Mathematical Modeling of Biological and Social Phases of Big History

Andrey V. Korotayev and Alexander V. Markov

The present article demonstrates that changes in biodiversity through the Phanerozoic correlate with a hyperbolic model (widely used in demography and macrosociology) much more strongly than with exponential and logistic models (traditionally used in population biology and extensively applied to fossil biodiversity as well). The latter models imply that changes in diversity are guided by a first-order positive feedback (more ancestors – more descendants) and/or a negative feedback arising from resource limitation. The hyperbolic model implies a second-order positive feedback. The authors demonstrate that the hyperbolic pattern of the world population growth arises from a second-order positive feedback between the population size and the rate of technological growth (this can also be identified with the collective learning mechanism). The feedback between the diversity and community structure complexity can also contribute to the hyperbolic character of biodiversity. This suggests that some mechanisms vaguely resembling the collective learning might have operated throughout the biological phase of Big History. Our findings suggest that we can trace rather similar macropatterns within both the biological and social phases of Big History which one can describe in a rather accurate way with very simple mathematical models.

Keywords: biological phase of Big History, social phase of Big History, mathematical modeling, collective learning, positive feedback, biodiversity, demography, sociology, paleontology, geology, hyperbolic growth.

In 2005, in the town of Dubna, near Moscow, at what seems to have been the first ever international conference devoted specifically to Big History studies, the two authors of the present article – sociologist/anthropologist Andrey Korotayev and biologist/paleontologist Alexander Markov – one after another demonstrated two diagrams.1 One of those diagrams illustrated the dynamics of the population of China between 700 BCE and 1851 CE, the other illustrated the dynamics of marine Phanerozoic biodiversity during the last 542 million years (see Fig. 1):

---

1 This research has been supported by the Russian Foundation for Basic Research (Project # 13-06-00501).

1 We would like to emphasize that we saw each other at that session for the first time, so we had no chance to arrange in advance the demonstration of those two diagrams.

Globalistics and Globalization Studies 2015 319–343
Nevertheless, one can hardly ignore the striking similarity between two diagrams depicting the development of rather different systems (human population, on the one hand, and biota, on the other) at different time scales (hundreds of years, on the one hand, and millions of years, on the other) studied by different sciences (Historical Demography, on the one hand, and Paleontology, on the other) using different sources (demographic estimates, on the one hand, and paleontological chronicles, on the other hand). What are the causes of this similarity in the development dynamics of rather different systems?
In 1960, von Foerster, Mora, and Amiot published a striking discovery in the journal *Science*. They showed that between 1 and 1958 CE the world’s population ($N$) dynamics can be described in an extremely accurate way with an astonishingly simple equation:

$$N_t = \frac{C}{t_0 - t}, \quad (1)$$

where $N_t$ is the world population at time $t$, and $C$ and $t_0$ are constants, with $t_0$ corresponding to an absolute limit (‘singularity’ point) at which $N$ would become infinite.

Of course, von Foerster and his colleagues did not imply that one day the world population would actually become infinite. The real implication was that prior to 1960 the world population growth for many centuries had followed a pattern which was about to come to an end and to transform into a radically different pattern. Note that this prediction started to come true only a few years after the ‘Doomsday’ paper had been published, because after the early 1970s the World System growth in general (and world population growth in particular) began to diverge more and more from the blow-up regime, and now it is not hyperbolic any more with its pattern being closer to a logistic one (see, e.g., Korotayev, Malkov, and Khaltourina 2006a, where we present a compact mathematical model that describes both the hyperbolic development of the World System in the period prior to the early 1970s, and its withdrawal from the blow-up regime in the subsequent period; see also Korotayev 2009).

Parameter $t_0$ was estimated by von Foerster and his colleagues as 2026.87, which corresponded to November 13, 2006; this allowed them to give their article an attractive and remarkable title – ‘Doomsday: Friday, 13 November, A.D. 2026’.

The overall correlation between the curve generated by the von Foerster equation and the most detailed series of empirical estimates looks as follows (see Fig. 2).

\footnote{To be exact, the equation proposed by von Foerster and his colleagues looked as follows: $N_t = \frac{C}{(t_0 - t)^{0.99}}$. However, as von Hoerner (1975) and Kapitza (1999) showed, it can be simplified as $N_t = \frac{C}{t_0 - t}$.}
Fig. 2. Correlation between empirical estimates of world population (in millions, AD 1000–1970) and the curve generated by the von Foerster equation

Note: black markers correspond to empirical estimates of the world population by McEvedy and Jones (1978) for the interval between 1000 and 1950 and the U.S. Bureau of the Census (2014) for 1950–1970. The grey curve has been generated by the von Foerster equation (1).

