The Universal Decay of Collective Memory and Attention

Cristian Candia[a,b,c,*], C. Jara-Figueroa[a], Carlos Rodriguez-Sickert[c], Albert-László Barabási[b], and César A. Hidalgo[a,*]

[a]Collective Learning Group, The MIT Media Lab, Massachusetts Institute of Technology

[b]Network Science Institute, Northeastern University, Boston, MA 02115, USA

[c]Centro de Investigación en Complejidad Social (CICS), Facultad de Gobierno, Universidad del Desarrollo, Santiago, Chile.

[*]Corresponding Authors: hidalgo@mit.edu, ccandiav@mit.edu

September 26, 2018

Abstract

Collective memory and attention are sustained by two channels: oral communication (communicative memory) and the physical recording of information (cultural memory). Here, we use data on the citation of papers and patents, and on the online attention received by songs, movies, and biographies, to study the temporal decay of the attention received by comparable cultural products. We find that the attention received by online and offline cultural products decays following a universal bi-exponential function. We propose a mathematical model that associates communicative and cultural memory to each exponential decay. This model fits the data better than previously proposed models. Finally, we use this model to measure the transition between communicative and cultural memory, finding that biographies remain in our communicative memory the longest (20-30 years) and music the shortest (about 5.6 years). These findings show that a universal decay function governs the average collective attention received by cultural products.

Introduction

In what is probably Pablo Neruda's most famous poem–Poema 20–he wrote: "Es tan corto el amor, y tan largo el olvido" (Love is so short, forgetting is so long). Neruda's words express elegantly the fact that when people are in love, they are constantly thinking of their loved ones, but once love fades, memories fade too. Inspired by Neruda, we ask whether society also experiences the two phases of memory: an initial phase of high attention, followed by a longer and slower phase of forgetting. In fact, there is a vast literature suggesting this might be the case, since collective memory is acknowledge to be a combination of two distinct processes[1–11]: communicative memory, normally sustained by the oral transmission of information, and cultural memory, which is sustained by the physical recording of information. This literature can provide inspiration for the construction of generative models for the attention received by cultural products.

Despite this progress, the theory of collective memory and attention is short on quantitative models that would allow us to connect it empirically to large-scale data, such as the data developed in the literature of knowledge diffusion. Indeed, the literature on knowledge diffusion models the adoption and diffusion of cultural content as a combination of two processes[12–18]: preferential attachment and temporal decay. Preferential attachment,[19,20] or cumulative advantage[21–23], refers to a process in which attention begets attention. Think of two scientific papers, one with 10,000 citations and another one with 100. The probability that the first paper receives a new citation is larger than the second one, simply because more people already know about it. This preferential attachment process needs to be properly discounted to measure temporal decay.

Recently, models combining preferential attachment and temporal decay have described the decay of attention (mostly paper and patent citations) using exponential and log-normal functions[12,13]. These models agree on the idea that attention should be modeled using a combination of preferential attachment and time decay. Yet, there is no consensus about the shape of the decay function or its universality across a variety of cultural domains.

Here, we use data on scientific publications, patents, songs, movies, and biographies to test the hypothesis that the decay of the attention received by a cultural product involves both the decay of communicative and cultural memory. Due to the properties ascribed to each type of memory–communicative memory being short-lived as compared to cultural memory[24]–we expect that the attention received by collective memory should decay fast at first, whereas that of cultural memory should follow a softer and longer decline. We formalize these ideas by constructing a mathematical model that predicts a bi-exponential decay function and validate it by showing that it is statistically better at explaining the empirically observed decay of attention than the exponential[13] and log-normal[12] functions used in the previous literature. This finding validates our hypothesis that the decay of the attention received by human collective memory is a process that results from the interplay between both communicative and cultural memory. The model also allows us to separate both mechanisms and generalizes well to multiple data-sets, suggesting that it captures a universal feature of the decay of human collective memory.

Collective Memory and Attention

Collective memories are sustained by communities, which could be as large as all the speakers of a language, or as small as a family. During the last century, scholars studying collective memory have advanced a large number of definitions, models, and processes, helping characterize different forms of collective memory and the mechanisms that contribute to their preservation[25].

Psychologists have explored both top-down and bottom-up approaches to memory formation and retention. Top-down approaches focus on how familiarity[26,27], narrative templates[28,29], and cultural attractors[30–32] contribute to the retention and formation of collective memories. Familiarity increases the memorability of events, even causing false memories, like that of people identifying Alexander Hamilton as a U.S. president[26]. Narrative templates, which are schemata that people use to describe multiple historical events, can also shape memories, like the memory of Russian exceptionalism that emerges from the narrative template of invasion, near defeat, and heroic triumph[28]. Cultural attractors, such as repetitive children songs or count-out rhymes, can increase the preservation of memories across generations[27].

Bottom-up approaches focus on how micro-level psychological processes can shape social outcomes[25]. For instance, forgetting can be induced through the selective retrieval of events, an effect known as retrieval-induced forgetting[33–35].

Also, social affinities, like belonging to the same social group, can increase the mnemonic power of conversations[36–39]. For instance, people are motivated to create coalitions[40] and shared realities with those that they perceive as belonging to their own group[37].

Scholars in computational social sciences have followed a different approach, focusing on how collective memory is expressed and created in the consumption of cultural content, from Wikipedia page views[11,41–44] to paper and patent citations[12,13,45,46]. Of course, these online and offline metrics are not direct measures of collective memory or attention, they are measures of the spillovers of attention that result in online searches or references. The idea is that movies, songs, or papers that are being talked about are of heightened interest, and hence, lead people to consult various data sources. When these cultural products move away from communicative memory, they lose the intense attention they had when they were being talked about.

