Evidence for soft bounds in Ubuntu package sizes and mammalian body masses

Marco Gherardi[a,b,1], Salvatore Mandrà[a,b,c], Bruno Bassetti[a,b], and Marco Cosentino Lagomarsino[d,e]

[a]Dipartimento di Fisica, Università degli Studi di Milano, I-20133 Milano, Italy; [b]Istituto Nazionale di Fisica Nucleare, Sezione di Milano, I-20133 Milan, Italy; [c]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138; [d]Genomic Physics Group, Unité Mixte de Recherche 7238 associée au Centre National de la Recherche Scientifique, "Microorganism Genomics," F-75006 Paris, France; and [e]Unité Mixte de Recherche 7238, Université Pierre et Marie Curie, F-75005 Paris, France

Edited by Giorgio Parisi, University of Rome, Rome, Italy, and approved November 13, 2013 (received for review June 18, 2013)

The development of a complex system depends on the self-coordinated action of a large number of agents, often determining unexpected global behavior. The case of software evolution has great practical importance: knowledge of what is to be considered atypical can guide developers in recognizing and reacting to abnormal behavior. Although the initial framework of a theory of software exists, the current theoretical achievements do not fully capture existing quantitative data or predict future trends. Here we show that two elementary laws describe the evolution of package sizes in a Linux-based operating system: first, relative changes in size follow a random walk with non-Gaussian jumps; second, each size change is bounded by a limit that is dependent on the starting size, an intriguing behavior that we call "soft bound." Our approach is based on data analysis and on a simple theoretical model, which is able to reproduce empirical details without relying on any adjustable parameter and generates definite predictions. The same analysis allows us to formulate and support the hypothesis that a similar mechanism is shaping the distribution of mammalian body sizes, via size-dependent constraints during cladogenesis. Whereas generally accepted approaches struggle to reproduce the large-mass shoulder displayed by the distribution of extant mammalian species, this is a natural consequence of the softly bounded nature of the process. Additionally, the hypothesis that this model is valid has the relevant implication that, contrary to a common assumption, mammalian masses are still evolving, albeit very slowly.

bounded diffusion | multiplicative processes | cladogenetic diffusion | macroevolutionary patterns

Software programs are embedded in the real world. As a consequence, the growth of a software package is characterized by inherent adaptive change in response to many factors of different natures. The multilevel feedback structure where programs and their environment evolve in concert is elusive and difficult to describe precisely; quantitative results in this direction are still erratic, despite the efforts made in the past few decades (1, 2). These very features make the subject attractive from the point of view of complex systems theory and analysis. Most of the traditional analyses concerned proprietary software, but a number of studies carried out within the past 10–15 y gathered a relevant amount of evidence concerning the evolution of Open Source Software (OSS) (3–5). The open source phenomenon has two specificities that make it particularly interesting. First, the goal of an open source project is to create a system that is useful or interesting to its developers and thus fills a social void rather than a commercial one. Second, large OSS projects are developed and maintained in a globally decentralized context, contrary to traditional softwarecontrary to traditional software. The emergent complex self-organizing structure challenges traditional theories of management and engineering (6–8). The OSS phenomenon is also affecting the daily lives of increasingly many people, because OSS operating systems and applications run on devices ranging from PCs to mobile phones and tablets.

Perhaps the simplest observable related to software growth is its size, which can be measured with different approaches (9). Despite its simplicity, the size of a piece of software encapsulates many of the features of its evolution and evolvability. Here, we consider the dynamics of package size in a widely used GNU/Linux system, the Debian-based Ubuntu distribution (www.ubuntu.com/project). We analyze systematically the available data and show that they are compatible with a multiplicative anomalous diffusion process. We study this process with the aid of a theoretical model and show that the combination of a "hard" lower cutoff and a more complex size-dependent "soft" upper cutoff on package size reproduces with extreme accuracy the observed distribution. The same model makes definite quantitative predictions for the future dynamics of Ubuntu packages. Finally, as we will see, the knowledge of these evolutionary patterns might lend a fresh perspective to the debate on the quantitative aspects of an a priori unrelated process, the cladogenesis that determines the mass distribution of mammalian species.

