Postmortem memory of public figures in news and social media

Robert West[a,1] (ID), Jure Leskovec[b] (ID), and Christopher Potts[c]

[a]School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland; [b]Department of Computer Science, Stanford University, Stanford, CA 94305; and [c]Department of Linguistics, Stanford University, Stanford, CA 94305

Deceased public figures are often said to live on in collective memory. We quantify this phenomenon by tracking mentions of 2,362 public figures in English-language online news and social media (Twitter) 1 y before and after death. We measure the sharp spike and rapid decay of attention following death and model collective memory as a composition of communicative and cultural memory. Clustering reveals four patterns of postmortem memory, and regression analysis shows that boosts in media attention are largest for premortem popular anglophones who died a young, unnatural death; that long-term boosts are smallest for leaders and largest for artists; and that, while both the news and Twitter are triggered by young and unnatural deaths, the news additionally curates collective memory when old persons or leaders die. Overall, we illuminate the age-old question of who is remembered by society, and the distinct roles of news and social media in collective memory formation.

computational social science | collective memory | news and social media analysis | forgetting

Being remembered after death has been an important concern for humans throughout history (1), and conversely, many cultures have considered *damnatio memoriae*—being purposefully erased from the public's memory—one of the most severe punishments conceivable (2). To reason about the processes by which groups and societies remember and forget, the French philosopher and sociologist Maurice Halbwachs introduced the concept of collective memory in 1925 (3), which has since been a subject of study in numerous disciplines, including anthropology, ethnography, philosophy, history, psychology, and sociology, and which gave rise to the new discipline of memory studies (4). Over the decades, collective memory has moved from being a purely theoretical construct to becoming a practical phenomenon that can be studied empirically (5), e.g., in order to quantify to what extent US presidents are remembered across generations (6) or how World War II is remembered across countries (7).

Whereas oral tradition formed the basis for collective memory in early human history, today the media play a key role in determining what and who is remembered, and how (8–11). Researchers have studied the role of numerous media in constructing the postmortem memory of deceased public figures. A large body of work has investigated the journalistic format of the obituary (12–16), which captures how persons are remembered around the time of their death (14). Taking a more long-term perspective, other work has considered how deceased public figures are remembered in the media over the course of years and decades (17–21). As ever more aspects of life are shifting to the online sphere, the Web is also gaining importance as a global memory place (22), which has led researchers to study, e.g., how social media users (23–27) and Wikipedia editors (28) react to the death of public figures. In the context of social media, the detailed analysis of highly visible individual cases, such as Princess Diana (24), pop star Michael Jackson (25, 26), or race car driver Dale Earnhardt (27), has revealed how people experience and overcome the collective trauma that can ensue following the death of celebrities.

Although such studies of individuals have led to deep insights at a fine level of temporal granularity, they lack breadth by excluding all but some of the very most prominent public figures. What is largely absent from the literature is a general understanding of patterns of postmortem memory in the media that goes beyond single public figures.

To bridge this gap, we draw inspiration from a body of related work that has studied the temporal evolution of collective memory using large-scale datasets—although, unlike our work, not with a focus on the immediate postmortem period of public figures. For instance, van de Rijt et al. (20) tracked thousands of person names in news articles, finding that famous people tend to be covered by the news persistently over decades. In a similar analysis, Cook et al. (19) further showed that the duration of fame had not decreased over the course of the last century. Beyond news corpora, the online encyclopedia Wikipedia has become a prime resource for the data-driven study of collective memory. Researchers have leveraged the textual content of Wikipedia articles (29), as well as logs of both editing (30) and viewing (31, 32), as proxies for the collective memory of traumatic events such as terrorist attacks or airplane crashes. Jatowt et al. (33) characterized the coverage and popularity of historical figures in Wikipedia, observing vastly increased page-view counts for people from the 15th and 16th centuries, a fact that Jara-Figueroa et al. (34) later attributed to the invention of the printing press. In addition to news and

---

Significance

Who is remembered by society after they die? Although scholars as well as the broader public have speculated about this question since ancient times, we still lack a detailed understanding of the processes at work when a public figure dies and their media image solidifies and is committed to the collective memory. To close this gap, we leverage a comprehensive 5-y dataset of online news and social media posts with millions of documents per day. By tracking mentions of thousands of public figures during the year following their death, we reveal and model the prototypical patterns and biographic correlates of postmortem media attention, as well as systematic differences in how the news vs. social media remember deceased public figures.

---

www.manaraa.com
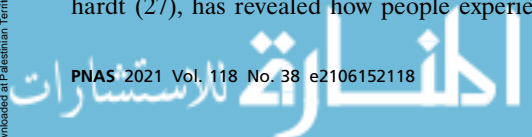
encyclopedic articles, books (35–37) and social media (38, 39) have also emerged as important assets for studying collective memory.

Whereas the above works are primarily descriptive in nature, researchers have also developed mathematical models of the growth and decay of collective memory. Notably, as part of a rich literature on the evolution of performance, fame, and success in the arts and sciences (40–45), Candia et al. (46) analyzed thousands of papers, patents, songs, movies, and athletes, showing that the decay of the intensity of collective memory can be well described by a biexponential function that captures two aspects of collective memory: communicative memory, which is "sustained by the oral transmission of information," and cultural memory, which is "sustained by the physical recording of information" (46).

We extend this literature by studying how the coverage of thousands of public figures in news and social media evolved during the year following their death. Our approach combines the Freebase knowledge base (47)—a comprehensive repository containing records for over 3 million public figures—with an extensive corpus of online news and social media compiled via the online media aggregation service Spinn3r (48), which comprises, for each day, hundreds of thousands of news articles from a complete set of all 6,608 English-language Web domains indexed by Google News and tens of millions of social media posts from Twitter, amounting to about one-third of full English Twitter (details in *Materials and Methods*; number of documents per day in *SI Appendix*, Fig. 1). The population of study consists of 2,362 public figures who died between 2009 and 2014 and received at least a minimum amount of premortem coverage both in the news and on Twitter. For each person, we tracked the daily frequency with which they were mentioned in the two media during the year before and the year after death, and operationalize postmortem memory via the resulting time series of mention frequency. (For details about data and preprocessing, see *Materials and Methods*.) Analyzing the mention time series allowed us to quantify the extreme spike and rapid decay of attention that tend to follow the death of public figures, a pattern well captured by a power law shifted by a constant additive offset. A cluster analysis of mention time series revealed four prototypical patterns of postmortem memory ("blip," "silence," "rise," and "decline"), and a regression analysis shed light on the biographic correlates of postmortem memory and on systematic differences between postmortem memory in mainstream news vs. social media. We conclude that the prototypical persona with the largest postmortem boost in English-language media attention can be described as an anglophone who was already well known before death and died a young, unnatural death. Long-term attention boosts are on average smallest for leaders and largest for artists. Finally, while both the mainstream news and Twitter are triggered by young and unnatural deaths, the mainstream news—but not Twitter—appears to also assume an additional role as stewards of collective memory when an old person or an accomplished leader dies. Overall, the present work helps illuminate an age-old question: Who is remembered by society?

Results
We strive to characterize the patterns by which postmortem memory evolves during the year immediately following the death of public figures. When considering this time frame, prior work has primarily taken a qualitative stance, asking how, linguistically, the mainstream and social media speak about small sets of deceased people (15, 16, 23–27). In contrast, enabled by a comprehensive corpus of news and social media posts, we take a quantitative stance, asking about whom the media speak how much after death.

At the core of our analysis are time series of mention frequency. A person i's "raw mention time series" specifies, for each day t relative to i's day of death ($t = 0$), the base-10 logarithm of the fraction $S_i(t)$ of documents in which person i was mentioned, out of all documents published on day t. To reduce noise, we also generated "smoothed mention time series" using a variable span smoother based on local linear fits (49). For each person, separate time series were computed for the news and for Twitter (examples in Fig. 1; additional details about mention time series construction in *Materials and Methods*).