Unfortunately, these aggregate approaches cannot distinguish between different forms of memory or attention loss, such as interference, suppression, or inhibition. They only provide an aggregate picture of the attention lost through all of these channels.

Nevertheless, the computational social science approach is closer to the definition of collective memory given by Jan Assmann[2,4], which focuses on the cultural products that communities or groups of people remember. Assmann–even though he focused on long-lived inter-generational memories–distinguishes between modes of *potentiality* and *actuality*. Potentiality being the existence of a record (an old book in a library's basement), and actuality being the attention received by that record when it becomes relevant to the community. The computational social science literature has focused on the use of big data to study the actuality of memories and the effects of language, technology, accomplishments, and triggers in the dynamics of collective memory and attention. For instance, historical figures born in countries with languages that are often translated to other languages receive more online attention than comparable historical figures born in less frequently translated languages[47]. Changes in communication technologies, such as the rise of the printing press, radio, and television, have also been shown to affect attention since they correlate with changes in the occupations of the people entering biographical records[42]. The edits and attention received by events in Wikipedia has also been seen to increase with related exogenous events[11,44], such as natural and human-made disasters, accidents, terrorism, and during anniversaries or commemorative events[48]. Moreover, the online attention received by past sports figures–a measure of their prevalence in present-day memory–has been shown to correlate with an age discounted measure of performance,[41,49], meaning that memorability and attention–at least in athletic activities–correlate with merit.

The approach presented in this paper is related more closely to the computational social science strand of literature since it uses cultural consumption data to study the dynamics of the attention received by cultural products and biographies. Yet, it is also an approach that is not completely unrelated to the psychological strand. By studying the dynamics of consumption of cultural products, from songs to scientific papers, we are exploring a form of selective retrieval, albeit not focused on how this selective retrieval shapes collective identity, but on its average temporal dynamics. Moreover, by proposing a model that describes the dynamics of attention, we are undertaking a bottom-up approach to the modeling of collective memory and attention. Finally, by looking at multiple cultural domains, we can explore the universality of average decay functions, rather than focusing on the mechanisms that make some events more or less memorable.

Results

The literature on collective memory[1,24,50] suggests that the decay of the attention received by a cultural product involves two mechanisms, an initial fast decay–a signature of communicative memory–followed by a softer decline–resulting from cultural memory–(Figure 1 A). Using the distinction between communicative and cultural memory[3,4,24,50] we propose a model where cultural memory and communicative memory co-exist, but decay at different rates. The decay of both types of memory, especially cultural memory, should be understood in relative terms: the share of the current attention occupied by a cultural product may stay the same, but because total memory is growing–more products are created each time. Hence the *relative* share of the current attention ($S(t)$) assigned to the said product will decay.

We model the attention received by a cultural product using several simplifying assumptions. First, we assume that the current attention, $S(t)$, of a cultural product is the sum of the attention it garners from both communicative memory u and cultural memory v. Hence, at any given time $S(t) = u(t) + v(t)$ (Fig. 1 A). Second, we assume that communicative and cultural memory decay, in relative terms, independently with decay rates $p+r$ for communicative memory and q for cultural memory, and that information flows from communicative memory to cultural memory at a rate r. Many processes are captured by the parameters p, r, and q, perhaps the most straightforward one is that because the total size of cultural memory is growing, the relative share occupied by a certain cultural product will shrink, which is captured in p. We acknowledge that these assumptions cannot capture the full complexity of the processes by which communicative and cultural memory decay, nor their interactions. The communicative and cultural memory may feed on each other on more complex ways than the assumed linear form (r). We adopt these simplifying assumptions with the goal of providing a tractable model with as few parameters as possible that can be used to capture the leading forces governing the dynamics of attention received by a cultural product. Given these assumptions, communicative memory decays as $u(t+1) = (1-p)u(t) - ru(t)$ and cultural memory as $v(t+1) = (1-q)v(t) + ru(t)$, together defining the following system of differential equations:

$$S(t) = u(t) + v(t) \tag{1}$$

$$\frac{du}{dt} = -(p+r)u \tag{2}$$

$$\frac{dv}{dt} = -qv + ru. \tag{3}$$

We set the initial communicative memory $u(t=0) = N$ and we assume that at the beginning of the process there is no cultural memory associated to a new cultural product ($v(t=0) = 0$)–although there are alternatives ways to initialize the model that does not change its aggregate behavior (Supplementary Model).

Using the initial conditions, we find that the solution of the equation system (1)-(3) is the bi-exponential function:

Figure 1: Scheme of the collective memory model. **A** The y-axis, $S(t)$, represents the normalized current level of attention received by a group of comparable cultural pieces. The x-axis, Age, represents the age of the cultural pieces, measured in years. The red curve shows the bi-exponential function predicted by our model in log-lin scale. The light-blue and light-green curves show the two exponential of communicative and cultural memory. The inset illustrates the basic mechanics of the model. At any time point t the total memory is the sum of communicative memory u and cultural memory v. Both communicate and cultural memory decay with their own respective decay rates $p+r$ and q, and cultural memory grows with r. **B** The bi-exponential model (Eq. 6) for various parameters p, q, and r, can account for a wide range of decays. **C** Comparison between the bi-exponential model (in red), and the exponential and log-normal models in log-log scale.

$$u(t) = Ne^{-(p+r)t}, \tag{4}$$

$$v(t) = \frac{Nr}{p+r-q}\left(e^{-qt} - e^{-(p+r)t}\right), \tag{5}$$

$$S(t) = \frac{N}{p+r-q}\left[(p-q)e^{-(p+r)t} + re^{-qt}\right]. \tag{6}$$

Figure 1 B illustrates $S(t)$ for different values of the parameters, with $N = 1$, and Figure 1 C compares the bi-exponential function with the exponential[13,14] and log-normal[12] decay functions explored previously in the literature (Supplementary Notes).

