Trends and fluctuations in the severity of interstate wars

Aaron Clauset[1, 2, 3, *]

[1]*Department of Computer Science, University of Colorado, Boulder, CO 80309, USA*
[2]*BioFrontiers Institute, University of Colorado, Boulder, CO 80303, USA*
[3]*Santa Fe Institute, Santa Fe, NM 87501, USA*

Since 1945, there have been relatively few large interstate wars, especially compared to the preceding 30 years, which included both World Wars. This pattern, sometimes called the long peace, is highly controversial. Does it represent an enduring trend, caused by a genuine change in the underlying conflict generating processes? Or, is it consistent with a highly variable but otherwise stable system of conflict? Using the empirical distributions of interstate wars sizes and onset times from 1823–2003, we parameterize stationary models of conflict generation that can distinguish trends from statistical fluctuations in the statistics of war. These models indicate that both the long peace, and the period of great violence that preceded it, are not statistically uncommon patterns in realistic but stationary conflict time series. This fact does not detract from the importance of the long peace, or the proposed mechanisms that explain it. However, the models indicate that the post-war pattern of peace would need to endure at least another 100–140 years to become a statistically significant trend. This fact places an implicit upper bound on the magnitude of any change in the true likelihood of a large war after the end of the Second World War. The historical patterns of war thus seem to imply that the long peace may be substantially more fragile than proponents believe, despite recent efforts to identify mechanisms that reduce the likelihood of interstate wars.

Over the next century, should we expect the occurrence of another international conflict that kills tens of millions of people, as the Second World War did in the last century? How likely is such an event to occur, and has that likelihood decreased in the years since the Second World War? What are the underlying social and political factors that govern this probability over time?

These questions reflect a central mystery in international conflict [2–5] and in the arc of modern civilization: are there trends in the frequency and severity of wars between nations, or, more controversially, is there is a trend specifically toward peace? If such a trend exists, what factors are driving it? If no such trend exists, what kind of processes govern the likelihood of such wars and how can they be stable despite changes in so many other aspects of the modern world? Scientific progress on these questions would help quantify the true odds of a large interstate war over the next one hundred years, and shed new light on whether the great efforts of the 20th century to prevent another major war have been successful, and whether the lack of such a conflict can be interpreted as evidence of a change in the true risk of war.

Early debates on trends in violent conflict tended to focus on the causes of wars, particularly those between nation states [6]. Some researchers argued that the risk of such wars is constant and fundamentally inescapable, while others argued that warfare is dynamic and its frequency, severity, and other characteristics depend on malleable social and political processes [2, 3, 7–10]. More recent debates have focused on whether there has been a real trend toward peace, particularly in the post-war period that began after the Second World War [4, 11–14].

*Electronic address: `aaron.clauset@colorado.edu`

In this latter debate, the opposing claims are associated with liberalism and realism perspectives in international relations [15]. The liberalism argument draws on multiple lines of empirical evidence, some spanning hundreds or even thousands of years (for example, Refs. [4, 16]), to identify a broad and general decline in human violence, and a specific decline in the likelihood of war, especially between the so-called "great powers." Arguments supporting this perspective often focus on mechanisms that reduce the risk of war, such as the spread of democracy [17, 18], peace-time alliances [10, 19–21], economic ties, and international organizations [10, 22].

The realism argument, in contrast, draws on empirical evidence, some of which also spans great lengths of time (for example, Ref. [23, 24]), to identify an absence of discernible trends toward peace within observed conflict time series. A key claim from this perspective is that the underlying conflict generating processes in the modern world are stationary, an idea advanced in the early 20th century by the English polymath Lewis Fry Richardson [25, 26], in his seminal work on the statistics of war sizes and frequencies.