The Shape of Postmortem Memory: Communicative and Cultural Memory. Averaging the 2,362 raw mention time series (Fig. 2) exposes a sharp spike in the interest in public figures in the immediate wake of their death (on days 0 and 1), followed by a steep drop up until around day 30, where the curves elbow into a long, much flatter phase, which is only slightly disrupted by a small secondary spike on day 365 after death. The main spike is so strong that, without logarithmically transforming the fractions $S_i(t)$ of mentioning documents, no interesting information besides the dominant main spike would be visually discernible.

In a model that is conceptually similar to Candia et al.'s biexponential model (46), we decompose the postmortem collective memory $S(t)$ into a sum of two components, $S(t) = u(t) + v(t)$, where $u(t)$ captures "communicative memory," and $v(t)$, "cultural memory" (see Introduction). Communicative memory is modeled as a decaying power law, i.e., $u(t) = at^{-b}$,* whereas cultural memory is modeled as constant during the time frame considered here (400 d after death), i.e., $v(t) = c$. We refer to this model as a "shifted power law." It fits the empirical average mention time series ($R^2 = 0.99$; details on model fitting in *Materials and Methods*) significantly better than any of eight alternative models from the literature (46, 50, 51) (details in *SI Appendix*, Fig. 2), including the biexponential model (46). The best shifted power law fit is shown as a black line in Fig. 2; the communicative and cultural memory components are plotted separately in green and red, respectively.

The fitted decay parameter of communicative memory is similar for the news ($b = 1.34$) and for Twitter ($b = 1.54$). Communicative memory starts high on day $t = 1$, but drops below cultural memory quickly, after 14 and 18 d in the news and on Twitter, respectively, and accounts for only 25% of total collective memory after 31 and 36 d, respectively, which constitutes an inflection point where communicative memory levels off strongly (*SI Appendix*, Fig. 3). Moreover, even though no premortem data were used in fitting the model, the constant cultural memory c closely approximates the average premortem fraction of mentioning documents in both media (cf. Fig. 2).

This suggests that, on average, public figures build up a certain baseline amount of (cultural) memory during their lifetime, on top of which a burst of quickly fading communicative memory is layered in the wake of death. Note that, although collective memory rapidly reverts to the premortem level when averaging over all people, it need not be so for individual people, as we shall see below (Fig. 3B).

Further evidence in support of two distinct memory modes comes from the fact that the average length of documents that mention a public figure dropped sharply with death (possibly due to brief death notes and obituaries) and reached the premortem level again after about 30 d (*SI Appendix*, Fig. 4), i.e., around the inflection point where communicative memory levels off according to the fitted model.

---

*With $u(t) = at^{-b}$, we have $du(t)/dt = -(b/t)\,u(t)$, so the forgetting rate b/t (fraction of memory lost at time t) decreases over time, in line with what is suggested by the psychological literature (50).

www.manaraa.com
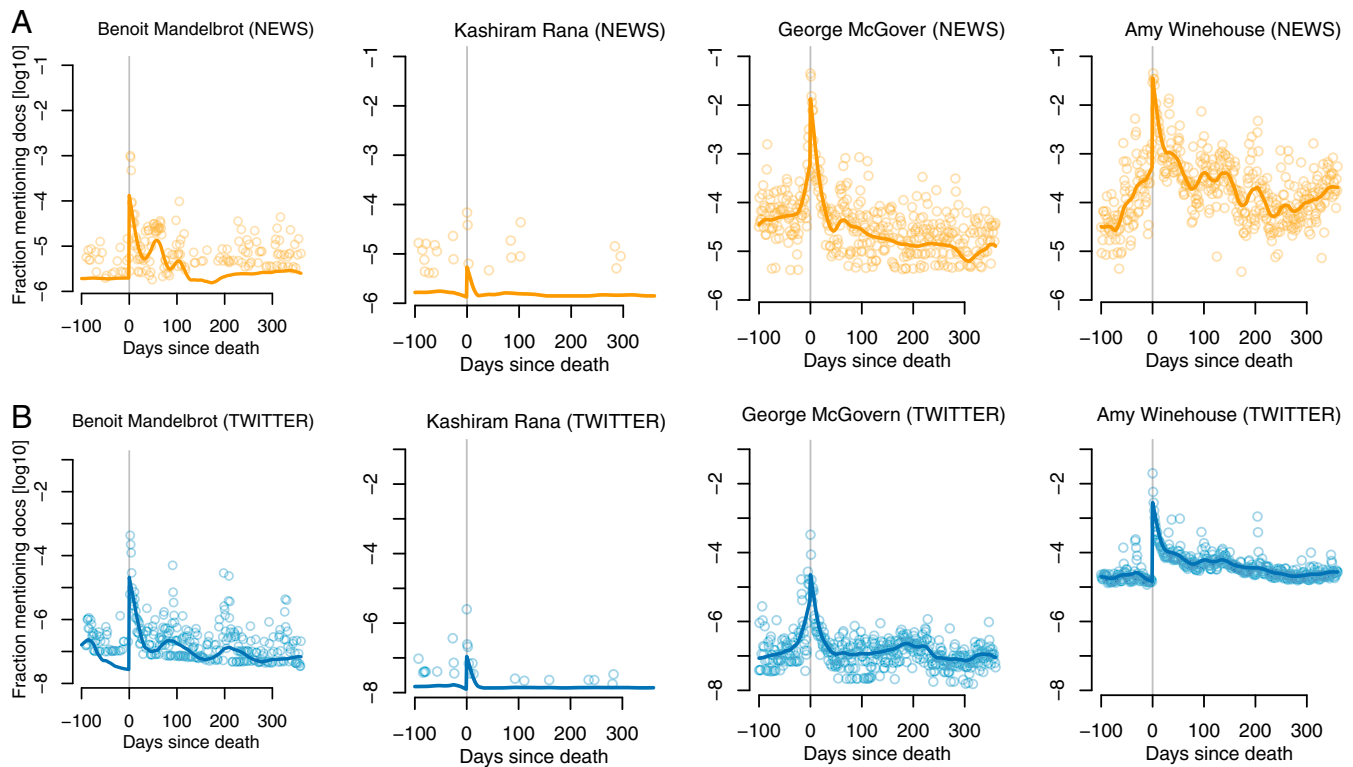
Fig. 1. Examples of mention time series for four deceased public figures, as observed (*A*) in the news and (*B*) on Twitter. In mention time series, the *x* axis specifies the number of days since death, and the *y* axis, the base-10 logarithm of the fraction of documents in which the person was mentioned that day, out of all documents published that day. Light circles correspond to raw mention time series, and dark curves, to their smoothed versions.

We hence divide the postmortem period into two phases: short-term (days 0 through 29) and long-term (days 30 through 360). Based on this distinction, in order to reason about the shape of mention time series, we summarize each time series by four characteristic numbers (depicted graphically in Fig. 3*A*):