The formal characteristics are as follows: $R = 0.998; R^2 = 0.996; p = 9.4 \times 10^{-17}$ $\approx 1 \times 10^{-16}$.

For readers unfamiliar with mathematical statistics we can explain that $R^2$ can be regarded as a measure of the fit between the dynamics generated by a mathematical model and the empirically observed situation, and can be interpreted as the proportion of the variation accounted for by the respective equation. Note that 0.996 also can be expressed as 99.6 per cent. Thus, the von Foerster equation accounts for an astonishing 99.6 per cent of all the macrovariation in the world population, from 1000 CE through 1970, as estimated by McEvedy and Jones (1978) and the U.S. Bureau of the Census (2014).

Note also that the empirical estimates of world population align in an extremely accurate way along the hyperbolic curve, which convincingly justifies the designation of the pre-1970s world population growth pattern as ‘hyperbolic’.

To start with, the von Foerster equation $N_t = \frac{C}{t_0 - t}$ is just a solution of the following differential equation (see, e.g., Korotayev, Malkov, Khaltourina 2006a: 119–20):

---

3 The second characteristic $(p)$, standing for ‘probability’, is a measure of the correlation's statistical significance. A bit counter-intuitively, the lower the value of $p$, the higher the statistical significance of the respective correlation. This is because $p$ indicates the probability that the observed correlation could be accounted solely by chance. Thus, $p = 0.99$ indicates an extremely low statistical significance, as it means that there are 99 chances out of 100 that the observed correlation is the result of a coincidence, and, thus, we can be quite confident that there is no systematic relationship (at least, of the kind that we study) between the two respective variables. On the other hand, $p = 1 \times 10^{-16}$ indicates an extremely high statistical significance for the correlation, as it means that there is only one chance out of $10000000000000000$ that the observed correlation is the result of pure coincidence (in fact, a correlation is usually considered as statistically significant with $p < 0.05$).

4 In fact, with slightly different parameters ($C = 164890.45; t_0 = 2014$) the fit ($R^2$) between the dynamics generated by the von Foerster equation and the macrovariation of world population for CE 1000–1970 as estimated by McEvedy and Jones (1978) and the U.S. Bureau of the Census (2014) reaches 0.9992 (99.92 per cent), whereas for 500 BCE – 1970 CE this fit increases to 0.9993 (99.93 per cent) (with the following parameters: $C = 171042.78; t_0 = 2016$).
This equation can be also written as:
\[ \frac{dN}{dt} = aN^2, \]  
(3)
where \( a = \frac{1}{C} \).

What is the meaning of this mathematical expression, \( \frac{dN}{dt} = aN^2 \)? In our case, \( dN/dt \) denotes an absolute population growth rate at a certain moment of time. Thus, this equation shows that at any moment of time an absolute population growth rate should be proportional to the square of population at this moment.

Note that this significantly demystifies the problem of the world population hyperbolic growth. Now to explain this hyperbolic growth, we should just explain why for many millennia the absolute rate of world population growth tended to be proportional to the square of population.


1) ‘the Malthusian (1978 [1798]) assumption that population is limited by the available technology, so that the growth rate of population is proportional to the growth rate of technology’ (Kremer 1993: 681–682). This statement seems rather convincing. Indeed, throughout most of human history the world population was limited by the technologically determined ceiling of land carrying capacity. For example, with foraging subsistence technologies the Earth could hardly support more than 8 million people, because the amount of naturally available useful biomass on the planet is limited, and the world population could overgrow this limit only when people started to apply various means to artificially increase the amount of available biomass, that is with a transition from foraging to food production. However, the extensive agriculture can only support a limited number of people, and world population further growth became possible only with the intensification of agriculture and other technological improvements (see, e.g., Turchin 2003; Korotayev, Malkov, and Khaltourina 2006a, 2006b; Korotayev and Khaltourina 2006).