Figure 2: The universal decay of collective memory. Average number of new citations received by: **A** all papers published in Physical Review B in 1980 ($n = 1,415$), **B** all papers published in Physical Review D in 1980 ($n = 803$), **C** all papers published in Physical Review Letters in 1990 ($n = 1,904$), **D** all papers published in Physical Review L in 1980 ($n = 1,202$), **E** all Mechanical patents granted in 1990 ($n = 20,296$), and **F** all Chemical patents granted in 1985 ($n = 14,749$). For cultural products we use the standardized levels of online attention for: **G** songs ($n = 18,320$) based on spotify's popularity index (y-axis) as a function of the date the song first appeared in the Billboard ranking (x-axis), **H** songs ($n = 15,275$) based on Last.fm's play counts (y-axis) as a function of the date the song first appeared in the Billboard ranking (x-axis), **I** movies ($n = 14,633$) based on YouTube's view counts (y-axis) as a function of the date the movie was released (x-axis), **J** tennis players ($n = 624$) based on Wikipedia's page views (y-axis) as a function of the date that the tennis player was included in the Top 600 International males singles tennis player (x-axis,. **K** olympic medalist ($n = 526$) based on Wikipedia's page views (y-axis) as a function of the date of the middle of the career of the Olympic medalist, and **L** basketball players ($n = 592$) based on Wikipedia's page views (y-axis) as a function of the date that the Basketball player starts his career (x-axis). Red lines show our bi-exponential model fit whereas the dashed lines and dotted lines show the log-normal decay used by Wang et al.[12] and the exponential decay used by Higham et al.[13].

We bring the bi-exponential model to our data by comparing it with the decay functions observed for paper and patent citations, and for the current online attention of past songs, movies, and biographies, with a comparable level of accomplishment. In the case of papers and patents, we group papers and patents with a similar number of cumulative citations. In the case of songs, movies, and biographies, these comparable sets are built into our selection criterion, since we study only songs that reached the Billboard ranking, biographies of award-winning athletes, and movies that have received over 1,000 votes on IMDB. By respectively grouping papers, patents, songs, movies, and biographies, with a similar level of accomplishment, we control for differences in preferential attachment, allowing us to isolate the temporal decay of collective memory statistically.

Figure 2 shows the average number of new citations obtained by scientific papers (A, B, C, and D) and patents (E and F) for different levels of accumulated citations k. The red lines show the fit of the bi-exponential model, whereas the dashed and dotted lines capture, respectively, the log-normal and exponential decays used in[12] and[13]. In all cases, we find that, after choosing papers and patents with the same level of cumulative citations, the bi-exponential model captures the temporal pattern of human collective forgetting accurately (see Methods section and Supplementary Tables for goodness of fit comparison for data on all years, journals, and categories). More importantly, in several of these empirical curves, the shoulder of the bi-exponential curve is clearly visible, allowing the model helps to unveil the point at which cultural memory takes over communicative memory.

We observe a similar behavior when we apply the bi-exponential model to data on music, movies, and biographies. Since we lack time series data for these three sources, we look at the present day online attention to music, movies, and biographies as a function of their age. For songs, we determine age using the year they first reached the Billboard ranking. For movies, we calculate age using their release year. For the biographies of athletes, we use as the age of the accomplishment the time when they were introduced in their respective international rankings. Once again, when we compare our model with the previously proposed log-normal (dashed lines) and exponential models (dotted lines) (Figure 2 G-L), we find that the bi-exponential model provides a more accurate fit to the data, due to its ability to capture the initial fast decay of communicative memory together with the slow decay of cultural memory. Also, it visible captures the transition from communicative to cultural memory.

Together, the data on papers, patents, songs, movies, and biographies, shows that this bi-exponential decay is universal across all domains. Yet, the parameters of the decay function are different for papers, patents, songs, movies, and biographies. We therefore, compared the model parameters (p, q, r, and t_c) across all studied domains (Figure 3). Here t_c is the time at which cultural memory overtakes communicate memory, which, according to the model can be approximated as (see Supplementary Model)

$$t_c = \frac{1}{p+r-q} \log\left(\frac{(p+r)(p-q)}{rq}\right). \tag{7}$$

Although our results suggest that the functional form of the decay in attention function is universal across multiple cultural domains, its parameters are informative of the domain-specific decay dynamics (Figure 3). When comparing the obtained parameters, we find that the decay rates of communicative memory are much larger than those of cultural memory ($p >> q$) as suggested by the literature[2] (Figure 3 A and C). Also, we find that communicative memory decays much faster for music and movies than for biographies (Figure 3 C), resulting in critical times that are relatively low for music, movies, and papers (5 to 10 years, Figure 3 D), and much longer for biographies (15 to 30 years). In other words, for biographies, the era dominated by communicative memory lasts longer than the era dominated by cultural memory.