1. Premortem mean: arithmetic mean of days 360 through 30 before death.

2. Short-term boost: maximum of days 0 through 29 after death, minus the premortem mean.

3. Long-term boost: arithmetic mean of days 30 through 360 after death, minus the premortem mean.

4. Halving time: number of days required to accumulate half of the total area between the postmortem curve (including the day of death) and the minimum postmortem value.
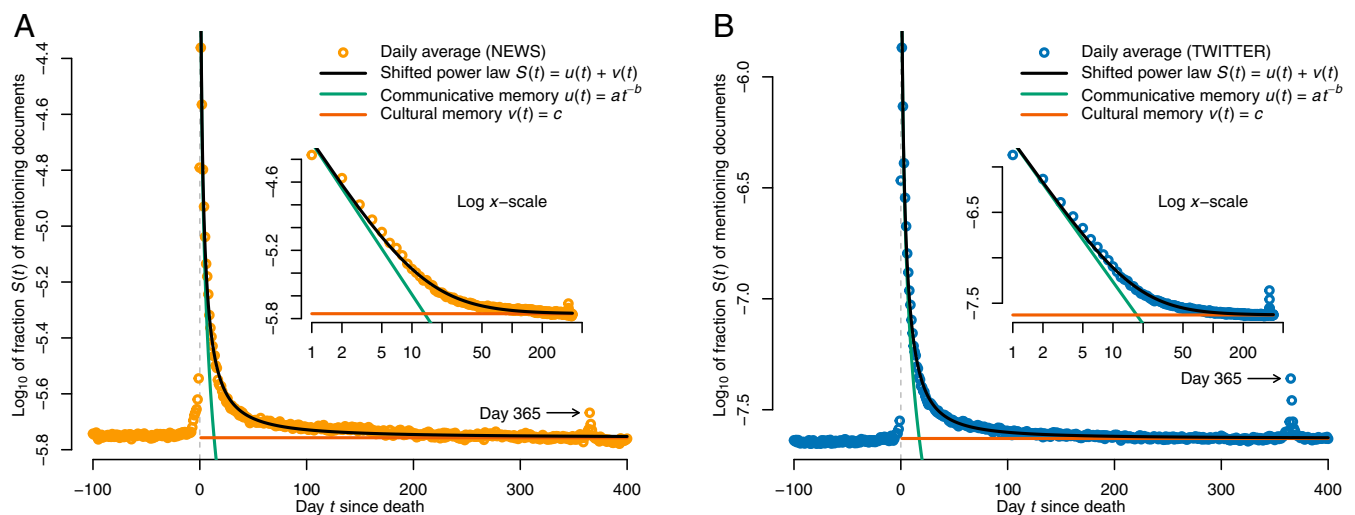
Fig. 2. Average mention time series, obtained via the arithmetic mean of the individual raw mention time series of the 2,362 people included in the study, (*A*) in the news and (*B*) on Twitter (see Fig. 1 for examples of individual mention time series). On average, the mention frequency of deceased public figures spikes by about 9,400% in the news, and by about 28,000% on Twitter, when they die, and fades quickly thereafter, with a minor secondary spike on the death anniversary. We also plot the best fit of the shifted power law model (black), which decomposes the total collective memory $S(t) = u(t) + v(t)$ on day t into a sum of communicative memory $u(t) = at^{-b}$ (green) and cultural memory $v(t) = c$ (red). *Insets* show the same data and fits on logarithmic *x* axes.

West et al.
Postmortem memory of public figures in news and social media
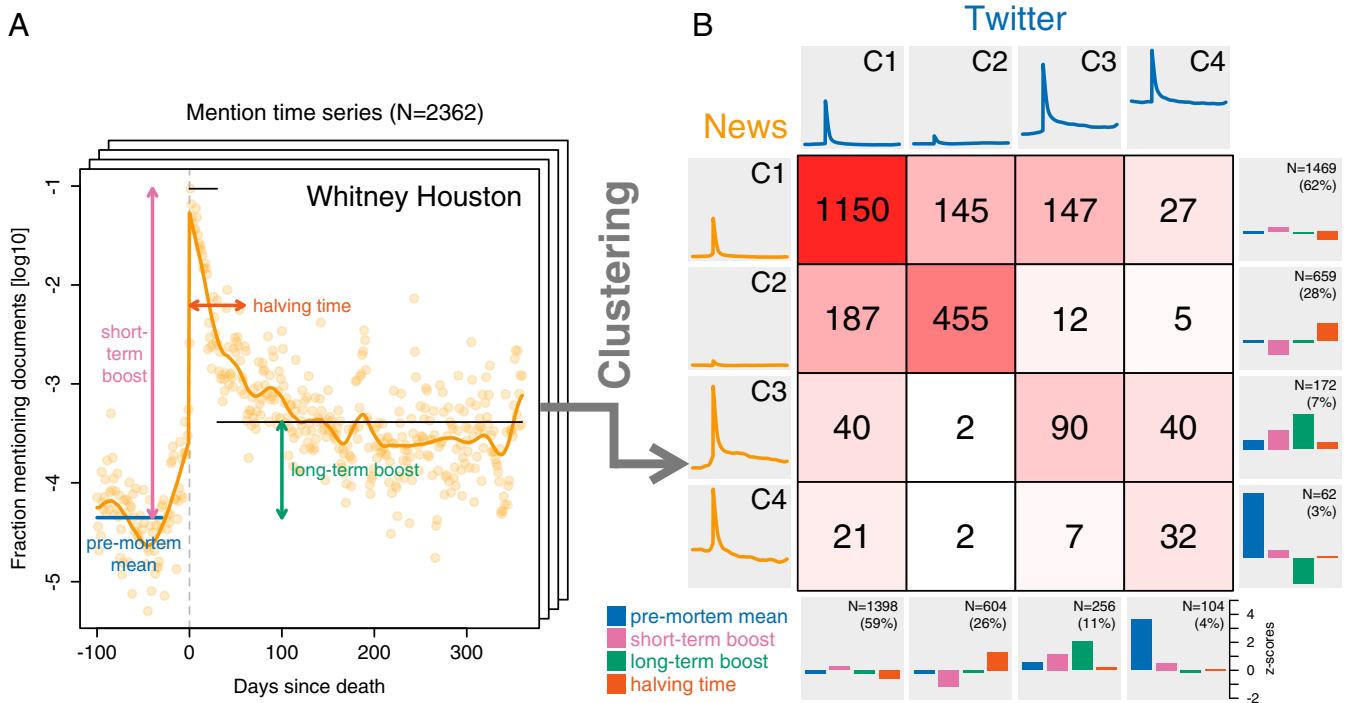
Fig. 3. Cluster analysis of mention time series. (*A*) From each mention time series, we extract four characteristic numbers: premortem mean, short-term boost, long-term boost, and halving time. In the resulting four-dimensional space, time series are clustered using the *k*-means algorithm. According to the average silhouette criterion, the optimal number of clusters is $k = 4$ both in the news and on Twitter. (*B*) Nearly identical clusters (C1 through C4) emerge independently in the news and on Twitter, in terms of both cluster centroids and cluster sizes. Cluster centroids are depicted as bar charts in the right (news) and bottom (Twitter) margins; average mention time series for each cluster, in the left (news) and top (Twitter) margins. As captured by the confusion matrix, whose diagonal entries are much larger than under a null model that assumes the two media to be independent, a given person tends to fall into the corresponding clusters in the two media.

All characteristics were computed on the smoothed time series, with the exception of the maximum used in the short-term boost, which was computed on the raw time series. The 29 d immediately before death were excluded from the premortem mean in an effort to exclude a potential rise in interest in people whose impending death might have been anticipated, e.g., due to illness. Since the time series capture logarithmic mention frequencies, the (arithmetic) premortem mean corresponds to the logarithm of the geometric mean mention frequency, and the short- and long-term boosts, to the logarithm of the multiplicative increase over the premortem geometric mean mention frequency.

Magnitude of Short- and Long-Term Boosts. Aggregating the short-term boost over all public figures allows us to quantify the sharp spike immediately after death observed in Fig. 2. The median short-term boost was 1.98 (95% CI [1.90, 2.03]) in the news, and 2.45 (95% CI [2.37, 2.50]) on Twitter. (All curve characteristics are summarized in *SI Appendix*, Table 2 and Fig. 5.) The boost was significantly stronger on Twitter (Wilcoxon's signed-rank test: $W = 477\,893$, two-sided $P < 10^{-15}$), where it approximately corresponded to a 28,000% increase on the linear scale ($10^{2.45} \approx 281$), compared to a 9,400% increase in the news ($10^{1.98} \approx 95$).

After the immediate spike, media interest tended to fade quickly. In the news, no important long-term boost was observed (median 0.000545, 95% CI [−0.000908, 0.00171]), whereas on Twitter, we measured a significantly larger (Wilcoxon's signed-rank test: $W = 881\,590$, two-sided $P < 10^{-15}$) long-term boost of 0.0160 in the median (95% CI [0.0133, 0.0175]), translating to a 3.8% increase on the linear scale ($10^{0.016} \approx 1.038$).

Cluster Analysis of Mention Time Series. Mention time series expose a great variety of curve shapes, a glimpse of which is given by the examples of Fig. 1. We hypothesized that, despite their diversity, mention time series could be grouped into distinct classes, a hypothesis that we explored in a cluster analysis. Time series were represented by their four characteristic numbers (premortem mean, short-term boost, long-term boost, halving time) in *z*-score-standardized form and clustered using the *k*-means algorithm. A separate clustering was performed for the news and for Twitter. Evaluating all numbers of clusters $k \in \{2, \ldots, 30\}$ via the average silhouette criterion (52) revealed a clear optimum for $k = 4$ clusters for both the news and Twitter (*SI Appendix*, Fig. 7).