However, it is well known that the technological level is not a constant, but a variable (see, e.g., Grinin 2007a, 2007b, 2012). And in order to describe its dynamics the second basic assumption is employed:

2) ‘High population spurs technological change because it increases the number of potential inventors…’ In a larger population there will be proportionally more people lucky or smart enough to come up with new ideas’ (Kremer 1993: 685), thus, ‘the growth rate of
technology is proportional to total population. In fact, here Kremer uses the main assumption of the Endogenous Technological Growth theory (Kuznets 1960; Grossman and Helpman 1991; Aghion and Howitt 1998; Simon 1977, 2000; Komlos and Nefedov 2002; Jones 1995, 2005, etc.). To our knowledge, this supposition was first put forward by Simon Kuznets (1960), so we will denote a corresponding type of dynamics as ‘Kuznetsian’, while the systems in which the ‘Kuznetsian’ population-technological dynamics combines with the ‘Malthusian’ demographic one will be denoted as ‘Malthusian-Kuznetsian’. In general, we find this assumption rather plausible – in fact, it is quite probable that, ceteris paribus, within a given period of time, a billion people will make approximately a thousand times more inventions than a million people.

This assumption was expressed by Kremer mathematically in the following way:

\[
\frac{dT}{dt} = kNT
\]  

Actually, this equation just says that the absolute technological growth rate at a given moment of time \((dT/dt)\) is proportional to the technological level \((T)\) observed at this moment (the wider is the technological base, the more inventions could be made on its basis), and, on the other hand, it is proportional to the population \((N)\) (the larger the population, the larger the number of potential inventors).8

The resultant models provide a rather convincing explanation of why throughout most of human history the world population followed the hyperbolic pattern with an absolute population growth rate tending to be proportional to \(N^2\). For example, why would the growth of population from, say, 10 million to 100 million, result in the hundredfold growth of \(dN/dt\)? The above mentioned models explain this rather convincingly. The point is that the growth of world population from ten to a hundred million implies that human subsistence technologies also grew approximately ten times (given that it will prove, after all, to be able to support a ten times larger population). On the other hand, the tenfold population growth also implies a tenfold growth of the number of potential inventors, and, consequently, a tenfold increase in a relative technological growth rate. Hence, the absolute technological growth rate would grow 10 \times 10 = 100 times (as Equation 4 shows that an order of magnitude larger number of people with an order of magnitude broader technological basis would likely make two orders of magnitude more inventions). And as throughout the Malthusian epoch the world population \((N)\) tended to the technologically determined carrying capacity ceiling of the Earth, we have good reason to expect that \(dN/dt\) will also grow just about 100 times.

In fact, one can demonstrate (see, e.g., Korotayev, Malkov, and Khaltourina 2006a, 2006b; Korotayev and Khaltourina 2006) that the hyperbolic pattern of the world's population growth can be explained by the nonlinear second order positive feedback mechanism that was shown long ago to generate just the hyperbolic growth, known also as the ‘blow-up regime’ (see, e.g., Kurdyumov 1999). In our case this nonlinear second order positive feedback looks as follows: more people – more potential inventors – a faster technological growth – a faster growth of the Earth's carrying capacity – a faster population growth – with more people you also have more potential inventors – hence, faster technological growth, and so on (see Fig. 3).

---

7 Note that ‘the growth rate of technology’ means here the relative growth rate (i.e. the level to which technology will grow in a given unit of time in proportion to the level observed at the beginning of this period).

8 Kremer did not test this hypothesis empirically in a direct way. Note, however, that our own empirical test of this hypothesis has supported it (see Korotayev, Malkov, Khaltourina 2006b: 141–146).
Note that the relationship between technological development and demographic growth cannot be analyzed through any simple cause-and-effect model, as we observe a true dynamic relationship between these two processes — each of them is both the cause and the effect of the other.

Note also that the process discussed above should be identified with the process of collective learning (on the notion of ‘collective learning’ see first of all Christian 2005: 146–148; see also David Christian’s and David Baker’s contributions to the present volume). Respectively, the mathematical models of the World System development discussed in this article can be interpreted as mathematical models of the influence of collective learning on the global social evolution. Thus, a rather peculiar hyperbolic shape of the acceleration of the global development observed prior to the early 1970s may be regarded just as a product of the global collective learning. Elsewhere we have also shown (Korotayev, Malkov, and Khaltourina 2006a: 34–66) that for the period prior to the 1970s the World System economic and demographic macrodynamics driven by the above mentioned positive feedback loops can be described mathematically in a rather accurate way with the following extremely simple mathematical model:

\[
\frac{dN}{dt} = aSN, \tag{5}
\]
\[
\frac{dS}{dt} = bNS, \tag{6}
\]
while the world GDP \(G\) can be calculated using the following equation:

\[
G = mN + SN, \tag{7}
\]

where \(G\) is the world GDP, \(N\) is population, and \(S\) is the produced surplus per capita, over the subsistence amount \(m\) that is minimally necessary to reproduce the population with a zero growth rate in a Malthusian system (thus, \(S = g - m\), where \(g\) denotes per capita GDP); \(a\) and \(b\) are parameters.

