that is responded to. In Fig. S1 we show the waiting time probability density obtained from numerical simulations, indicating that it follows a power law with exponent $\alpha = 3/2$. Thus the supercritical regime follows the same scaling behavior as the critical regime, but only for the letters that are responded to. The rest of the letters wait indefinitely in the list $(\tau = \infty)$. In practice these letters are forgotten and thus removed from the list.

We wish to thank Diana Kormos Buchwald and Tilman Sauer for providing us the dataset capturing the Einstein correspondence, and for helping us understand many aspects of Einstein's life and correspondence patterns. Also, we thank Tilman Sauer for providing us the translation of Einstein's letter to Ralph De Laer Kronig. Similarly, we wish to thank Alison M Pearn from the Darwin Correspondence Project for providing us the record of Darwin's communications. We have also benefited from useful discussions with Alexei Vázquez and Zoltán Dezsö. A-L. B. is supported by NSF ITR 0426737, NSF ACT/SGER 0441089 awards and by an award from the James S. McDonnell Foundation. J.G.O. is supported by FCT (Portugal) grant No. SFRH/BD/14168/2003.

[1] Cobham, A., *J. Oper. Res. Soc. Amer.* **2,** 70–76 (1954).
[2] Abate, J. & Whitt, W. *Queuing Systems* **25,** 173–233 (1997).
[3] Cohen, J. W. *The Single Server Queue* (North Holland, Amsterdam, 1969).
[4] Havlin, S. & Ben-Avraham, D., Diffusion in Disordered Media, *Adv. Phys.* **36,** 695 (1987).

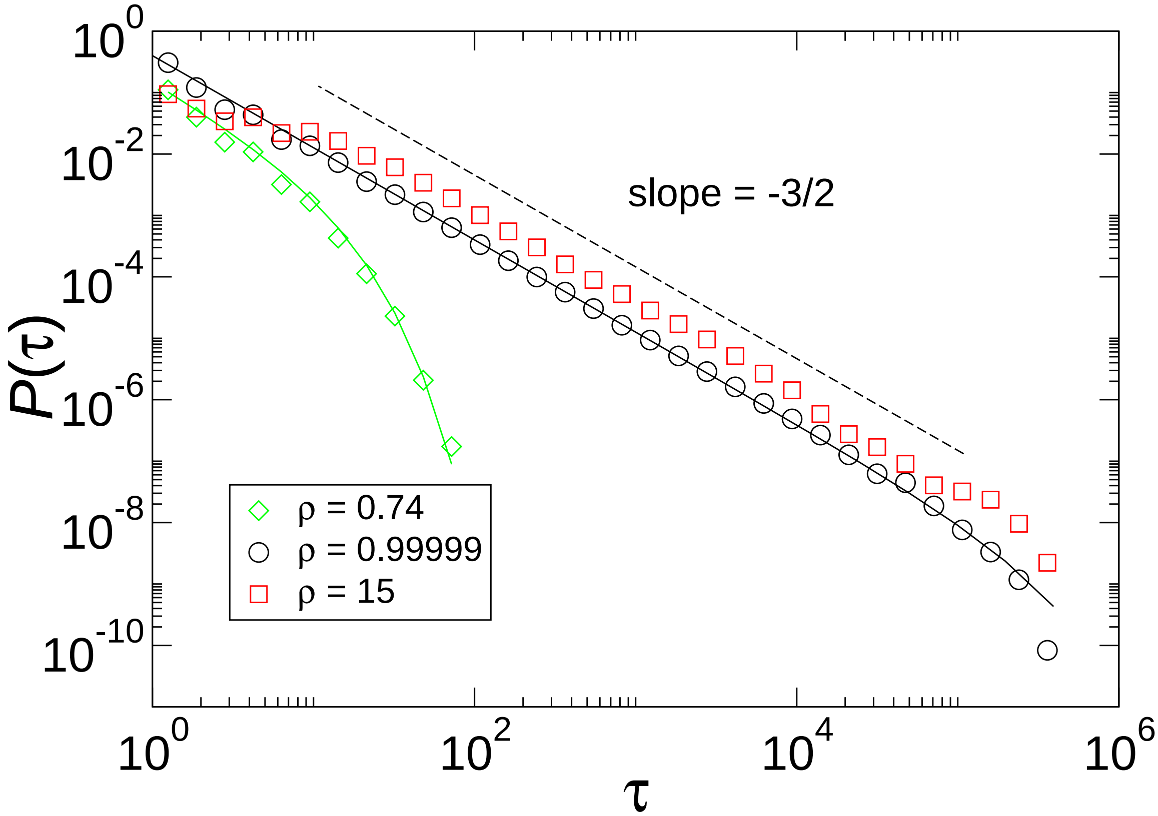

Figure S1: Waiting time distribution for tasks in the discussed model with continuous priorities. The numerical simulations were performed as follows: At each time step (*i*) with probability p a letter arrives to the queue and is assigned a random priority $x \in [0,1]$ chosen from the uniform distribution, increasing the queue length $\ell(t)$ by one. (*ii*) With probability q a letter is responded to and removed from the queue, decreasing the queue length $\ell(t)$ with one. Therefore, p and q play the role of λ and μ, respectively, the relationship between them being $\lambda = \ln[1/(1-p)]$ and $\mu = \ln[1/(1-q)]$. The waiting time distribution is plotted for three $a = p/q$ values, corresponding to the $\rho = \lambda/\mu$ values shown in the figure. While for $\rho << 1$ (subcritical regime) the exponential decay prevails over the power law one, as we approach $\rho = 1$ (critical regime) the power law regime becomes visible, and at $\rho = 1$ the exponential cutoff disappears. The distribution of waiting times for the responded letters has a power law tail even for $\rho >> 1$ (supercritical regime). Note, however, that in this regime a high fraction of letters is never replied to, staying forever on the priority list whose length increases linearly with time. The continuous lines are fits according to Eq. (1).