The cluster centroids are visualized in the right and bottom margins of Fig. 3*B*; the right margin shows the centroids for the news, the bottom margin, for Twitter. Moreover, we plot the average smoothed mention time series for each cluster in the left (news) and top (Twitter) margins. (An overlay of all time series per cluster is plotted in *SI Appendix*, Fig. 8) Strikingly, although the clustering was performed independently for the news and for Twitter, respectively, the centroids that emerged—as well as the number of data points in each cluster—are nearly identical. The resulting clusters, which we name C1 through C4 in order of decreasing size, can be described as follows:

C1 ("blip"): Average mention frequency pre- as well as post-mortem, with a short-term boost of average magnitude in between (62% of people in the news; 59% on Twitter).
C2 ("silence"): Average mention frequency pre- as well as post-mortem, with a faint short-term boost of below-average magnitude in between (28% in the news; 26% on Twitter).

C3 ("rise"): High premortem mention frequency, large short-term boost, followed by an extreme long-term boost (7% in the news; 11% on Twitter).

C4 ("decline"): Extreme premortem mention frequency, above-average short-term boost, followed by a below-average long-term boost (3% in the news; 4% on Twitter).

In both media, over half of the people (59–62%) fall into cluster C1; their time series resemble the overall average (Fig. 2), with a brief spike after death and a quick drop to the—usually low—premortem level. About half of the remaining people (26–28%) fall into cluster C2; their time series are similar to those of C1, with the difference that the death of people in C2 went largely unnoticed. About half of the people outside of C1 and C2 (7–11%) fall into C3, which mostly contains people who were popular already before death and experienced a large boost in attention in both the short and the long term. The final cluster, C4, is composed of a tiny elite (3–4%) of people of an extreme premortem popularity that tended to fade postmortem. The long-term decrease was considerably stronger in the news than on Twitter in this cluster.

Not only do nearly identical clusters of nearly identical size emerge in the news as on Twitter; a given person also tends to fall into the corresponding clusters in the two media, as captured by the cluster confusion matrix (Fig. 3B), which counts, for all $i, j \in \{1, 2, 3, 4\}$, the number of people falling into news cluster i and Twitter cluster j. Using Pearson's χ^2 test, we reject the null hypothesis under which cluster membership is assumed to be independent in the news vs. Twitter, given the empirical marginal cluster sizes ($\chi^2 = 11739$, $P < 10^{-5}$). In particular, all diagonal entries of the confusion matrix are strongly overrepresented,[†] whereas all but two off-diagonal entries are underrepresented, and "C3 in news, C4 on Twitter" is the only off-diagonal entry to be significantly overrepresented (one-sample proportions test with continuity correction: $\chi^2 = 135$, $P < 10^{-15}$).

Biographic Correlates of Postmortem Memory. Next, we aim to understand what premortem properties of a person are associated with their postmortem mention frequency. A naive correlational analysis would not suffice for this purpose, as personal properties are correlated with one another; e.g., leaders (politicians, CEOs, etc.) in the dataset are more likely to have died old and of a natural death, and are more likely to be men, compared to artists. In order to disentangle such correlations, we performed a regression analysis. We fitted linear regression models for two outcomes:

1. short-term boost,
2. long-term boost,

and with six predictors in either case:

1. premortem mean mention frequency,
2. age at death (factor with eight levels: 20–29, 30–39, ..., 90–99),
3. manner of death (factor with two levels: natural, unnatural),
4. notability type (factor with six levels, specifying a profession or role for which the person was most known: arts, sports, leadership [including politicians, business/organization leaders, religious leaders, military, etc.], known for death [including disaster victims], general fame, academia/engineering),
5. language (factor with three levels: anglophone, non-anglophone, unknown), and
6. gender (factor with two levels: female, male).

Out of all 2,362 people, the regression analysis only included the 870 people for whom all factors took one of the above-defined levels (details on definition of biographic features in *Materials and Methods*; distribution summarized in *SI Appendix, Table 1*). All factor variables were dummy-coded as binary indicators. The premortem mean was first rank-transformed and then linearly scaled and shifted to the interval $[-0.5, 0.5]$. Additionally, 70–79 y (which contains the mean and median age at death) was chosen as the default age level, and the most frequent level was chosen as the default for all other factors, such that the regression intercept captures the average outcome for a "baseline persona" representing male anglophone artists of median premortem popularity who died a natural death at age 70–79. With the above, a coefficient β for a binary predictor corresponds to an additive boost increase of β with respect to the baseline persona, or—since boosts are base-10 logarithms (of postmortem-to-premortem ratios of mention frequencies)—to a multiplicative postmortem-to-premortem ratio increase of 10^β. A separate regression model was fitted for each combination of medium (news or Twitter) and outcome (short- or long-term boost), for a total of four models.

The model coefficients (summarized in Table 1) paint a largely consistent picture for the news vs. Twitter. We observe that, in both media, *ceteris paribus*, both the short- and the long-term boost were larger for people who died an unnatural death, for people with an anglophone background, and for people who were already popular premortem. No significant gender variation was detected, with the exception of the long-term boost in the news, which was slightly larger for women. The only significant notability type was leadership, whose long-term boost was smaller than that of the baseline (arts) in both media.

The dependence of short- and long-term boosts on the age at death is displayed visually in Fig. 4 A and B. In order to determine whether the above finding that attention increased more for people who died an unnatural death holds across age brackets, the plots are based on a slightly modified model with an additional "age by manner of death" interaction term. This allows us to estimate the average postmortem attention boost attained by each age bracket separately for people who died natural vs. unnatural deaths (as before, the estimates are for male anglophone artists of median premortem popularity). Inspecting the curves of Fig. 4 A and B, we make two observations. First, across age brackets, people who died an unnatural death received larger boosts, both short- and long-term, and both in the news and on Twitter. Second, the curves have a nonmonotonic U-shape for the news (Fig. 4A), but are monotonically decreasing for Twitter (Fig. 4B); i.e., the news increased attention most for those who died either very young or very old, whereas Twitter increased attention more the younger the deceased.

News vs. Twitter. The above analyses were done separately for the news and Twitter. In order to understand how postmortem memory of the same person differed between the two media, we conducted a pairwise analysis. We again fitted linear regression models with the same predictors as above, but this time with outcomes defined by the news-minus-Twitter difference in short- and long-term boosts. Accordingly, the rank-transformed and scaled premortem mean predictor was replaced with the news-minus-Twitter difference in rank-transformed and scaled premortem means. Given this setup, large positive coefficients mark groups of people who received particularly strong boosts in the news compared to Twitter, and large negative coefficients mark groups of people who received particularly strong boosts on Twitter compared to the news.

The model coefficients (summarized in Table 2) reveal that those who died an unnatural death, as well as leaders, received

---

[†]The empirical trace (the sum of diagonal elements) is 1,727. Under the null hypothesis, the expected trace would be 1,059. Given the marginal constraints, the minimum and maximum traces that could possibly be attained are 505 and 2,236, respectively (determined via linear programming). That is, the empirical trace attains 71% of the range between the minimum and the maximum, whereas the null model attains only 32%.