Note that the mathematical analysis of the basic model (not presented here) suggests that up to the 1970s the amount of \(S\) (per capita surplus produced at the given level of World System development) should be proportional, in the long run, to the World System's population: \(S = kN\). Our statistical analysis of the available empirical data has confirmed this theoretical proportionality (Korotayev, Malkov, and Khaltourina 2006a: 49–50). Thus, in the right-hand side of equation (6) \(S\) can be replaced with \(kN\), and as a result we arrive at the following equation:

\[
\frac{dN}{dt} = kaN^2. \tag{3}
\]
As we remember, the solution of this type of differential equations is
\[ N_t = \frac{C}{(t_0 - t)}, \]
and this produces simply a hyperbolic curve.

As, according to our model, \( S \) can be approximated as \( kN \), its long-term dynamics can be approximated with the following equation:
\[ S = \frac{kC}{t_0 - t}. \]
Thus, the long-term dynamics of the most dynamic component of the world GDP, \( SN \), ‘the world surplus product’, can be approximated as follows:
\[ SN = \frac{kC^2}{(t_0 - t)^2}. \]
Of course, this suggests that the long-term world GDP dynamics up to the early 1970s must be approximated better by a quadratic hyperbola than by a simple one; and, as we could see below (see Fig. 4), this approximation works very effectively indeed:

![Fig. 4. World GDP Dynamics, 1–1973 CE (in billions of 1990 international dollars, PPP): the fit between predictions of a quadratic-hyperbolic model and the observed data](image)

Note: \( R = .9993, R^2 = .9986, p << .0001 \). The black markers correspond to Maddison’s (2001) estimates (Maddison’s estimates of the world per capita GDP for 1000 CE has been corrected on the basis of [Melientsev 2004]). The grey solid line has been generated by the following equation:
\[ G = \frac{17749573.1}{(2006 - t)^2}. \]

Thus, up to the 1970s the hyperbolic growth of the world population was accompanied by the quadratic-hyperbolic growth of the world GDP, just as our model suggests. Note that the hyperbolic growth of the world population and the quadratic-hyperbolic growth of the world GDP are tightly interconnected processes, actually two sides of the same coin, two dimensions of one process propelled by the nonlinear second order positive feedback loops between the technological development and demographic growth (see Fig. 5).
We have also demonstrated (Korotayev, Malkov, and Khaltourina 2006a: 67–80) that the dynamics of the World System population's literacy ($l$) is rather accurately described by the following differential equation:

$$\frac{dl}{dt} = aSl (1-l),$$

where $l$ is the proportion of the population that is literate, $S$ is per capita surplus, and $a$ is a constant. In fact, this is a version of the autocatalytic model. It has the following sense: the increasing literacy is proportional to the fraction of the population that is literate, $l$ (potential teachers), to the fraction of the population that is illiterate, $(1-l)$ (potential pupils), and to the amount of per capita surplus $S$, since it can be used to support educational programs (in addition to this, $S$ reflects the technological level $T$ that implies, among other things, the level of development of educational technologies). Note that, from a mathematical point of view, Equation 9 can be regarded logistic where saturation is reached at literacy level $l = 1$, and $S$ is responsible for the speed with which this level is approached.

It is important to emphasize that with low values of $l$ (which correspond to most part of human history except for the recent decades), the increasing rate of the world literacy generated by this model (against the background of hyperbolic growth of $S$) can be approximated rather accurately as hyperbolic (see Fig. 6).
Fig. 6. World Literacy Dynamics, 1 – 1980 CE (%): the fit between predictions of the hyperbolic model and the observed data

Note: R = 0.997, \( R^2 = 0.994 \), \( p << 0.0001 \). Black dots correspond to UNESCO/World Bank (2014) estimates for the period after 1970, and to Meliantsev’s (2004) estimates for the earlier period. The grey solid line has been generated by the following equation:

\[
I_t = \frac{3769.264}{(2040 - t)^2}. 
\]

The best-fit values of parameters \( C (3769.264) \) and \( t_0 (2040) \) have been calculated with the least squares method.

The overall number of literate people is proportional both to the literacy level and to the overall population. As both of these variables experienced a hyperbolic growth until the 1960s/1970s, one has sufficient grounds to expect that until recently the overall number of literate people in the world \( (L) \) grew not just hyperbolically, but rather in a quadratic-hyperbolic way (as the world GDP did). Our empirical test has confirmed this – the quadratic-hyperbolic model describes the growth of the literate population of the planet with an extremely good fit indeed (see Fig. 7).