Figure 3: Model's parameters described by Eq. 6, and for the same data deployed in figure 2. Each box correspond to a model's parameter and colors represent the type of cultural product: red for songs, blue for movies, green for biographies, purple for patents, and orange for papers. The y-axis for parameters q, r, and p represent the change rate, measured in number of citations over time (years). The y-axis for T_c represents the critical time and it is calculated by Eq. 7, and it is measured in years. The x-axis represents the cultural domain analyzed. Bars represent the standard deviation of the coefficient estimation. (See Supplementary Tables)

Together, these results show that the bi-exponential decay predicted from formalizing the mechanisms suggested by the literature on collective memory provide a universally good approximation for the decay of memory across a wide variety of cultural domains.

Discussion

Inspired by Neruda's observation, that love was short and intense while forgetting lingered, we build on the ideas of communicative and cultural memory to show that the decay of the attention received by cultural products and biographies follows a universal decay function that is characterized by two phases: a short-lived and fast decaying phase connected to communicative memory, and a longer-lived and slower decaying phase connected to cultural memory. We find that the shape of this function is universal across multiple cultural domains and that its parameters are informative of the attention dynamics characterizing each domain. These findings provide quantitative evidence to validate the concepts of communicative and cultural memory and allow us to better understand how societies forget.

For decades scholars have been using papers and patents citations to study the spread and adoption of ideas and cultural content[12–14,45,46,51–58]. Mathematically, the temporal decay curves describing the number of citations or attention $A(t)$ received by a paper, patent, or cultural product, can be expressed as a function of two parameters: (i) its age t, and (ii) the cumulative citations received by that paper, patent, or cultural product k. Formally, it has been shown that $A(t)$ is separable[12,13,15–18], as $A(t) = c(k) \times S(t)$, where $c(k)$ captures the effects of preferential attachment and $S(t)$ captures the temporal decay (see Methods section and Supplementary Notes). Yet, while there is consensus on the fact that preferential attachment processes contribute to the spread of cultural products with high levels of attention, there is no consensus on the nature of the functional form capturing the decay of attention. The data shows an initially fast decay followed by a milder decline. What gives rise to this unorthodox decay function?

Our results indicate that the fast decay followed by a mild decline observed in these decay functions is a universal bi-exponential curve that can be derived from a model that builds on two fundamental concepts from the literature on collective memory: communicative and cultural memory[1–10]. The agreement between this model and the empirical data validates these theoretical mechanisms and offers a mean to quantify them.

While the shape of the decay function is universal, its parameters are informative of the decay dynamics of specific systems (Figure 3). For instance, athlete biographies have relatively large critical times compared to songs, movies, papers, and patents; meaning that athletes are remembered mainly through oral culture for as much as a couple of decades after their main accomplishment. Songs, on the other hand, have a relatively high rate of transfer (r) from communicative to cultural memory and are short-lived in communicative memory. In fact, both songs and movies are short-lived in communicative memory, but movies live in communicative memory a bit longer than songs, probably because of the high output of the music industry.

But why would communicative memory feed cultural memory? The rationale behind communicative memory feeding cultural memory is that after each communicative act the probability that a record is created increases. The parameter r intends to capture, on average, how communicative memory translates into cultural memory. We acknowledge that there are more complex mechanisms associated with this process, for instance, cultural memory should also feed communicative memory. Yet, despite these simplifying assumptions, the model employed here still explains much of the variation observed in the data.

According to our model, in the beginning, most of the attention comes from acts of communication, but this changes over time. Indeed, after a critical time (t_c), cultural products receive more attention from records than from acts of communication. For instance, soon after their release, scientific papers are discussed at conferences, media, magazines, and the news. This generates an excess of attention for newer cultural products and the creation of new records referring to that product. Yet, once the conversation is over, the attention coming from the consultation of these records becomes dominant.

Nevertheless, it is interesting to think about the mechanisms that could contribute to the reduction of communicative memory or the *flattening* of the bi-exponential function. For example, the level of coordinated consumption of cultural goods (e.g., how much people like to go to movies together), could modulate how much that cultural good is discussed, and hence, the size of the communicative bump. Also, exogenous effects, such as the cancellation of a conference due to weather could reduce the communicative memory effects for the papers discussed in those conferences.

Our results support the hypotheses that the decay of human collective memory involves the combined decay of communicative and cultural memory, and that the decay function is universal across multiple cultural domains. These findings allow us to explain the dynamics of the attention received by a cultural product during its lifetime, and suggest that the dynamics of human collective memory follows a universal decay function.

Methods

Data

We use two types of data sources: time series data for scientific papers and patents, and cross-section data for songs, movies, and biographies summarized in Table 1. The American Physical Society (APS) corpus collects data about the attention pattern of physics articles from twelve different journals, between 1896 and 2016. For our analysis, we use a prospective approach (See Supplementary Methods) for all papers published between 1970 and 2003 in Physical Review Letters (PRL), and in Physical Review A to E[14,54,59] ($n = 485,105$). The United States Patent and Trademark Office (USPTO)[60,61] contains information about patents granted between 1976 and 2005. We use all patents granted between 1976 and 1995 in all categories ($n = 1,681,690$): Chemical (CAT 1), Computers & Computation (CAT 2), Drugs & Medical (CAT 3), Electrical & Electronic (CAT 4), Mechanical (CAT 5) and Others (CAT 6). For both patents and papers, we construct two time-series, one for the number of citations obtained in each time window, and another for the accumulated citations obtained up to a given time. Because we are interested in characterizing the dynamics of relative attention, we adjust the time series by normalizing it by the number of papers published in a journal each year[13,14] (See Supplementary Methods).