COMPUTER SCIENCES

SOCIAL SCIENCES

West et al.
Postmortem memory of public figures in news and social media

PNAS | 5 of 9
https://doi.org/10.1073/pnas.2106152118

www.manaraa.com

Table 1. Linear regression models of short- and long-term boosts in news and Twitter

	Short-term boost (news)	Short-term boost (Twitter)	Long-term boost (news)	Long-term boost (Twitter)
(Intercept)	2.322 (0.063)***	2.670 (0.067)***	0.088 (0.014)***	0.095 (0.015)***
Premortem mean (relative rank)	0.804 (0.093)***	0.948 (0.100)***	0.031 (0.020)	0.086 (0.022)***
Manner of death: unnatural	0.618 (0.095)***	0.282 (0.100)**	0.097 (0.021)***	0.106 (0.022)***
Language: non-anglophone	−0.316 (0.074)***	−0.116 (0.078)	−0.061 (0.016)***	−0.037 (0.017)*
Language: unknown	−0.446 (0.086)***	−0.325 (0.091)***	−0.079 (0.019)***	−0.081 (0.020)***
Gender: female	0.083 (0.072)	−0.034 (0.076)	0.034 (0.016)*	0.006 (0.017)
Notability type: academia/engineering	0.181 (0.197)	0.340 (0.208)	−0.032 (0.043)	0.023 (0.046)
Notability type: general fame	0.070 (0.124)	0.132 (0.131)	−0.010 (0.027)	−0.008 (0.029)
Notability type: known for death	−0.107 (0.099)	−0.088 (0.106)	−0.021 (0.022)	0.008 (0.023)
Notability type: leadership	0.152 (0.083)	0.113 (0.087)	−0.058 (0.018)**	−0.040 (0.019)*
Notability type: sports	0.049 (0.083)	0.072 (0.088)	−0.034 (0.018)	−0.034 (0.020)
Age: 20–29	0.162 (0.170)	0.718 (0.180)***	0.060 (0.037)	0.192 (0.040)***
Age: 30–39	0.400 (0.167)*	0.649 (0.177)***	0.028 (0.037)	0.118 (0.039)**
Age: 40–49	−0.046 (0.126)	0.351 (0.133)**	−0.001 (0.028)	0.100 (0.030)***
Age: 50–59	−0.075 (0.099)	0.181 (0.104)	−0.058 (0.022)**	−0.024 (0.023)
Age: 60–69	−0.109 (0.082)	0.130 (0.086)	−0.050 (0.018)**	−0.025 (0.019)
Age: 80–89	0.022 (0.078)	0.021 (0.082)	−0.018 (0.017)	−0.013 (0.018)
Age: 90–99	0.174 (0.098)	0.034 (0.103)	−0.011 (0.021)	−0.024 (0.023)
R^2	0.213	0.192	0.123	0.178
Adj. R^2	0.197	0.176	0.106	0.161
No. obs.	870	870	870	870
RMSE	0.772	0.815	0.169	0.181

SEs of coefficients are in parentheses. ***$P < 0.001$, **$P < 0.01$, and *$P < 0.05$.

particularly large short-term attention boosts in the news compared to Twitter. Conversely, premortem popular people and those with a non-anglophone background received particularly large short-term attention boosts on Twitter compared to the news, the latter possibly because English is the most globally connected language (53), such that Twitter posts, even though all written in English, stemmed from a more geographically and culturally diverse set of writers than news articles, which originated exclusively from anglophone outlets by design. Other than leaders, no further notability type was significantly associated with either outcome, and no significant gender variation was observed. Finally, the age dependence is visualized in Fig. 4*C*, which shows that, the older a person, the larger the news-minus-Twitter difference in boosts, confirming that news media favored older people more than Twitter did, both short- and long-term.

Discussion

Our analysis of mention frequencies over time revealed that, for the majority of public figures, a sharp pulse of media attention immediately followed death, whereby mention frequency increased by 9,400% in the news, and by 28,000% on Twitter, in the median. The average mention frequency then declined sharply, with an inflection point around 1 mo after death, from where on it decayed more slowly, eventually converging toward the premortem level. These two stages are consistent with a model that posits two components of collective memory: a constant baseline level of cultural memory built up during life, and an added layer of communicative memory that is sparked by death and usually decays in a matter of days according to a power law. A cluster analysis of the mention time series revealed a set of four prototypical memory patterns ("blip," "silence," "rise," and "decline"). The same set of patterns emerged independently in the news and on Twitter, and the same person tended to fall into the same cluster across the two media.

In our regression analysis of biographic correlates of postmortem memory, out of all notability types (arts, sports, leadership, known for death, general fame, academia/engineering),

only leadership (politicians, business leaders, etc.) stood out significantly, being associated with a particularly low boost in long-term memory (Table 1). One might wonder if this fact could simply be explained by a regression to the mean, since leaders had the highest premortem mention frequencies in the news (*SI Appendix*, Fig. 6*A*). Note, however, that on Twitter, too, leaders saw the lowest long-term boosts, despite the fact that, on Twitter, it is artists—not leaders—who had the highest premortem mention frequencies (*SI Appendix*, Fig. 6*B*). We thus consider a regression to the mean an unlikely explanation for the low long-term boosts of leaders, as it could not simultaneously explain the situation in both media. Rather, we speculate that the lives and legacies of people of different notability types might differ systematically. Considering that nearly all (8 out of 10) long-term boost coefficients for notability types are negative in Table 1 (and none are significantly positive), the distinction to be made is in fact not that between leadership and the rest, but rather that between arts (the default notability type) and the rest. Based on this observation, we hypothesize that artists remain more present in the collective memory because they not only tend to be active performers during their lifetime, but also frequently leave a legacy of artwork that can long survive them, whereas leaders, athletes, etc., are noteworthy primarily for the actions they take during their lifetime, and are of much decreased interest to the media once they cannot take action anymore—an effect that seems to be most pronounced for leaders.[‡]

The low coefficients of determination (adjusted R^2) of the linear regression models, ranging from 0.106 to 0.197 (Table 1), serve as a testimony of the richness of human lives and legacies, which cannot be captured by statistical models relying on just a few biographic variables. Given the inherent unpredictability of

[‡]In terms of interesting exceptions to this rule, certain leaders saw large long-term boosts in the news (e.g., Zimbabwean military Solomon Mujuru, who died in a fire in circumstances considered suspicious by some), on Twitter (e.g., World War II veterans Richard Winters and Maurice Brown), or in both media (e.g., entrepreneur and activist Aaron Swartz, suicide at age 26; Governor of Punjab Salman Taseer, assassinated at age 66).

www.manaraa.com
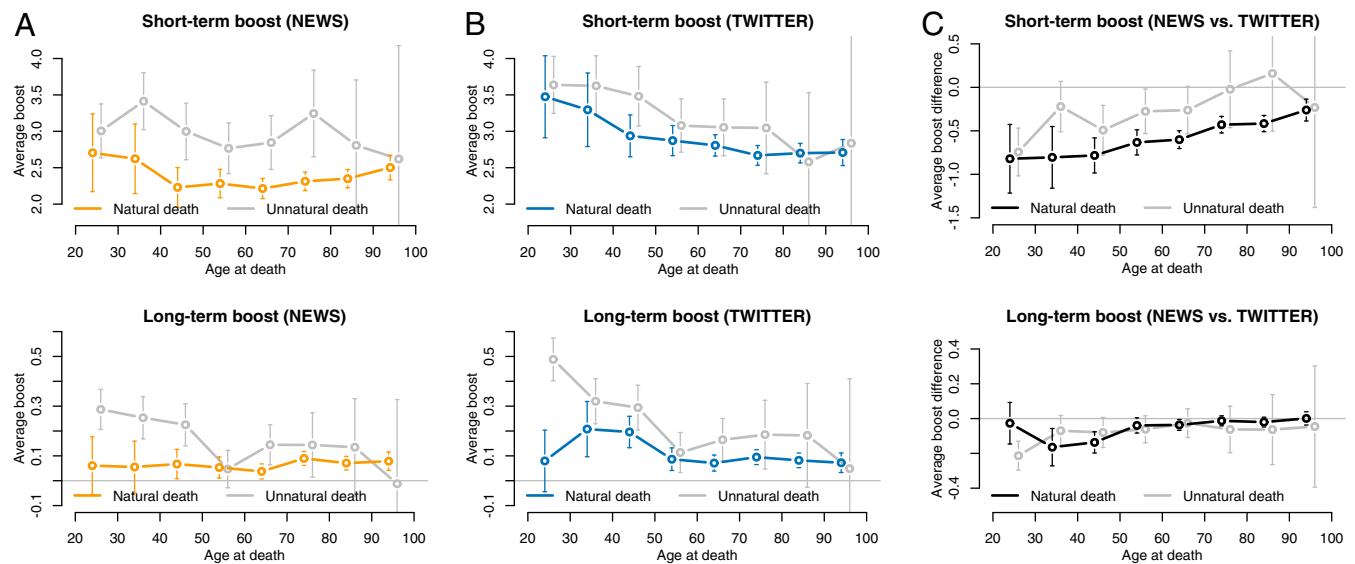
Fig. 4. Dependence of postmortem mention frequency on age at death for (*A*) the news and (*B*) Twitter, in terms of short-term (*Top*) and long-term (*Bottom*) boost, defined as the base-10 logarithm of the postmortem-to-premortem ratio of fractions of mentioning documents. Results were obtained via linear regression models that controlled for premortem mean, notability type, language, and gender. (*C*) Per-person news-minus-Twitter difference in short-term (*Top*) and long-term (*Bottom*) boosts. Error bars capture 95% CIs (approximated as ± 2 SEs). These plots show that unnatural deaths lead to larger attention boosts across age brackets, both short- and long-term; that the news increases attention most for those who die very young or very old, whereas Twitter increases attention more the younger the deceased person; and that the difference between attention boosts in news vs. Twitter is larger for those who die older (short- and long-term) and for those who die an unnatural death, across age brackets (short-term).