---

Since literacy appeared, almost all of the Earth's literate population has lived within the World System; hence, the literate population of the Earth and the literate population of the World System have been almost perfectly synonymous.
Fig. 7. World Literate Population Dynamics, 1–1980 CE ($L_t$, millions): the fit between predictions of the quadratic-hyperbolic model and the observed data

Note: $R = 0.9997$, $R^2 = 0.9994$, $p << 0.0001$. The black dots correspond to UNESCO/World Bank (2014) estimates for the period since 1970, and to Meliantsev’s (2004) estimates for the earlier period; we have also taken into account the changes of age structure on the basis of UN Population Division (2014) data. The grey solid line has been generated by the following equation:

$$L_t = \frac{4958551}{(2033 - t)^2}.$$ 

The best-fit values of parameters $C$ (4958551) and $t_0$ (2033) have been calculated with the least squares method.

Similar processes are observed with respect to world urbanization, whose macro dynamics appears to be described by the differential equation:

$$\frac{du}{dt} = bSu \left( u_{\text{lim}} - u \right),$$

where $u$ is the proportion of the population that is urban, $S$ is per capita surplus produced with the given level of the World System's technological development, $b$ is a constant, and $u_{\text{lim}}$ is the maximum possible proportion of the urban population. Note that this model implies that during the blow-up regime of the ‘Malthusian-Kuznetsian’ era, the hyperbolic growth of world urbanization must have been accompanied by a quadratic-hyperbolic growth of the urban population of the world, which is supported by our empirical tests (see Figs 8–9).
Fig. 8. World Megaurbanization Dynamics (% of the world population living in cities with > 250 thousand inhabitants), 10000 BCE – 1960 CE: the fit between predictions of the hyperbolic model and empirical estimates.

Note: $R = 0.987$, $R^2 = 0.974$, $p << 0.0001$. The black dots correspond to Chandler’s (1987) estimates, UN Population Division (2014), Modelski (2003), and Gruebler (2006). The grey solid line has been generated by the following equation:

$$u_t = \frac{403.012}{(1990 - t)}.$$

The best-fit values of parameters $C(403.012)$ and $t_0 (1990)$ have been calculated with the least squares method. For comparison, the best fit ($R^2$) obtained here for the exponential model is 0.492.

Fig. 9. Dynamics of World Urban Population Living in Cities with more than 250,000 Inhabitants (mlns), 10000 BCE – 1960 CE: the fit between predictions of the quadratic-hyperbolic model and the observed data.

Note: $R = 0.998$, $R^2 = 0.996$, $p << 0.0001$. The black markers correspond to estimates of Chandler (1987) and UN Population Division (2014). The grey solid line has been generated by the following equation:

$$U_t = \frac{912057.9}{(2008 - t)^2}.$$

The best-fit values of parameters $C(912057.9)$ and $t_0 (2008)$ have been calculated with the least squares method. For comparison, the best fit ($R^2$) obtained here for the exponential model is 0.637.
Within this context it is hardly surprising that the general macro dynamics of the size of the largest settlement within the World System is also quadratic-hyperbolic (see Fig. 10).

\[ U_{\text{max}} = \frac{104020618.573}{(2040 - t)^2}. \]

The best-fit values of parameters \( C \) (104020618.5) and \( t_0 \) (2040) have been calculated with the least squares method. For comparison, the best fit \( (R^2) \) obtained here for the exponential model is 0.747.

As has been demonstrated by cross-cultural anthropologists (see, e.g., Naroll and Divale 1976; Levinson and Malone 1980: 34), for pre-agrarian, agrarian, and early industrial cultures the size of the largest settlement is a rather effective indicator of the general sociocultural complexity of a social system. This, of course, suggests that in the ‘Malthusian-Kuznetsian’ era the World System’s general sociocultural complexity also increased, in a generally quadratic-hyperbolic way.

As we have noted in the beginning, the dynamics of marine biodiversity is strikingly similar to the population dynamics in China, the country with the best-known demographic history.