Results

Ubuntu Package Sizes. Ubuntu packages are bundled files comprising the pieces of software that make up the whole system. Since Ubuntu was first released in October 2004, the number of packages increased from a few hundred to tens of thousands. Since then, one new release every 6 mo has been issued. This chronological regularity is valuable for a systematic quantitative study. The first, second, and third releases were christened *Warty Warthog*, *Hoary Hedgehog*, and *Breezy Badger*; from then on, the naming followed alphabetical order, encompassing 17 different real and imaginary animals, up to *Quantal Quetzal* (October 2012), the latest release we consider here. Analysis of empirical data for approximately 370,000

Significance

Not unlike a big city, a large software project grows in a complex way, involving many developers and even more users, but a predictive framework to understand these temporal patterns is lacking. We focus on software size and analyze the changes of the Ubuntu open source operating system, finding two quantitative laws. First, growth is driven by changes in scale rather than by addition–subtraction; second, evolution toward larger sizes between two consecutive releases is limited by bounds that depend on the starting size of a package. Strikingly, a stochastic model that implements these two laws is predictive. Finally, we provide evidence that similar principles could be in place for the evolution of body mass in mammals.

Author contributions: M.G., B.B., and M.C.L. designed research; M.G. and S.M. performed research; M.G. analyzed data; and M.G. and M.C.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: marco.gherardi@mi.infn.it.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1311124110/-/DCSupplemental.

changes in package size between all successive Ubuntu releases reveals striking regularity (Fig. 1). The logarithm $\log\delta$ of the multiplicative change $\delta = s'/s$ between the sizes s and s' of a package in consecutive releases appears to follow an "α-stable" distribution, independently of the initial size s and of time (the distribution is centered in $\delta = 1$ and has power law exponent $\alpha \approx 0.7$). α-stable distributions [widely used in many modeling contexts (10–13)] are the most general class of probability distributions followed by the sum of a large number of independent identically distributed random variables (it is therefore a generalization of the Gaussian, which is recovered for $\alpha = 2$). It is interesting to note that the average change in package size is roughly symmetric, implying that packages are generally equally likely to get larger or smaller (so long as they are far from the boundaries).

Notably, events belonging to the tails appear to be bounded in a size-dependent way (Fig. 1). No package can shrink to sizes smaller than a global cutoff $s_{\min}$. This hard bound is easily rationalized by the existence of minimum requirements from the package management system. Consequently, the largest possible decrease, starting from s, is $\delta_{\min} = s_{\min}/s$. Note that a multiplicative diffusion process with a hard lower bound is known to reproduce asymptotically, under certain assumptions, a power law distribution, which is truncated for finite times (14). In our case, the presence of an upper cutoff can modify this dynamic behavior and generate distributions resembling power laws only sufficiently far from the boundaries.