social systems (54, 55), this would be unlikely to change even if more biographic variables and more data points became available, and if more complex statistical models were to be used. We emphasize, however, that despite the inherent limits of predictability all model fits were highly significant ($p < 10^{-15}$ for the *F*-statistics of all models of Table 1; cf. *Regression modeling* in *SI Appendix*). Also, and most importantly, the effects were not only significant, but also large. As mentioned, a coefficient β for a binary predictor corresponds to a multiplicative increase of 10^{β} in the postmortem-to-premortem ratio of mention frequencies, compared to the baseline persona, a male anglophone artist of median premortem popularity who died a natural death at age 70–79. For example, *ceteris paribus*, an unnatural death quadrupled ($10^{0.618} \approx 4.15$) the short-term postmortem-to-premortem mention-frequency ratio in the news, and nearly doubled it ($10^{0.282} \approx 1.92$) on Twitter. The effect of age at death was also large. For instance, on Twitter, *ceteris paribus*, the short-term postmortem-to-premortem mention-frequency ratio for the 30–39 age bracket was twice that of the neighboring, 40–49 age bracket ($10^{0.649-0.351} \approx 1.99$); and that of the youngest age bracket was nearly five times that of the oldest age bracket ($10^{0.718-0.034} \approx 4.83$).

One of the key contributions of this study is the comparison between mainstream news and Twitter—a prominent social media platform—on a fixed set of attention subjects, thus extending a rich literature on the interplay between the two media (56–59). Despite the striking similarity of prototypical mention time series emerging from the cluster analysis (Fig. 3), the regression analysis revealed several noteworthy differences between postmortem memory in the news vs. Twitter. First, whereas on Twitter the postmortem boost was monotonically and negatively associated with age at death (Fig. 4*B*), we observed a nonmonotonic U-shaped relationship in the news (Fig. 4*A*), which provided the largest postmortem boost to both those who died very young and to those who died very old, an effect that even held for a fixed person (Fig. 4*C*). Second, the increased short-term boost associated with unnatural deaths was even more pronounced in the news than on Twitter (Table 2), across age groups (Fig. 4*C*). Third,

leaders were boosted more by the news than by Twitter, both short- and long-term (Table 2). Taken together, these findings could be interpreted as the result of two simultaneous roles played by mainstream news media: On the one hand, as heralds catering to the public curiosity stirred by a young or unnatural death; on the other hand, as stewards of collective memory when an old person or an accomplished leader dies after a life of achievement. On the contrary, the extent to which Twitter plays both roles is weaker: On the one hand, unnatural deaths were followed by a less pronounced short-term boost on Twitter than in the news; on the other hand, Twitter users paid less attention when an old public figure or a leader died.

The present study showed that even the simple counting of mentions yields nuanced insights into who is remembered after death. Future studies may go further by also asking how deceased public figures are remembered, by studying how the language, tone, and attitude toward them change in the wake of death. By considering a diverse set of thousands of public figures such as ours, future work will be able to quantify, e.g., to what extent news and social media abide by the old Latin adage "*De mortuis nihil nisi bonum*" ("Of the dead, speak no evil"). The analysis could be further enriched by going beyond the coarse biographic categories considered here and leveraging manually curated repositories of more fine-grained information about public figures (60). We also emphasize that media attention cannot capture all aspects of collective memory, so we encourage researchers to apply our methodology to further measures of popularity, in particular those capturing the consumption, rather than production, of content, including songs, movies, books, Wikipedia articles, etc.

Finally, this study started from an elite of people considered worthy of being included in the Freebase knowledge base (which roughly equals the set of people with a Wikipedia article). This notability bias was further increased by discarding people whose premortem mention frequency was too low in the news or on Twitter (*Materials and Methods*), a restriction necessary in order to compare the coverage of a fixed person across the two media. Since the bar for being mentioned in the news (61)

Table 2. Linear regression models of news-minus-Twitter difference in short- and long-term boosts

	Short-term boost	Long-term boost
(Intercept)	−0.427 (0.047)***	−0.015 (0.014)
Premortem mean (relative-rank diff.)	−0.212 (0.083)*	−0.034 (0.025)
Manner of death: unnatural	0.348 (0.070)***	−0.008 (0.021)
Language: non-anglophone	−0.219 (0.054)***	−0.023 (0.017)
Language: unknown	−0.052 (0.063)	0.012 (0.019)
Gender: female	0.091 (0.053)	0.029 (0.016)
Notability type: academia/engineering	−0.048 (0.146)	−0.046 (0.044)
Notability type: general fame	−0.016 (0.092)	0.002 (0.028)
Notability type: known for death	0.105 (0.074)	−0.015 (0.022)
Notability type: leadership	0.200 (0.062)**	−0.006 (0.019)
Notability type: sports	0.059 (0.062)	0.009 (0.019)
Age: 20–29	−0.577 (0.126)***	−0.135 (0.038)***
Age: 30–39	−0.235 (0.124)	−0.089 (0.038)*
Age: 40–49	−0.374 (0.093)***	−0.101 (0.028)***
Age: 50–59	−0.204 (0.073)**	−0.029 (0.022)
Age: 60–69	−0.175 (0.061)**	−0.021 (0.018)
Age: 80–89	0.014 (0.058)	−0.006 (0.018)
Age: 90–99	0.164 (0.072)*	0.015 (0.022)
R^2	0.101	0.052
Adj. R^2	0.083	0.034
No. obs.	870	870
RMSE	0.571	0.174

SEs of coefficients are in parentheses. ***$P < 0.001$, **$P < 0.01$, and *$P < 0.05$.

as well as for being included in Freebase and Wikipedia (62) is higher for women than for men, the women included in the study are likely to be more inherently noteworthy than the men included. This might in turn affect our estimate of the association of gender with postmortem memory: Although we identified only small and mostly insignificant effects, it is conceivable that different effects might appear if the inherent noteworthiness was equalized across genders in the dataset by lowering the bar for inclusion for women or raising it for men—an important methodological challenge.

Going forward, researchers should also strive to remove the bar for inclusion in a study of postmortem memory altogether, by moving from a noteworthy elite of public figures to a representative set of regular people. With the widespread adoption of social media, we may, for the first time in history, not only ask, but also answer, who is remembered after they die.

Materials and Methods

News and Twitter Corpora. We compiled a large corpus of media coverage via the online media aggregation service Spinn3r, which provides "social media, weblogs, news, video, and live web content" (48). We had full access to the Spinn3r data stream and collected a complete copy over the course of more than 5 y (June 2009 to September 2014) via the Spinn3r API, for a total of around 40 terabytes of data. Besides the main text content, documents consist of a title, a URL, and a publication date.