The similarity probably stems from the fact that both curves are produced by the interference of the same three components (general hyperbolic trend, as well as cyclical and stochastic dynamics). In fact, there is a lot of evidence that some aspects of biodiversity dynamics are stochastic (Raup et al. 1973; Sepkoski 1994; Markov 2001a; Markov 2001b; Cornette and Lieberman 2004), while others are periodic (Raup and Sepkoski 1984; Rohde and Müller 2005). On cyclical and stochastic components of the long-term population dynamics of China (as well as other complex agrarian societies) see, e.g., Korotayev and Khaltourina 2006; Korotayev, Malkov, and Khaltourina 2006b; Chu and Lee 1994; Nefedov 2004; Turchin 2003, 2005a, 2005b; Turchin and Korotayev 2006; Turchin and Nefedov 2009; Usher 1989; Komlos and Nefedov 2002; Grinin, Korotayev and Malkov 2008; Grinin et al. 2009; Grinin 2007c; Korotayev 2006; Korotayev, Khaltourina, and Bozhevolnov 2010; Korotayev et al. 2010; van Kessel-Hagesteijn 2009; Abel 1980; Braudel 1973; Goldstone 1991; Grinin, Korotayev 2012 etc.).
In fact, similarly to what we have observed with respect to the world population dynamics, even before the start of its intensive modernization, the population dynamics of China was characterized by a pronounced hyperbolic trend – as we can see below (see Figs 11 and 12), the hyperbolic model describes traditional Chinese population dynamics much more accurately than either linear or exponential models do:

**Fig. 11.** Population Dynamics of China (million people), 57–1851 CE: fit with linear and exponential models


**Fig. 12.** Population Dynamics of China (million people), 57–1851 CE: fit with a hyperbolic model

The hyperbolic model turns out to describe mathematically the population dynamics of China in an especially accurate way with respect to the modern period (see Fig. 13).
In a rather similar way the hyperbolic model turns out to describe the marine biodiversity (measured by number of genera) through the Phanerozoic much more accurately than the exponential one (see Fig. 14):

When measured in terms of species number the fit between the empirically observed marine biodiversity dynamics and the hyperbolic model becomes even better (see Fig. 15):
The hyperbolic model describes the continental biodiversity in an especially accurate way (see Fig. 16).

However, the highest fit between the hyperbolic model and the empirical data is observed when the hyperbolic model is used to describe the dynamics of total (marine and continental) global biodiversity (see Fig. 17).
As we see, the hyperbolic dynamics is most prominent when both marine and continental biotas are considered together. This fact can be interpreted as a proof of the integrated nature of the biosphere.

But why throughout the Phanerozoic did the global biodiversity tend to follow the hyperbolic trend (similarly to what we observed within social World System in general and China in particular)?

As we have noted above, in macrosociological models, the hyperbolic pattern of the world population growth arises from a non-linear second-order positive feedback (more or less identical with the mechanism of collective learning) between the demographic growth and technological development (more people – more potential inventors – faster technological growth – the carrying capacity of the Earth grows faster – faster population growth – more people – more potential inventors, and so on).

Based on the analogy with macrosociological models and diverse paleontological data, we suggest that the hyperbolic character of biodiversity growth can be similarly accounted for by a non-linear second-order positive feedback between the diversity growth and community structure complexity (more genera – higher alpha diversity – the communities become more stable and ‘buffered’ – average life span of genera grows; extinction rate decreases – faster diversity growth – more genera – higher alpha diversity, and so on).

The growth of genus richness through the Phanerozoic was mainly due to the increase of average longevity of genera and gradual accumulation of long-lived (stable) genera in the biota. This pattern reveals itself in the decrease of extinction rate. Interestingly, in both biota and humanity, growth was facilitated by the decrease in mortality rather than by the increase in birth rate. The longevity of newly arising genera was growing in a stepwise manner. The most short-lived genera appeared during the Cambrian; more long-lived genera appeared in Ordovician to Permian; the next two stages correspond to the Mesozoic and Cenozoic (Markov 2001a, 2002). We suggest that diversity growth can facilitate the increase in genus longevity via the progressive stepwise changes in the structure of communities.

One wonders if it cannot be regarded as a (rather imperfect) analogue of the collective learning mechanism that plays such an important role within the social macroevolution.
Most authors agree that there were three major biotic changes that resulted in fundamental reorganization of community structure during the Phanerozoic: Ordovician radiation, end-Permian extinction, and end-Cretaceous extinction (Bambach 1977; Sepkoski et al. 1981; Sepkoski 1988, 1992; Markov 2001a; Bambach et al. 2002). Generally, after each major crisis the communities became more complex, diverse and stable. The stepwise increase of alpha diversity (average number of species or genera in a community) through the Phanerozoic was demonstrated by Bambach (1977) and Sepkoski (1988). Although Powell and Kowalewski (2002) argued that the observed increase in alpha diversity might be an artifact caused by several specific biases that influenced the taxonomic richness of different parts of the fossil record, there is evidence that these biases largely compensated each other, so that the observed increase in alpha diversity was probably underestimated rather than overestimated (Bush and Bambach 2004).