For songs, movies, and online biographies we use cross-section data, it means, data collected by observing songs movies and biographies at the same point in time. We use different inclusion criteria–what cultural products are included in our sample–for each type of cultural content. For songs, we use weekly ranking data from the "Hot-100 Billboard's ranking"[62] between October 1958 and July 2017. To measure online attention, we use Spotify's popularity index[63] (a direct function of play counts) taken on October 2016 and July 2017 ($n = 18,320$), and last.fm's ($n = 15,275$) play counts[64] for the last week of July 2017 (see Supplementary Methods). We also collect data on $14,633$ movies released between 1937 and 2017 that have obtained more than 1,000 votes in the Internet

Table 1: Cultural products and their measurements of present day levels of attention (current attention) and measurements to account by cumulated advantage effect (Accomplishment).

Cultural Products	Attention Metric	Preferential Attachment Metric
APS Papers	Citations received in the last 6 months	Cumulative citations
USPTO Patents	Citations received in the last 6 months	Cumulative citations
Music	Spotify popularity Last.fm play counts	Entered at least once in the Hot-100 Billboard ranking
Movies	Trailer play counts in YouTube	More than 1,000 votes on IMDB
Biographies	Wikipedia page views	Highly performing athletes in tennis, basketball, and the Olympics

Figure 4: Universal patterns in the decay of human collective memory. **A** Average number of citations received each semester by papers published in physical review B ($A(t)$). **B** Average number of citations received by a paper as a function of the cumulative citations received by that paper ($\propto c(k)$). Different curves represent different ages, **C** Average number of citations received by papers with the same number of cumulative citations as a function of their age ($\propto S(t)$). Different curves represent groups of papers with a different number of total citations.

Movie Database[65] as of July 2017. To measure the current online attention of movies we use the play counts for the trailer of each movie taken from YouTube[66] ($n = 14,633$). For online biographies, we focus on basketball, tennis, and Olympic medal winners. For basketball players, we consider the "Slam 500 Greatest NBA Players of All Times ($n = 592$)," for tennis players we consider the "Top 600 International males singles tennis player ($n = 624$)," and for Olympic medal winners we consider athletes who have won more than three gold medals ($n = 526$). Current online attention was measured using the number of page views received by the Wikipedia biography[67] of each athlete between July 2016 and June 2017–for more information see Supplementary Methods.

Decomposition of Citing Curve

Mathematically, in our approach the temporal decay curves describing the number of citations or attention $A(t)$ received by a paper, patent, or piece of cultural content (Figure 4 A) can be expressed as a function of two parameters: (i) its age t, and (ii) the cumulative citations received by that paper, patent, or cultural piece k. Formally, it has been shown that $A(t)$ is separable[12,13,15–18], as $A(t) = c(k) \times S(t)$, where $c(k)$ captures the effects of preferential attachment (Figure 4 B) and $S(t)$ captures the temporal decay (Figure 4 C).

The solid line (Figure 4 A) shows the average number of citations received by papers published in Physical Review B in 1990 ($A(t)$) as a function of their age. $A(t)$ describes the traditional increase and decline known to characterize knowledge diffusion or cultural product adoption curves[13,15–18].

Figure 4 B shows the preferential attachment component, by presenting the number of new citations (Δc) received by a paper as a function of its cumulative citations ($c(k)$)[19,20]. Figure 4 C shows the temporal decay component ($S(t)$), representing the number of new citations received by papers with the same number of cumulative citations $k = k^*$ as a function of their age. That is, the dashed lines show papers for which the effect of preferential attachment is kept constant: $A(t)|_{k=k^*} = c(k^*) \times S(t)$. Here, we observe the initially fast decay followed by a milder decline.

Model

Here, we formalize this intuition by proposing a model for the decay of the attention received by a cultural piece. We took inspiration from collective memory studies and nuclear decay. We solve the model analytically as follow:

$$\frac{du}{dt} = -pu - ru = -(p+r)u \tag{8}$$
$$\frac{dv}{dt} = ru - qv = ru - qv \tag{9}$$
$$\tag{10}$$

We can write this using matrix representation:

$$\begin{pmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{pmatrix} = \begin{pmatrix} -(p+r) & 0 \\ r & -q \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Where the initial conditions are:

$$\begin{pmatrix} u(0) \\ v(0) \end{pmatrix} = \begin{pmatrix} N \\ 0 \end{pmatrix}$$

Then, to solve the equation system we first have to find the eigen values of the 2x2 matrix, this is:

$$det(A - \lambda I) = 0 \tag{11}$$

Solving for A, we find $\lambda_1 = -(p+o)$ and $\lambda_2 = -q$. Now we have to find the eigen vectors, this is:

$$\begin{pmatrix} -(p+r) - -(p+r) & 0 \\ r & -q - -(p+r) \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Using both $\lambda_1 = -(p+r)$ and $\lambda_2 = -q$, we find that eigen-vectors are:

$$\vec{\eta_{\lambda_1}} = \left(1, \frac{r}{q-p-r}\right) \tag{12}$$
$$\vec{\eta_{\lambda_2}} = (0, 1) \tag{13}$$

Now, we have that the general solution is:

$$\vec{x(t)} = C_1 e^{\lambda_1 t} \vec{\eta_{\lambda_1}} + C_2 e^{\lambda_2 t} \vec{\eta_{\lambda_2}} \tag{14}$$