Twitter posts (tweets) were easy to recognize automatically in the Spinn3r data, whereas news articles were not explicitly labeled as such. In order to identify news articles, we started from a comprehensive list of all 151K online news articles about Osama bin Laden's killing (May 2, 2011) indexed by Google News (63). Assuming that every relevant news outlet had reported on bin Laden's death, we labeled as news articles all documents in the Spinn3r crawl that were published on one of the 6,608 Web domains that also published an English news article about bin Laden's death according to the Google News list.

We included in our analysis all English-language news articles and tweets collected between June 11, 2009, and September 30, 2014. The resulting corpus comprises, for each day, hundreds of thousands of news articles and tens of millions of tweets (*SI Appendix*, Fig. 1).

Although Spinn3r does not publicly disclose its data collection strategy, we assess the corpus as highly comprehensive (*SI Appendix*, Table 3).

Detecting People Mentions. In order to construct mention time series (Fig. 1), we had to identify documents that contain the names of dead public figures. This is not a trivial task, since names may be ambiguous. Entity disambiguation is a well-studied task, but unfortunately natural language processing–based methods were too resource-intensive for our 40-terabyte corpus, so we resorted to a simpler method: in addition to fully unambiguous names and aliases (henceforth simply "names") belonging to a single entity, we identified a set of mostly unambiguous names, which refer to the same entity at least 90% of the time in English Wikipedia, and we mapped each mention of such a name in the Spinn3r corpus to the entity it most frequently referred to in Wikipedia (people without any highly unambiguous name were excluded). For Twitter, we considered a tweet to mention a given person if the tweet mentioned at least one full name of the person. For the news, we additionally required at least one additional mention of the person (via a full name or a token-based prefix or suffix of a full name), in order to reduce spurious mentions (e.g., when the person was merely mentioned in a link to another article, rather than in the core article content).

Inclusion Criteria. In order to compile a set of dead public figures, we started from the 33,340 people listed in the September 28, 2014, version of the Freebase knowledge base as having died during the period spanned by our media corpus (June 11, 2009, to September 30, 2014). On 86 out of these 1,936 d, Spinn3r failed to provide data due to technical problems. We excluded people for whom the 100 d immediately following death included at least one of the 86 missing days (in order to obtain better estimates of short- and long-term boosts), who died within less than 360 d of the corpus boundary dates (in order to compute premortem means and long-term boosts in the same way for everyone), who were mentioned on fewer than 5 d in either of the news or Twitter during the 360 d before death (in order to avoid extremely noisy premortem means), or whose names on English Wikipedia contained parentheses, e.g., "John Spence (Trinidad politician)" (as such names are unlikely to be used in prose). These criteria reduced the set of people from 33,340 to 2,362.

Biographic Features. Each public figure was described by the following biographic features, extracted or computed from Freebase: age at death, gender, manner of death (natural or unnatural, inferred from the more detailed cause of death; cf. *Taxonomy of causes of death* in *SI Appendix*), language ("anglophone" for citizens of the United States, Canada, the United Kingdom, Ireland, Australia, New Zealand, or South Africa; "non-anglophone" for citizens of other countries; "unknown" for people with no nationality listed in Freebase), and notability type (a profession or role for which the person was most known, e.g., "singer" for Whitney Houston; manually grouped into six classes: arts, sports, leadership, known for death, general fame, academia/engineering; cf. *Taxonomy of notability types* in *SI Appendix*). The distribution of these features is summarized in *SI Appendix*, Table 1 for all public figures, for the people included in the study, and for the subset retained for the regression analysis.

Mention Time Series. To avoid taking logarithms of zero when constructing mention time series, a constant value of ε was added to each individual time series before taking logarithms, where ε was the minimum nonzero value across all individual time series (but note that the raw time series of Fig. 1 were drawn without adding ε). Separate values of ε were computed for the news and for Twitter. When smoothing time series via the variable span smoother (49), we considered the pre- and postmortem periods separately, in order to not smooth out the spike that usually immediately followed death. Missing days were interpolated linearly.

Model Fitting. In order to fit the shifted power law model to the data, we operate on the logarithmic scale, by finding the nonlinear least-squares estimates:

$$\arg\min_{a,b,c} \sum_{t=1}^{400} \left(\langle \log S_i(t) \rangle - \log \left(at^{-b} + c \right) \right)^2, \qquad [1]$$

where $\langle \log S_i(t) \rangle$ is the arithmetic mean of the empirically measured $\log S_i(t)$ over all persons i. The following optimal parameters were obtained:

News: $a = 5.58 \times 10^{-5}$, $b = 1.34$, $c = 1.75 \times 10^{-6}$ **[2]**

Twitter: $a = 1.90 \times 10^{-6}$, $b = 1.54$, $c = 2.35 \times 10^{-8}$ **[3]**

Data and Code Availability. All analysis code, as well as mention frequency data and supplementary data, are publicly available on GitHub (64).