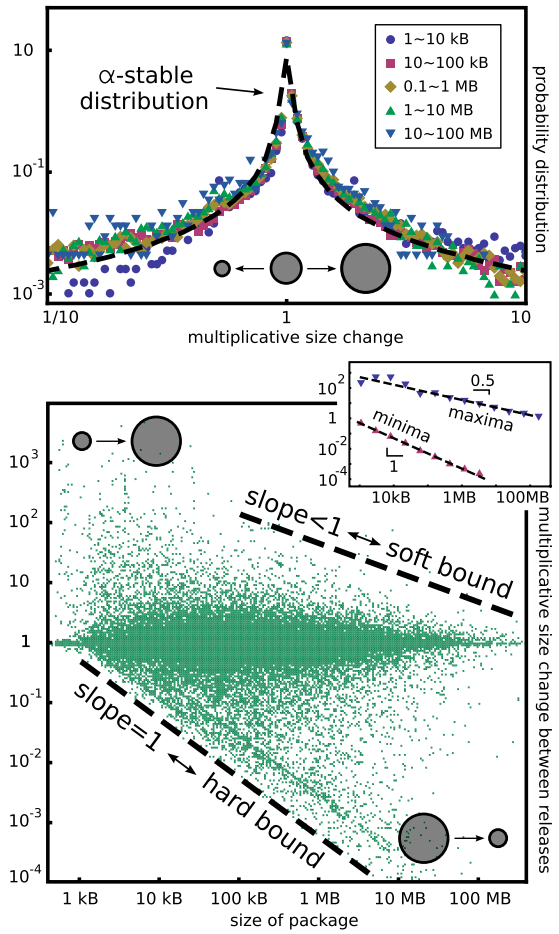

Fig. 1. The changes in package size between Ubuntu releases follow an α-stable distribution with size-dependent bounds. (*Top*) The distribution is independent of size for small multiplicative size changes (x axis; symbols represent different size ranges). (*Bottom*) A scatterplot of multiplicative size change vs. initial package size in the whole range reveals a hard lower bound, due to a minimum attainable size, and a soft upper bound, with a nontrivial size dependence (P values $<10^{-4}$); the signature of a hard bound in such a scatterplot is a slope of modulus 1, whereas anything less steep is named soft. (*Inset*) Binned averages of maximum and minimum size changes (dashed lines are power law fits yielding exponents 0.5 and 1, respectively). The visible linear structure parallel to the hard-bound line is composed of packages whose size dropped to around 4 kb; these are mainly "transitional" packages, i.e., dummy packages only used as pointers to newer versions of the same piece of software. Removing all jumps involving transitional packages (and those they point to) does not affect the results.

Expansion to larger package sizes manifests a more intriguing and complex behavior: the largest size that a package can attain between two consecutive releases depends on its starting size. Specifically, the largest possible increase is $\delta_{\max} = (s_{\max}/s)^{\gamma}$, with an exponent γ approximately equal to 1/2. We call this a soft bound, meaning that the larger a package is, the shorter its maximum jump can be, but packages of different initial sizes do not behave as if a unique maximal size were present. The same behavior is found consistently throughout the history of Ubuntu releases (*SI Appendix S2.B*, Fig. S6). This indicates that the soft-bound behavior cannot be reduced to a time-evolving hard bound caused by extrinsic factors changing in time, such as technological constraints. To simulate the model, one can use rejection sampling to draw a value δ from the bulk jump distribution and then update s with $s' = \delta s$ using the acceptance criteria $s_{\min}/s \le \delta \le (s_{\max}/s)^{\gamma}$. Importantly, a hard bound can be reached in one step from any given size, whereas the maximum $s_{\max}$ in the definition of the soft bound cannot be reached from any initial size. To the best of our knowledge, the phenomenology of such soft bound has no analog in the existing literature (*SI Appendix* provides further evidence supporting the existence of hard and soft bounds).

Based on the foregoing empirical observations, we define a stochastic model of package size evolution, which relies on three assumptions: *i*) At every new release, each package (of size s) assumes the new size $s' = s\delta$ (multiplicative size changes). *ii*) Each package has probability q of also "duplicating", i.e., branching and adding a "spinoff" copy of itself to the new release [This move has no impact on size distributions (*SI Appendix S1.C*) but is included for completeness, as code reuse appears to be the driving force of innovation (*Discussion* and *SI Appendix S2.B*)]. *iii*) The logarithms of the growth factors δ are independent α-stable random variables conditioned on two size-dependent cutoffs, a lower hard bound and an upper soft bound, whose parameters $s_{\min}$, $s_{\max}$, and γ are obtained from the data. This model has no free parameters, as all of the quantities needed to specify the distribution are estimated by data analysis. Technically, it is realized as a branching multiplicative diffusion process. We do not explicitly consider package deletion, as its role for the evolution of package size distributions is irrelevant (*SI Appendix S1.C*).