Using initial conditions we find that $C_1 = N$ and $C_2 = \frac{Nr}{P+r-q}$. Therefore:

$$u(t) = Ne^{-(p+r)t} \tag{15}$$
$$v(t) = \frac{Nr}{p+r-q}\left(e^{-qt} - e^{-(p+r)t}\right) \tag{16}$$

Finally, the bi-exponential model is:

$$S(t) = N\left[e^{-(p+r)t} + \frac{r}{p+r-q}\left(e^{-qt} - e^{-(p+r)t}\right)\right] \tag{17}$$

Transition Time

An interesting parameter here is the critical time, this is the time when the temporal scale occurs. We will calculate the critical time defined as:

$$\left.\frac{d\log S}{dt}\right|_{t=t_c} = -(1+\delta)q, \tag{18}$$

where $\delta \sim 1$.

$$\frac{d\log S}{dt} = \frac{1}{S}\left[-(p+r)e^{-(p+r)t_c} + \frac{r}{p+r-q}\left(-qe^{-qt_c} + (p+r)e^{-(p+r)t_c}\right)\right] \tag{19}$$
$$\frac{d\log S}{dt} = \frac{1}{S}\left[(p+r)e^{-(p+r)t_c}(\frac{r}{p+r-q} - 1) + \frac{rq}{p+r-q}e^{-qt_c}\right] \tag{20}$$

By definition, $S(t_c) \approx \frac{r}{p+r-q}e^{-qt_c}$. Therefore,

$$\left.\frac{d\log S}{dt}\right|_{t=t_c} \approx \frac{(p+r)(p+r-q)}{r}e^{-(p+r-q)}(\frac{-p+q}{p+r-q}) + q = -(1+\delta)q \tag{21}$$
$$\implies -(1+\delta)q = \frac{(p+r)}{r}e^{-(p+r-q)}(-p+q) + q \tag{22}$$
$$\implies q\delta = \frac{(p+r)}{r}e^{-(p+r-q)}(p-q) \tag{23}$$
$$\implies t_c(\delta) = \frac{1}{p+r-q}\left[\log\left(\frac{(p+r)(p-q)}{rq}\right) - \log\delta\right] \tag{24}$$

In the main text we have used $\delta = 1$, meaning that we have defined the critical time t_c as the time when the decay rate of S is equal to $2q$.

Figure 5: Goodness of fit for all cohorts of APS papers ($n = 485,105$) and USPTO patents ($n = 1,681,690$). A) Difference of the AICc for Log-normal and Bi-exponential decay function for APS papers. B) Difference of the R^2 for Bi-exponential and Log-normal decay functions for APS papers. C) Difference of the AICc for Log-normal and Bi-exponential decay function for USPTO patents. D) Difference of the R^2 for Bi-exponential and Log-normal decay functions for USPTO patents. The gray zones on A and C represent the non-significant difference between two models. Black lines on B and D represent equal goodness of fit. We note that bi-exponential model outperforms log-normal model, especially in the long-term description. All the R^2 in B and D have a $p - value < 0.001$

Model Fitting

We fit our model to papers, patents, songs, movies and biographies data. In particular, and motivated for accuracy, we fit the logarithm of the equation 6, this means we fit the follow equation:

$$log(\overline{S(t)}) = log\left[\frac{N}{p+r-q}\left[(p-q)e^{-(p+r)t} + re^{-qt}\right]\right], \tag{25}$$

where $\overline{S(t)}$ corresponds to the average of new citations for papers and patents, and $\overline{S(t)} = \big(S(t) - \overline{Pop}\big)/\sigma_{Pop}$ corresponds to the standardized current popularity for songs, movies and biographies, represents. $\overline{Pop}$ is the average popularity and σ_{Pop} is the standard deviation of current popularity of the decay curves. Those results are shown in the main text for songs (Figures 2 G and H), movies (Figure 2 I), tennis players (Figure 2 J), Olympic medalist (Figure 2 K), and basketball players (Figure 2 L).

Goodness of Fit

We analyzed three level of accomplishment, (k^*), for each cohort of APS papers (507 groups of papers) and USPTO patents (480 groups of patents). We compute the Akaike information criterion (Figure 5 A and C), to compare the bi-exponential and log-normal models, corrected by the number size of the sample as follow:

$$AICc = 2k - 2\ln(\hat{L}) + \frac{2k^2 + 2k}{n - k - 1}, \tag{26}$$

where $\hat{L}$ is the maximum value of the likelihood function for the model. Also, we calculate the R^2 as the square of the correlation between the observed value and the predicted value. We observe in Fig. 5 A and B that the difference for AICc in both, papers and patents, is significantly bigger than two. It means that after accounting by the size of the sample and the number of parameters of the model, the bi-exponential decay present substantial evidence to be better describing the whole decay (We note that a lower AICc means less information lost in the fit, that's the reason why the difference is positive in the figure). The gray stripe represents the zone where both, log-normal and bi-exponential are equally good at describing the behavior. We observe that even after correcting by the size of the sample and by penalizing the number of parameters, the bi-exponential model offers a more accurate description of the decay function. We also calculate the difference of the adjusted pseudo-R^2, Figure 5 B and D, between Bi-exponential and log-normal decay. We observe in both, papers and patents that the R^2 is bigger for Bi-exponential decay. We observe that the bi-exponential model is always better than log-normal, especially when it comes to the long-term behavior of the decay. All models presented in Figure 3 are summarized in Supplementary Tables 1, 2, and 3.

Data Availability

The datasets from the American Physical Society, analyzed during the current study is available in the APS Data Sets for Research repository, under request https://journals.aps.org/datasets.