Another important symptom of progressive development of communities is the increase in evenness of distribution of species (or genus) abundances. In the primitive, pioneer or suppressed communities, this distribution is strongly uneven (community is overwhelmingly dominated by a few very abundant species). In more advanced, climax or flourishing communities, this distribution is more even (Magurran 1988). The former type of community is generally more vulnerable. Evenness of distribution of species richness in communities increased substantially during the Phanerozoic (Powell and Kowalewski 2002; Bush and Bambach 2004). Most probably there was also an increase in habitat utilization, total biomass and rate of trophic flow in biota through the Phanerozoic (Powell and Kowalewski 2002).

The more complex the community, the more stable it is due to the development of effective interspecies interactions and homeostatic mechanisms based on the negative feedback principle. In a complex community, when the abundance of a species decreases, many factors arise that facilitate its recovery (e.g., there will be more food and fewer predators). Even if a species becomes extinct, its vacant niche may ‘recruit’ another species, most probably a related one that may acquire morphological similarity with its predecessor and thus, the taxonomists will assign it to the same genus. So a complex community can facilitate the stability (and longevity) of its components, such as niches, taxa and morphotypes. This effect reveals itself in the phenomenon of ‘coordinated stasis’: the fossil record shows many examples of persistence of particular communities for many million years while the rates of extinction and taxonomic turnover are minimized (Brett et al. 1996, 2007).

Selective extinction leads to accumulation of ‘extinction-tolerant’ taxa in the biota (Sepkoski 1991b). Although there is evidence that mass extinctions can be non-selective in some aspects (Jablonski 2005), they are obviously highly selective with respect to the ability of taxa to endure unpredictable environmental changes. This can be seen, for instance, from the selectivity of the end-Cretaceous mass extinction with respect to the time of the first occurrence of genera. In younger cohorts the extinction level was higher compared to the older cohorts (see Markov and Korotayev 2007a: Fig. 2). The same pattern can be observed during the periods of ‘background’ extinction as well (Markov 2000). This means that genera differ in their ability to survive the extinction events, and that in the course of time the extinction-tolerant genera accumulate in each cohort. Thus, taxa generally become more stable and long-lived in the course of evolution, apart from the effects of
communities. The communities composed of more stable taxa would be, in turn, more sta-
ble themselves, thus creating a positive feedback.

The stepwise change of dominant taxa plays a major role in biotic evolution. This pat-
tern is maintained not only by the selectivity of extinction (discussed above), but also by
the selectivity of the recovery after crises (Bambach et al. 2002). The taxonomic structure
of the Phanerozoic biota was changing in a stepwise way, as demonstrated by the concept of
three sequential ‘evolutionary faunas’ (Sepkoski 1992). There were also stepwise changes in
the proportion of major groups of animals with different ecological and physiological param-
eters. There was a stepwise growth in proportion of motile genera compared to non-motile;
‘physiologically buffered’ genera compared to ‘unbuffered’, and predators compared to prey
(Bambach et al. 2002). All these trends should have facilitated the stability of communities
(e.g., diversification of predators implies that they become more specialized; a specialized
predator regulates its prey’s abundance more effectively than a non-specialized predator).

There is also another possible mechanism of the second-order positive feedback be-
tween the diversity and its growth rate. Recent research has demonstrated a shift in typical
relative-abundance distributions in paleocommunities after the Paleozoic (Wagner et al.
2006). One possible interpretation of this shift is that the community structure and the in-
teractions between species in the communities became more complex. In the post-
Paleozoic communities, new species probably increase ecological space more efficiently,
either by facilitating opportunities for additional species or by niche construction (Wagner
et al. 2006; Solé et al. 2002; Laland et al. 1999). This possibility makes the mechanisms
underlying the hyperbolic growth of biodiversity and human population even more similar,
because the total ecological space of the biota is analogous to the ‘carrying capacity of the
Earth’ in demography. As far as new species can increase ecological space and facilitate
opportunities for additional species entering the community, they are analogous to the ‘in-
ventors’ of the demographic models whose inventions increase the carrying capacity of the
Earth.