Starting from the population of packages in the first Ubuntu release, *Warty*, and evolving their sizes for 16 steps (8 y), the model predicts very accurately the package size distribution in the latest release, *Quantal* (Fig. 2). Sensitivity analysis shows (*SI Appendix S2.C*, Fig. S7) that the results are robust with respect to variation of the parameters. Moreover, as shown by Fig. 3, the accordance of model and data are not dependent on the particular initial shape of the distribution; in fact, arbitrarily chosen subsets of packages can be followed through their evolution, and the size proportions they assume in *Quantal* are predicted very well by the model (*SI Appendix S2.D*). In particular, the plots in Fig. 3 show that the model is able to capture accurately the time course of divergence of initially similarly sized packages over the whole period of 8 y. This also shows that the agreement between model and data is not an accident due to specific behavior of the packages found at the distribution tails. It is then appealing to attempt to forecast future evolution. For instance, we find that the current distribution is

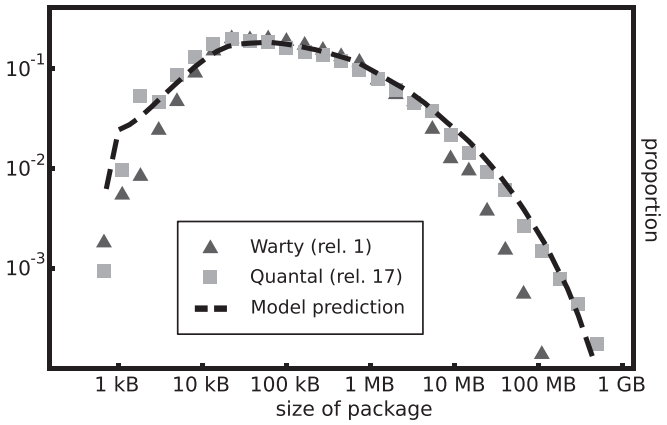

Fig. 2. The dynamics of package size distribution is captured by a bounded multiplicative diffusion process. Starting from the initial pool of packages constituting *Warty Warthog* (▲), with all parameters fixed by data analysis, the model yields the distribution traced by the dashed line, which nicely reproduces size proportions in *Quantal Quetzal* (■). Notice that the tails of the two empirical distributions differ by almost one order of magnitude; furthermore, the ramp at small sizes for *Quantal* (which was not present in *Warty*) is correctly predicted.

very far from stationary; at this rate, assuming constant parameters, a stationary state would be reached in ∼2–400 y (*SI Appendix S2.D*, Fig. S11). In 10 y the largest package should weigh ∼1 Gb, and the average package size is predicted to nearly double from the current 1.2 Mb to about 2.3 Mb; the most common size, instead, will have slightly increased only by around 10 kb (it is currently 22 kb).

Mammalian Body Masses. We found that the knowledge of the modeling framework with soft bounds described above may suggest a different perspective on the debate around a distant scientific problem. In fact, similar models to the one described here have been used to explain the evolution of species body masses in mammals and other taxa (15, 16). In this case, the branching process represents cladogenesis, i.e., the lineage splitting event generating new species (clades in the phylogenetic tree) whose average body mass is related to the ancestor's. A simple scaling form recently discovered for intraspecific size variability (17) justifies the use of the mean species mass as the sole relevant variable. The model proposed by Clauset and Erwin (16) [and further developed in subsequent publications (18, 19)] assumes multiplicative diffusion on evolutionary time scales, with a lower hard bound due to metabolic constraints and an explicit bias toward larger sizes [the controversial Cope's rule (20–22)], whose strength must increase for lower masses [although there appears to also be evidence for the opposite tendency (15)]. Moreover, the introduction of a size-dependent extinction rate is necessary to approximate the large-mass tail of the empirical distribution of extant mammals.

In the framework suggested by software evolution, it seems natural to characterize the low propensity of large species to generate larger descendant species (and the tendency of small species to generate larger ones) through a soft, i.e., size-dependent, cutoff instead. Fossil data of ancestor–descendant size ratios are not abundant and are susceptible to noise and bias (23). We used a compilation by Alroy (15) of 1,109 North American terrestrial mammals up to the late Pleistocene, obtained by a highly conservative method. Despite the great amount of work behind these data, they do not allow an estimate of parameters nearly as precise as what was attained for Ubuntu packages; nonetheless, our analysis shows that the changes in body size are compatible with an α-stable distribution of exponent $\alpha \approx 1.8$ and with upper and lower soft cutoffs with γ-values around 0.2 and 0.6, respectively (Fig. 4 and *SI Appendix S2.E*). Furthermore, uncertainties on these estimates are not a big inconvenience, as the results are fairly robust to variation of these parameters (*SI Appendix S2.F*, Fig. S13). Note that the exponent α in this case takes a very different value than the one observed for Ubuntu packages.