The datasets of United State Patents analyzed during the current study is available in the NBER repository, http://www.nber.org/patents/.

The datasets for Songs, Movies, and Biographies generated during and analyzed during the current study are available from the corresponding authors on reasonable request.

Code Availability

The entire analysis, data processing, and fitting was done using the standard R libraries (https://www.r-project.org/).

Acknowledgments

C. C. and C. R-S. acknowledges the financial support from Centro de Investigación en Complejidad Social (CICS) and Universidad del Desarrollo. C. J-F. and C.A.H. acknowledge support from the MIT Media Lab Consortia. The authors also thank Flavio Pinheiro, Tarik Roukny, Gustavo Castro-Dominguez, the Centro de Investigación en Complejidad Social (CICS), the Collective Learning Group at the MIT Media Lab, and the Center for Complex Network Research (CCNR) at Northeastern University for the helpful insights and discussions.

Author Contributions

C.C., C.A.H., A.-L. B, contributed to the study conception and design, interpretation of data, and drafting the manuscript. C. C. and C.J-F contributed to the acquisition of data, data analysis, modeling, and drafting the manuscript; C. R-S. contributed to study conception and design, and interpretation of data.

Competing Interests

We declare that the authors have no competing financial or non-financial interests as defined by Nature Research.

References

1. Halbwachs, M. *La mémoire collective* (Albin Michel, 1997).
2. Assmann, J. Collective Memory and Cultural Identity. *New German Critique* **65,** 125–133 (1995).
3. Assmann, J. *Das kulturelle Gedächtnis. Schrift, Erinnerung und politische Identität in frühen Hochkulturen* 9–27 (Verlag C.H. Beck, 2007).
4. Assmann, J. in *Cultural Memories* 15–27 (Springer, 2011).

5. Wertsch, J. V. *Voices of Collective Remembering* (Cambridge University Press, 2002).

6. Goldhammer, A., Nora, P. & Kritzman, L. D. *Realms of Memory: The Construction of the French Past* (Columbia University Press, 1998).

7. Roediger, H. & DeSoto, K. Forgetting the presidents. *Science* **346,** 1106–1109 (2014).

8. Roediger, H., Zaromb, F. M. & Butler, A. C. The role of repeated retrieval in shaping collective memory. *Memory in mind and culture,* 29–58 (2009).

9. Zaromb, F., Butler, A. C., Agarwal, P. K. & Roediger III, H. L. Collective memories of three wars in United States history in younger and older adults. *Memory & cognition* **42,** 383–399 (2014).

10. Rubin, D. C. How quickly we forget. *Science* **346,** 1058–1059 (2014).

11. García-Gavilanes, R., Mollgaard, A., Tsvetkova, M. & Yasseri, T. The memory remains: Understanding collective memory in the digital age. *Science Advances* **3** (2017).

12. Wang, D., Song, C. & Barabási, A.-L. Quantifying long-term scientific impact. *Science (New York, N.Y.)* **342,** 127–32 (2013).

13. Higham, K. W., Governale, M., Jaffe, A. B. & Zülicke, U. Fame and obsolescence: Disentangling growth and aging dynamics of patent citations. *Physical Review E* **95** (2017).

14. Higham, K., Governale, M., Jaffe, A. & Zülicke, U. Unraveling the dynamics of growth, aging and inflation for citations to scientific articles from specific research fields. *Journal of Informetrics* **11,** 1190 –1200 (2017).

15. Valverde, S., Solé, R. V., Bedau, M. A. & Packard, N. Topology and evolution of technology innovation networks. *Physical Review E* **76,** 056118 (Nov. 2007).

16. Csárdi, G., Strandburg, K. J., Zalányi, L., Tobochnik, J. & Érdi, P. Modeling innovation by a kinetic description of the patent citation system. *Physica A: Statistical Mechanics and its Applications* **374,** 783–793 (Feb. 2007).

17. Golosovsky, M. & Solomon, S. Stochastic Dynamical Model of a Growing Citation Network Based on a Self-Exciting Point Process. *Physical Review Letters* **109,** 098701 (Aug. 2012).

18. Dorogovtsev, S. N. & Mendes, J. F. F. Evolution of networks with aging of sites. *Phys. Rev. E* **62,** 1842–1845 (2 Aug. 2000).

19. Barabási, A.-L. & Albert, R. Emergence of Scaling in Random Networks. *Science* **286,** 509–512 (Oct. 1999).

20. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics* **74,** 47–97 (Jan. 2002).

21. Price, D. D. S. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science* **27,** 292–306 (Sept. 1976).

22. Yule, G. U. A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* **213,** 21–87 (1925).

23. Merton, R. K. The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property. *Isis* **79,** 606–623 (1988).

24. Assmann, J. Communicative and Cultural Memory. *Cultural Memory Studies. An International and Interdisciplinary Handbook,* 109–118 (2008).

25. Hirst, W., Yamashiro, J. K. & Coman, A. Collective Memory from a Psychological Perspective. *Trends in Cognitive Sciences* **22,** 438–451 (2018).

26. Roediger, H. L. & DeSoto, K. A. Recognizing the Presidents: Was Alexander Hamilton President? *Psychological Science* **27,** 644–650 (2016).

27. Rubin, D. C. *Memory in oral traditions: The cognitive psychology of epic, ballads, and counting-out rhymes.* (1995).