1. C. Carr, Mortuary practices: Their social, philosophical-religious, circumstantial, and physical determinants. *J. Archaeol. Method Theory* **2**, 105–200 (1995).
2. H. I. Flower, *The Art of Forgetting: Disgrace and Oblivion in Roman Political Culture* (University of North Carolina Press, 2006).
3. M. Halbwachs, *Les Cadres Sociaux de la Mémoire* (Librairie Félix Alcan, Paris, 1925).
4. H. L. Roediger III, J. V. Wertsch, Creating a new discipline of memory studies. *Mem. Stud.* **1**, 9–22 (2008).
5. H. L. Roediger III, M. Abel, Collective memory: A new arena of cognitive study. *Trends Cogn. Sci.* **19**, 359–361 (2015).
6. H. L. Roediger III, K. A. DeSoto, Forgetting the presidents. *Science* **346**, 1106–1109 (2014).
7. H. L. Roediger III *et al.*, Competing national memories of World War II. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 16678–16686 (2019).
8. G. R. Edgerton, P. C. Rollins, *Television Histories: Shaping Collective Memory in the Media Age* (University Press of Kentucky, 2001).
9. N. Kligler-Vilenchik, Y. Tsfati, O. Meyers, Setting the collective memory agenda: Examining mainstream media influence on individuals' perceptions of the past. *Mem. Stud.* **7**, 484–499 (2014).
10. M. Neiger, O. Meyers, E. Zandberg, *On Media Memory: Collective Memory in a New Media Age* (Springer, 2011).
11. B. Zelizer, *Covering the Body: The Kennedy Assassination, the Media, and the Shaping of Collective Memory* (University of Chicago Press, 1992).
12. E. Barry, From epitaph to obituary: Death and celebrity in eighteenth-century British culture. *Int. J. Cult. Stud.* **11**, 259–275 (2008).
13. B. Fowler, *The Obituary as Collective Memory* (Routledge, 2007).
14. J. Hume, *Obituaries in American Culture* (University Press of Mississippi, 2000).
15. N. Starck, Death can make a difference: A comparative study of "quality quartet" obituary practice. *Journal. Stud.* **9**, 911–924 (2008).
16. F. Hanusch, Valuing those close to us: A comparison of German and Australian quality newspapers' reporting of death in foreign news. *Journal. Stud.* **9**, 341–356 (2008).
17. Cm. Au Yeung, A. Jatowt, "Studying how the past is remembered: Towards computational history through large scale text mining" in *Proceedings of the 20th ACM International Conference on Information and Knowledge Management* (Association for Computing Machinery, 2011), pp. 1231–1240.
18. T. Huet, J. Biega, F. M. Suchanek, "Mining history with Le Monde" in *Proceedings of the 2013 Workshop on Automated Knowledge Base Construction* (Association for Computing Machinery, 2013), pp. 49–54.
19. J. Cook, A. Das Sarma, A. Fabrikant, A. Tomkins, "Your two weeks of fame and your grandmother's" in *Proceedings of the 21st International Conference on World Wide Web* (Association for Computing Machinery, 2012), pp. 919–928.
20. A. van de Rijt, E. Shor, C. Ward, S. Skiena, Only 15 minutes? The social stratification of fame in printed media. *Am. Sociol. Rev.* **78**, 266–289 (2013).
21. F. M. Suchanek, N. Preda, Semantic culturomics. *Proceedings VLDB Endowment* **7**, 1215–1218 (2014).
22. C. Pentzold, Fixing the floating gap: The online encyclopaedia Wikipedia as a global memory place. *Mem. Stud.* **2**, 255–272 (2009).
23. K. Z. Gach, C. Fiesler, J. R. Brubaker, "Control your emotions, Potter:" An analysis of grief policing on Facebook in response to celebrity death. *Proc. ACM Hum. Comput. Interact.* **1**, 1–18 (2017).
24. L. D. Stone, J. W. Pennebaker, Trauma in real time: Talking and avoiding online conversations about the death of Princess Diana. *Basic Appl. Soc. Psych.* **24**, 173–183 (2002).
25. J. Sanderson, P. Hope Cheong, Tweeting prayers and communicating grief over Michael Jackson online. *Bull. Sci. Technol. Soc.* **30**, 328–340 (2010).
26. D. Hoe-Lian Goh, C. Sian Lee, An analysis of tweets in response to the death of Michael Jackson. *Aslib Proceedings* **63**, 432–444 (2011).
27. S. K. Radford, P. H. Bloch, Grief, commiseration, and consumption following the death of a celebrity. *J. Consum. Cult.* **12**, 137–155 (2012).
28. B. C. Keegan, J. R. Brubaker, "'Is' to 'was': Coordination and commemoration in posthumous activity on Wikipedia biographies" in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing* (Association for Computing Machinery, 2015), pp. 533–546.
29. M. Ferron, P. Massa, "Psychological processes underlying Wikipedia representations of natural and manmade disasters" in *Proceedings of the 8th Annual International Symposium on Wikis and Open Collaboration* (Association for Computing Machinery, 2012), pp. 1–10.
30. M. Ferron, P. Massa, Beyond the encyclopedia: Collective memories in Wikipedia. *Mem. Stud.* **7**, 22–45 (2014).
31. N. Kanhabua, T. N. Nguyen, C. Niederée, "What triggers human remembering of events? A large-scale analysis of catalysts for collective memory in Wikipedia" in *Proceedings of the 2014 IEEE/ACM Joint Conference on Digital Libraries* (Association for Computing Machinery, 2014), pp. 341–350.
32. R. García-Gavilanes, A. Mollgaard, M. Tsvetkova, T. Yasseri, The memory remains: Understanding collective memory in the digital age. *Sci. Adv.* **3**, e1602368 (2017).
33. A. Jatowt, D. Kawai, K. Tanaka, "Digital history meets Wikipedia: Analyzing historical persons in Wikipedia" in *Proceedings of the 2016 IEEE/ACM Joint Conference on Digital Libraries* (Association for Computing Machinery, 2016), pp. 17–26.
34. C. Jara-Figueroa, A. Z. Yu, C. A. Hidalgo, How the medium shapes the message: Printing and the rise of the arts and sciences. *PLoS One* **14**, e0205771 (2019).
35. J. B. Michel *et al.*; Google Books Team, Quantitative analysis of culture using millions of digitized books. *Science* **331**, 176–182 (2011).
36. P. M. Greenfield, The changing psychology of culture from 1800 through 2000. *Psychol. Sci.* **24**, 1722–1731 (2013).
37. E. A. Pechenick, C. M. Danforth, P. S. Dodds, Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. *PLoS One* **10**, e0137041 (2015).
38. A. Jatowt, E. Antoine, Y. Kawai, T. Akiyama, "Mapping temporal horizons: Analysis of collective future and past related attention in Twitter" in *Proceedings of the 24th International Conference on World Wide Web* (Association for Computing Machinery, 2015), pp. 484–494.
39. Y. Sumikawa, A. Jatowt, M. Düring, "Digital history meets microblogging: Analyzing collective memories in Twitter" in *Proceedings of the 18th ACM/IEEE Joint Conference on Digital Libraries* (Association for Computing Machinery, 2018), pp. 213–222.
40. S. Fortunato *et al.*, Science of science. *Science* **359**, eaao0185 (2018).
41. D. Wang, C. Song, A. L. Barabási, Quantifying long-term scientific impact. *Science* **342**, 127–132 (2013).
42. S. P. Fraiberger, R. Sinatra, M. Resch, C. Riedl, A. L. Barabási, Quantifying reputation and success in art. *Science* **362**, 825–829 (2018).
43. R. Sinatra, D. Wang, P. Deville, C. Song, A. L. Barabási, Quantifying the evolution of individual scientific impact. *Science* **354**, aaf5239 (2016).
44. L. Liu *et al.*, Hot streaks in artistic, cultural, and scientific careers. *Nature* **559**, 396–399 (2018).
45. B. Yucesoy, A. L. Barabási, Untangling performance from success. *EPJ Data Sci.* **5**, 1–10 (2016).
46. C. Candia, C. Jara-Figueroa, C. Rodriguez-Sickert, A. L. Barabási, C. A. Hidalgo, The universal decay of collective memory and attention. *Nat. Hum. Behav.* **3**, 82–91 (2019).
47. K. Bollacker, C. Evans, P. Paritosh, T. Sturge, J. Taylor, "Freebase: A collaboratively created graph database for structuring human knowledge" in *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data* (Association for Computing Machinery, 2008), pp. 1247–1250.
48. Spinn3r, Documentation. https://web.archive.org/web/20170105130816/https://www.spinn3r.com/documentation/. Accessed 1 May 2017.
49. J. Friedman, "A variable span smoother" Tech. Rep. No. SLAC-PUB-3477, Stanford Linear Accelerator Center, 1984).
50. J. T. Wixted, E. B. Ebbesen, On the form of forgetting. *Psychol. Sci.* **2**, 409–415 (1991).
51. D. C. Rubin, A. E. Wenzel, One hundred years of forgetting: A quantitative description of retention. *Psychol. Rev.* **103**, 734–760 (1996).
52. P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
53. S. Ronen *et al.*, Links that speak: The global language network and its association with global fame. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E5616–E5622 (2014).
54. T. Martin, J. M. Hofman, A. Sharma, A. Anderson, D. J. Watts, "Exploring limits to prediction in complex social systems" in *Proceedings of the 25th International Conference on World Wide Web* (Association for Computing Machinery, 2016), pp. 683–694.
55. M. J. Salganik, P. S. Dodds, D. J. Watts, Experimental study of inequality and unpredictability in an artificial cultural market. *Science* **311**, 854–856 (2006).
56. N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, Fake news on Twitter during the 2016 U.S. presidential election. *Science* **363**, 374–378 (2019).
57. A. Bovet, H. A. Makse, Influence of fake news in Twitter during the 2016 US presidential election. *Nat. Commun.* **10**, 7 (2019).
58. W. X. Zhao *et al.*, "Comparing Twitter and traditional media using topic models" in *Proceedings of the European Conference on Information Retrieval* (Association for Computing Machinery, 2011), pp. 338–349.
59. H. Kwak, C. Lee, H. Park, S. Moon, "What is Twitter, a social network or a news media?" in *Proceedings of the 19th International Conference on World Wide Web* (Association for Computing Machinery, 2010), pp. 591–600.
60. A. Z. Yu, S. Ronen, K. Hu, T. Lu, C. A. Hidalgo, Pantheon 1.0, a manually verified dataset of globally famous biographies. *Sci. Data* **3**, 150075 (2016).
61. E. Shor, A. van de Rijt, B. Fotouhi, A large-scale test of gender bias in the media. *Sociol. Sci.* **6**, 526–550 (2019).
62. C. Wagner, E. Graells-Garrido, D. Garcia, F. Menczer, Women through the glass ceiling: Gender asymmetries in Wikipedia. *EPJ Data Sci.* **5**, 1–24 (2016).
63. K. Bharat, Google News and the coverage of bin Laden. https://web.archive.org/web/20170105164742/https://news.googleblog.com/2011/05/google-news-and-coverage-of-bin-laden.html. Accessed 1 May 2017.
64. R. West, J. Leskovec, C. Potts, Code and data repository for "Postmortem memory of public figures in news and social media." GitHub. https://github.com/epfl-dlab/postmortem-memory. Deposited 28 May 2021.

COMPUTER SCIENCES

SOCIAL SCIENCES