We simulated the in silico evolution of body masses throughout mammalian history, starting from the mass of the founder species *Hadrocodium wui*, a small mammaliaform from the Early Jurassic weighing 2 g (24). Remarkably, the characteristically skewed and wide distribution of extant terrestrial mammals (25) is recovered with good precision by this model (Fig. 4). The (softly) bounded nature of the diffusion, together with the asymmetry of the initial condition, are the key ingredients that account for the shape of the empirical distribution (*SI Appendix S2.G*, Fig. S15). It must be said that the agreement is not completely parameter-free as in the case of Ubuntu packages: model time is chosen as the one that best recovers the expected distribution, because it cannot be estimated directly. However, one or more free parameters were present also in the previous studies (16, 18).

Discussion

To sum up, the analysis allows us to uncover two relevant quantitative laws. First, package sizes vary following a process driven by changes in scale, rather than by addition–subtraction. Similar behavior, with an α-stable distribution for the jumps, has been observed in other systems, e.g., related to economics (26), but it is not to be expected a priori. Second, and more important, evolution toward larger sizes is such that the largest change that a package can attain in an elementary update depends on its

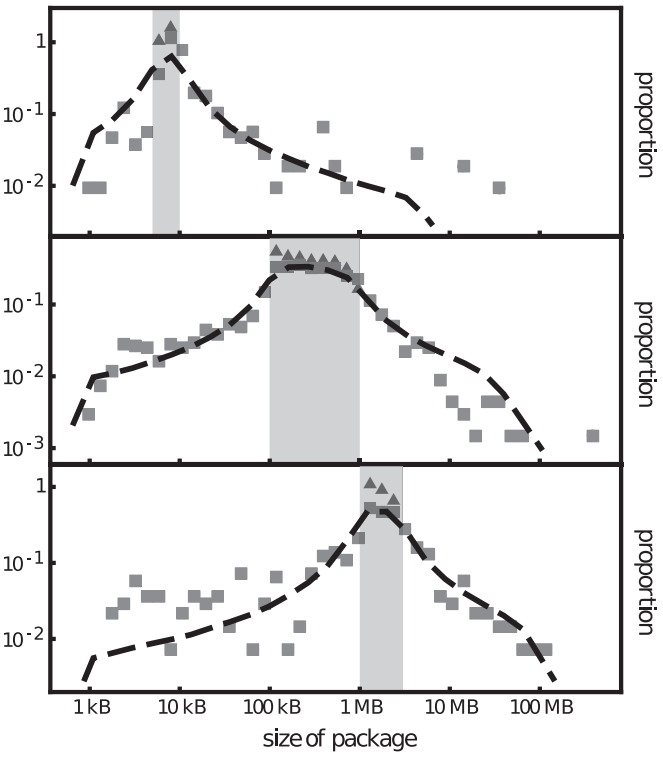

Fig. 3. The model accurately predicts the divergence of similarly sized packages. By starting from a subset of packages in *Warty*, defined by three different size intervals (shaded areas, actual size distributions are represented by ▲), the model generates the distributions shown by dashed lines, which agree well with those observed empirically for the same subsets of packages in *Quantal* (■).