28. Wertsch, J. V. *Voices of Collective Remembering* (Cambridge University Press, Cambridge, 2002).

29. Hammack, P. L. Narrative and the politics of meaning. *Narrative Inquiry* (2011).

30. Sperber, D. & Hirschfeld, L. A. *The cognitive foundations of cultural stability and diversity* 2004.

31. Buskell, A. What are cultural attractors? *Biology & Philosophy* **32,** 377–394 (2017).

32. Richerson, P. J. & Boyd, R. *Not by Genes Alone: How Culture Transformed Human Evolution* (2005).

33. Storm, B. C., Bjork, E. L. & Bjork, R. A. On the durability of retrieval-induced forgetting. *Journal of Cognitive Psychology* **24,** 617–629 (2012).

34. Garcia-Bajos, E., Migueles, M. & Anderson, M. Script knowledge modulates retrieval-induced forgetting for eyewitness events. *Memory* (2009).

35. Cuc, A., Koppel, J. & Hirst, W. Silence is not golden: A case for socially shared retrieval-induced forgetting. *Psychological Science* (2007).

36. Echterhoff, G., Higgins, E. T. & Levine, J. M. Shared Reality: Experiencing Commonality with others' Inner States about the World. *Perspectives on Psychological Science* (2009).

37. Coman, A. & Hirst, W. Social identity and socially shared retrieval-induced forgetting: The effects of group membership. *Journal of Experimental Psychology: General* (2015).

38. Coman, A., Stone, C. B., Castano, E. & Hirst, W. Justifying atrocities: The effect of moral-disengagement strategies on socially shared retrieval-induced forgetting. *Psychological Science* (2014).

39. Stone, C. B., Barnier, A. J., Sutton, J. & Hirst, W. Building consensus about the past: Schema consistency and convergence in socially shared retrieval-induced forgetting. *Memory* (2010).

40. Kurzban, R., Tooby, J. & Cosmides, L. Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences* **98,** 15387–15392 (2001).

41. Yu, A. Z., Ronen, S., Hu, K., Lu, T. & Hidalgo, C. A. Pantheon 1.0, a manually verified dataset of globally famous biographies. *Scientific Data* **3,** 150075 (Jan. 2016).

42. Jara-Figueroa, C., Yu, A. Z. & Hidalgo, C. A. Estimating technological breaks in the size and composition of human collective memory from biographical data. *arXiv.org* (2015).

43. Skiena, S & Ward, C. B. *Who's Bigger?: Where Historical Figures Really Rank* (2014).

44. Kanhabua, N., Nguyen, T. N. & Niederée, C. *What triggers human remembering of events? A large-scale analysis of catalysts for collective memory in Wikipedia* in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries* (2014).

45. Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. Atypical Combinations and Scientific Impact. *Science* **342,** 468–472 (2013).

46. Mukherjee, S., Romero, D. M., Jones, B. & Uzzi, B. The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: The hotspot. *Science Advances* **3** (2017).

47. Ronen, S. *et al.* Links that speak: The global language network and its association with global fame. *Proceedings of the National Academy of Sciences* (2014).

48. Ferron, M. & Massa, P. Beyond the encyclopedia: Collective memories in Wikipedia. *Memory Studies* **7,** 22–45 (2014).

49. Yucesoy, B. & Barabási, A.-L. Untangling performance from success. *EPJ Data Science* **5,** 17 (2016).

50. Halbwachs, M. *On Collective Memory* 244 (University of Chicago Press, 1992).

51. Fortunato, S. *et al.* Science of science. *Science* **359** (2018).

52. Radicchi, F., Fortunato, S. & Castellano, C. Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences* **105,** 17268–17272 (2008).

53. Kuhn, T., Perc, M. c. v. & Helbing, D. Inheritance Patterns in Citation Networks Reveal Scientific Memes. *Phys. Rev. X* **4,** 041036 (4 Nov. 2014).

54. Sinatra, R., Wang, D., Deville, P., Song, C. & Barabasi, A.-L. Quantifying the evolution of individual scientific impact. *Science* **354** (Nov. 2016).

55. King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J. & West, J. D. Men Set Their Own Cites High: Gender and Self-citation across Fields and over Time. *Socius* **3,** 2378023117738903 (2017).

56. Chu, J. S. G. & Evans, J. A. *Too Many Papers? Slowed Canonical Progress in Large Fields of Science* Mar. 2018.

57. Yook, S.-H., Radicchi, F. & Meyer-Ortmanns, H. Self-similar scale-free networks and disassortativity. *Phys. Rev. E* **72,** 045105 (4 Oct. 2005).

58. Mukherjee, S., Uzzi, B., Jones, B. F. & Stringer, M. in *Knowledge and Networks* (eds Glückler, J., Lazega, E. & Hammer, I.) 243–267 (Springer International Publishing, Cham, 2017).

59. Shen, H.-W. & Barabasi, A.-L. Collective credit allocation in science. *Proceedings of the National Academy of Sciences* **111,** 12325–12330 (Aug. 2014).

60. Hall, B. H., Jaffe, A. B. & Trajtenberg, M. *The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools* NBER Working Papers 8498 (National Bureau of Economic Research, Inc, Oct. 2001).

61. Jaffe, A. B., Trajtenberg, M. & Henderson, R. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics* **108,** 577–598 (Aug. 1993).

62. Billboard. *Billboard Magazine The Hot 100 Ranking* 2016.

63. Spotify. *Spotify Web API* 2017.

64. Last.fm. *Last.fm Web API* 2017.

65. IMDB. *The Internet Movie Database* 2017.

66. YouTube. *YouTube Web Platform* 2017.

67. Wikipedia. *The Free Encyclopedia* 2017.