Exponential and logistic models of biodiversity imply several possible ways in which
the rates of origination and extinction may change through time (Sepkoski 1991a). For
instance, exponential growth can be derived from constant per-taxon extinction and origi-
nation rates the latter being higher than the former. However, actual paleontological data
suggest that origination and extinction rates did not follow any distinct trend through the
Phanerozoic, and their changes over time look very much like chaotic fluctuations (Cor-
nette and Lieberman 2004). Therefore, it is more difficult to find a simple mathematical
approximation for origination and extinction rates than for the total diversity. In fact, the
only critical requirement of the exponential model is that the difference between the origi-
nation and extinction through time should be proportional to the current diversity level:

\[(N_o - N_e)/\Delta t \approx kN, \quad (11)\]

where \(N_o\) and \(N_e\) are the numbers of genera with, respectively, first and last occurrences
within the time interval \(\Delta t\), and \(N\) is mean diversity level in the interval. The same is true
for the hyperbolic model. It does not predict the exact way in which origination and extin-
tion should change, but it does predict that their difference should be roughly propor-
tional to the square of the current diversity level:

\[(N_o - N_e)/\Delta t \approx kN^2. \quad (12)\]

In demographic models discussed above, the hyperbolic growth of the world population
was not decomposed into separate trends of birth and death rates. The main driving force
of this growth is presumably the increase of the Earth’s carrying capacity and the way this
capacity is realized – either by decreasing death rate, or by increasing birth rate, or both –
depends upon many factors and may vary from time to time.

The same is probably true for biodiversity. The overall shape of the diversity curve
depends mostly on the differences in the mean rates of diversity growth in the Paleozoic
(low), Mesozoic (moderate), and Cenozoic (high). The Mesozoic increase was mainly due
to lower extinction rate (compared to the Paleozoic), while the Cenozoic increase was
largely due to higher origination rate (compared to the Mesozoic) (see Markov and Korot-
tayev 2007a: 316, Figs 3a, 3b). This probably means that the acceleration of diversity
growth during the last two eras was driven by different mechanisms of positive feedback
between diversity and its growth rate. Generally, the increment rate \((\frac{N_o - N_e}{\Delta t})\) was
changing in a more regular way than the origination rate \(N_o/\Delta t\) and extinction rate \(N_e/\Delta t\).
The large-scale changes in the increment rate correlate better with \(N^2\) than with \(N\) (Ibid.:
figs 3c and 3d), thus supporting the hyperbolic rather than the exponential model.

Conclusion
In macrosociological models the hyperbolic pattern of the world population growth arises
from a non-linear second-order positive feedback between the demographic growth and
technological development (more people – more potential inventors – faster technological
growth – the carrying capacity of the Earth grows faster – faster population growth – more
people – more potential inventors, and so on, which is more or less identical with the
working of the collective learning mechanism). Based on the analogy with macrosociologi-
cal models and diverse paleontological data, we suggest that the hyperbolic character of
biodiversity growth can be similarly accounted for by a non-linear second-order positive
feedback between the diversity growth and community structure complexity (which sug-
gests the presence within the biosphere of a certain analogue of the collective learning
mechanism). The feedback can work via two parallel mechanisms: 1) decreasing extinc-
tion rate (more taxa – higher is the alpha diversity, or mean number of taxa in a communi-
ity – communities become more complex and stable – extinction rate decreases – more
taxa, and so on), and 2) increasing origination rate (new taxa facilitate niche construction;
newly formed niches can be occupied by the next ‘generation’ of taxa). The latter makes
the mechanisms underlying the hyperbolic growth of biodiversity and human population
even more similar, because the total ecospace of the biota is analogous to the ‘carrying
capacity of the Earth’ in demography. As far as new species can increase ecospace and
facilitate opportunities for additional species entering the community, they are analogous
to the ‘inventors’ in the demographic models whose inventions increase the carrying ca-
pacity of the Earth. The hyperbolic growth of the Phanerozoic biodiversity suggests that
‘cooperative’ interactions between taxa can play an important role in evolution, along with
generally accepted competitive interactions. Due to this ‘cooperation’ (~ ‘collective learn-
ing’?), the evolution of biodiversity acquires some features of a self-accelerating process.
The same naturally refers to cooperation/collective learning as regards the global social
evolution. The discussed above suggests that we can trace rather similar macropatterns
within both the biological and social phases of Big History that produce rather similar
curves in diagrams and that can be described in rather accurate way with rather simple
mathematical models.

References
Abel, W. 1980. Agricultural Fluctuations in Europe from the Thirteenth to the Twentieth Centuries. New
York: St. Martin’s.