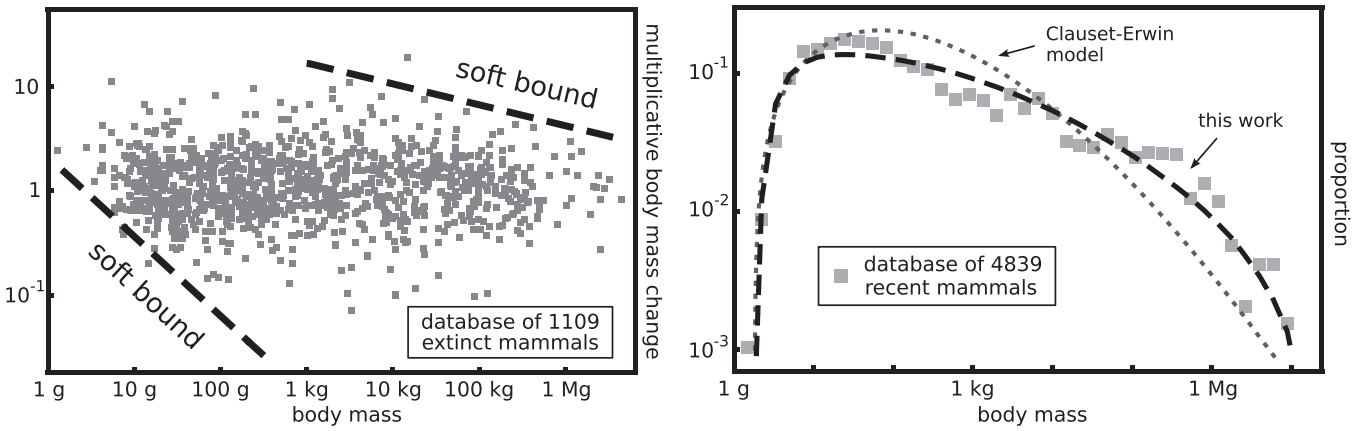

Fig. 4. Application of the bounded diffusion framework to mammalian body-mass data. (*Left*) Multiplicative changes in body mass for 1,109 mammalian species, plotted as a function of ancestor's body mass (■); data are compatible with the existence of soft bounds (dashed lines). (*Right*) The distribution of mammalian body masses is well reproduced by the model (dashed line), and some features appear to be improved with respect to the Clauset–Erwin model (dotted line). Note that the dotted line corresponds to a stationary state, whereas the dashed line does not.

starting size (as a power law), the soft bound. A third instructive result is that the two above laws, implemented in an otherwise fully stochastic model, are sufficient to define a statistical predictive framework for Ubuntu package size changes. The upper cutoffs on size jumps and their soft nature appear to have no counterpart in the previous literature. Furthermore, the distribution of the size changes is precisely estimated from data, under the sole assumption that they are independent (which is also suggested by the data).

This phenomenology casts a quantitative light on the laws by which software packages expand and contract, in the spirit of earlier investigations by Lehman and coworkers (1). The relevant quantities necessary to capture the evolution of size are size ratios rather than size differences. This suggests that the dominant route of expansion is the forking and reuse of submodules, with new code being largely produced by copying and modifying old code. Birth of new packages, also a relevant driving process for the dynamics of software evolution, supports this interpretation: newborn packages appear with size proportions approximately equal to those of the preceding release (*SI Appendix S2.B*, Fig. S4).

In another area of human interactions, namely the evolution of business firms' sizes, multiplicative processes are also found to emerge from the dynamics of smaller modules (26, 27). Variants of existing models developed in this context (28, 29) might help provide a microscopic interpretation for the bulk of the size-change distribution. For instance, we can consider the following heuristic argument. Let us describe a piece of software through the dependency network of its components (e.g., modules or classes). We suppose that a node changing its size by a factor λ propagates the need for maintenance to its k (direct and indirect) dependencies, resulting in a similar size change. Then the effect of this cascade of events for the whole package would be summarized by the jump $\log\delta = \log[(s-k+\lambda k)/s] = \log[1+(\lambda-1)k/s]$, which can be approximated by $(\lambda-1)k/s$, when the latter is small. Therefore, assuming that λ is a sufficiently compact random variable centered around 1, the distribution of $\log\delta$ will resemble that of k. A mechanism driving the evolution of dependency networks (software in particular) has been recently proposed (8), based on a simple process where new nodes attach to a fixed number D of existing nodes. The distribution of k in this case is $P(k)\sim k^{-(1+1/D)}$, which implies a power law-distributed $\log\delta$, with exponent $1+\alpha = 1+1/D$. Note that the value $D\approx 2$, which seems to be ubiquitous in dependency networks (8), yields $\alpha\approx 0.5$, not far from the observed $\alpha\approx 0.7$.

On the other hand, this framework does not seem to be able to account for the soft bounds. We speculate that the emergence of the soft bound could be related to allometric scaling (30, 31), where the system size and its network of dependency grow jointly and are subject to global constraints. Therefore, a more elaborate microscopic model would be needed to account for this behavior. Rather than including detailed code-production mechanisms, the approach taken here assumes that their effect on intermediate time scales can be summarized by two elementary processes. First, the generation of new packages proceeds by copy and modification of old packages; second, packages evolve under a simple constraint of minimum size and a complex constraint for large sizes. Finally, the explanation of the observed $\gamma \simeq 0.5$ exponent for the upper soft bound remains an open question. We speculate that tradeoffs between increase in complexity and cost of deployment might be responsible for this law.

Regarding the application to mammalian body masses, one important remark is that the present model relaxes the common assumption that the body-mass distribution is stationary at present time. Consequently, different initial conditions can produce markedly different distributions. If the initial mass is sufficiently large, left-skewed distributions can be obtained; such a shape is less common but is nonetheless found in some taxa (32). Note, however, that the fitting distribution for mammals is very nearly stationary (*SI Appendix*, Fig. S14). Mining the literature, we were not able to find any conclusive evidence that could rule out a mild nonstationarity of the extant distribution, and therefore we hope that our findings may be useful to stimulate the debate in this direction.

A second remark is that the bounds on the diffusion process in the context of mammalian body masses are realized by a size-dependent extinction rate (23). In our approach, the soft nature of the constraint for large masses is interpreted as the result of the competition between the short-term selective advantages of an increased body size and the corresponding long-term extinction risk, as concluded in previous studies. This macroevolutionary tradeoff mechanism is quantitatively robust across all mammalian species (33) and also in other taxa (19). The observation that the lower boundary is soft as well suggests that a similar tradeoff might be present also for small body masses.

As already stated above, the soft-bound mechanism implies that larger masses require a higher number of generations to be reached, whereas the lower bound can be reached in a single step from any mass. This pattern has two notable consequences. First, it qualitatively predicts a macroevolutionary asymmetry between large increases and large decreases, while preserving the symmetry for small size changes, a phenomenon that has been

recently observed for mammals (34). Second, it accounts for a slowly saturating evolution of the maximum body mass as a function of time, which is quantitatively in line with recent findings (35) (*SI Appendix S2.H*). Finally, we note that a reasonable reparameterization of the bounds is sufficient to recover the body-mass distribution of fully aquatic mammals as well (*SI Appendix S2.I*).

ACKNOWLEDGMENTS. We are grateful to Aaron Clauset for discussions and help with the Alroy dataset; and to Amos Maritan, Miguel Fortuna, Alberto Vailati, Vincenzo Gino Benza, Felisa Smith, Laurence Hurst, Kunihiko Kaneko, Michele Caselle, and Matteo Osella for exchanges and discussions. M.G. acknowledges financial support from Fondo Sociale Europeo (Regione Lombardia), through the grant "Dote Ricerca." S.M. also acknowledges the Air Force Office of Scientific Research (prime sponsor) and the University of California, San Diego, for partial support to this work under Grants FA9550-12-1-0046 and 10323836-SUB.

1. Lehman M, Ramil J (2003) Software evolution—background, theory, practice. *Inf Process Lett* 88(1):33–44.
2. Mens T, Demeyer S, eds (2008) *Software Evolution* (Springer, Berlin).
3. Ermann L, Chepelianskii AD, Shepelyansky DL (2011) Fractal Weyl law for Linux kernel architecture. *Eur Phys J B* 79(1):115–120.
4. Maillart T, Sornette D, Spaeth S, von Krogh G (2008) Empirical tests of Zipf's law mechanism in open source Linux distribution. *Phys Rev Lett* 101(21):218701.
5. Godfrey M, Tu Q (2000) Evolution in open source software: A case study. *Proceedings of the International Conference on Software Maintenance* (IEEE Computer Society, Alamitos, CA), pp 131–142.
6. Madey G, Freeh V, Tynan R (2002) The open source software development phenomenon: An analysis based on social network theory. *Americas Conference on Information Systems* (Association for Information Systems), pp 1806–1813.
7. Fortuna MA, Bonachela JA, Levin SA (2011) Evolution of a modular software network. *Proc Natl Acad Sci USA* 108(50):19985–19989.
8. Pang TY, Maslov S (2013) Universal distribution of component frequencies in biological and technological systems. *Proc Natl Acad Sci USA* 110(15):6235–6239.
9. Kemerer C, Slaughter S (1999) An empirical approach to studying software evolution. *IEEE Trans Softw Eng* 25(4):493–509.
10. Mandelbrot B (1983) *The Fractal Geometry of Nature* (W.H. Freedman and Co., New York).
11. Mantegna R, Stanley H (1995) Scaling behaviour in the dynamics of an economic index. *Nature* 376:46–49.
12. Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. *Nature* 439(7075):462–465.
13. Barthelemy P, Bertolotti J, Wiersma DS (2008) A Lévy flight for light. *Nature* 453(7194): 495–498.
14. Sornette D, Cont R (1997) Convergent multiplicative processes repelled from zero: Power laws and truncated power laws. *J Phys I France* 7(3):431–444.
15. Alroy J (1998) Cope's rule and the dynamics of body mass evolution in North American fossil mammals. *Science* 280(5364):731–734.
16. Clauset A, Erwin DH (2008) The evolution and distribution of species body size. *Science* 321(5887):399–401.
17. Giometto A, Altermatt F, Carrara F, Maritan A, Rinaldo A (2013) Scaling body size fluctuations. *Proc Natl Acad Sci USA* 110(12):4646–4650.
18. Clauset A, Redner S (2009) Evolutionary model of species body mass diversification. *Phys Rev Lett* 102(3):038103.
19. Clauset A, Schwab DJ, Redner S (2009) How many species have mass M? *Am Nat* 173(2):256–263.
20. Cope E (1887) *The Origin of the Fittest* (Appleton, New York).
21. Gould S (1997) Cope's rule as psychological artefact. *Nature* 385(6613):199–200.
22. Van Valkenburgh B, Wang X, Damuth J (2004) Cope's rule, hypercarnivory, and extinction in North American canids. *Science* 306(5693):101–104.
23. Liow LH, et al. (2008) Higher origination and extinction rates in larger mammals. *Proc Natl Acad Sci USA* 105(16):6097–6102.
24. Luo ZX, Crompton AW, Sun AL (2001) A new mammaliaform from the early Jurassic and evolution of mammalian characteristics. *Science* 292(5521):1535–1540.
25. Smith F, et al. (2003) Body mass of late Quaternary mammals. *Ecology* 84(12):3403.
26. Stanley M, et al. (1996) Scaling behavior in the growth of companies. *Nature* 379(6568):804–806.
27. Fu D, et al. (2005) The growth of business firms: Theoretical framework and empirical evidence. *Proc Natl Acad Sci USA* 102(52):18801–18806.
28. Amaral LAN, et al. (1997) Scaling behavior in economics: I. Empirical results for company growth. *J Phys I France* 7(4):621–633.
29. Yan K-K, Fang G, Bhardwaj N, Alexander RP, Gerstein M (2010) Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. *Proc Natl Acad Sci USA* 107(20):9186–9191.
30. West GB, Brown JH, Enquist BJ (1997) A general model for the origin of allometric scaling laws in biology. *Science* 276(5309):122–126.
31. Banavar JR, Maritan A, Rinaldo A (1999) Size and form in efficient transportation networks. *Nature* 399(6732):130–132.
32. Kozłowski J, Gawelczyk A (2002) Why are species' body size distributions usually skewed to the right? *Funct Ecol* 16(4):419–432.
33. Clauset A (2013) How large should whales be? *PLoS ONE* 8(1):e53967.
34. Evans AR, et al. (2012) The maximum rate of mammal evolution. *Proc Natl Acad Sci USA* 109(11):4187–4190.
35. Smith FA, et al. (2010) The evolution of maximum body size of terrestrial mammals. *Science* 330(6008):1216–1219.