This debate has been difficult to resolve because the evidence is not overwhelming, war is an inherently rare event, and there are the multiple ways to formalize the notion of a trend [5, 6, 18, 27, 28]. Should we focus exclusively on international conflicts, or should other types of conflict be included, such as civil wars, peacekeeping missions, or even the simple use of military force? Should we focus on conflicts between major powers, or between geographically close nations, or should all nations be included? What conflict variables should we consider? How should we account for the changing number of states, which have increased by nearly an order of magnitude over the past 200 years? What about other non-stationary characteristics of conflicts, such as the in-

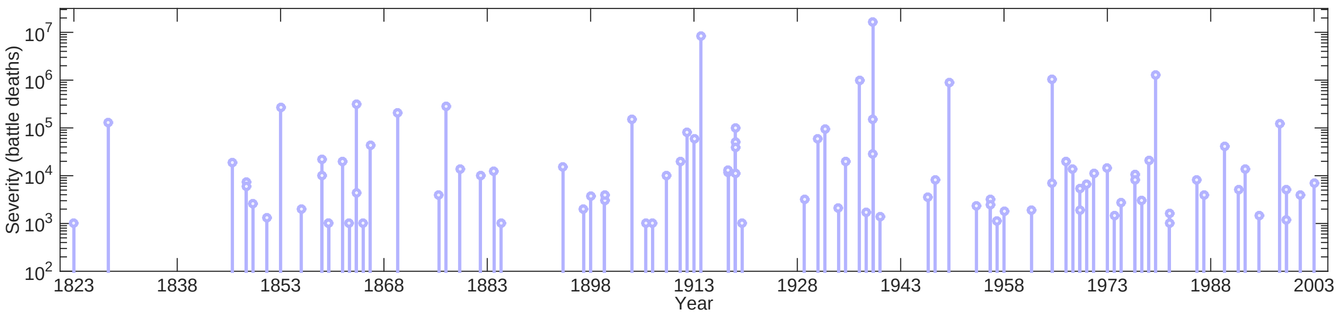

FIG. 1: The Correlates of War interstate war data [1] as a conflict time series, showing both severity (battle deaths) and onset year for the 95 conflicts in the period 1823–2003.

creasing frequency of asymmetric or unconventional conflicts, the use of insurgency, terrorism, and radicalization, improvements in the technology of war and communication, or changes in economic development? Different answers to these questions can lead to opposite conclusions about the existence or direction of a trend in conflict [2, 3, 8, 9, 23, 24, 27–29].

Ultimately, the question of identifying trends in war is inherently statistical. Answering it depends on distinguishing a lasting change in the dynamics of some conflict variable from a temporary pattern—a fluctuation—in an otherwise stable conflict generating process. More formally, a trend exists if there is a measurable shift in the parameters of the underlying process that produces wars, relative to a model with constant parameters. Identifying such trends is a kind of change-point detection task [30], in which one tests whether the distributions observed before and after a change point are statistically different. The ease of making such distinctions depends strongly on the natural variance of the observed data.

This article presents a data-driven analysis of the general evidence for trends in the sizes of and years between interstate wars worldwide, and uses the resulting models to characterize the plausibility of a trend toward peace since the end of the Second World War. This analysis focuses on the 1823–2003 period and on the interstate wars in the Correlates of War interstate conflict dataset [1, 31] (Fig. 1), which provides comprehensive coverage in this period, with few artifacts and relatively low measurement bias. The underlying variability in these data is captured using an ensemble approach, which then specifies a stationary process by which to distinguish trends from flucutations in the timing of war onsets, the severity of wars, and the joint distribution of onsets and severity.

The so-called long peace [11]—the widely recognized pattern of few or no large wars since the end of the Second World War—is an important and widely claimed example of a trend in war. However, the analysis here demonstrates that periods like the long peace are, in fact, a statistically common occurrence under the stationary model, and even periods of profound violence, like that of and between the two World Wars, are within expectations for statistical fluctuations. Hence, even if there have been genuine changes in the processes that generate wars over the past 200 years, data on the frequency and severity of wars alone is insufficient to detect those shifts. In fact, the long peace pattern would need to endure for at least another 100–150 years before it could plausibly be called a genuine trend. These findings place an upper bound on the magnitude of any underlying changes in the conflict generating processes for wars, if they are to be consistent with the observed statistics. These results imply that the current peace may be substantially more fragile than proponents believe.

I. WAR SIZES AND WAR ONSETS

Trends in war are implicitly statements about changing likelihoods of rare events, and their rarity necessarily induces uncertainty in any statistical estimate. To control for this uncertainty in the analysis here, the entire distribution of a conflict variable is modeled and ensembles of these models are used to quantify the uncertainty in the distribution's shape. For concreteness, the analysis focuses on the sizes of wars and their timing in the historical record. Initially, these variables are considered independently, and subsequently a joint model is formulated to numerically estimate the likelihood of historical patterns, like the long peace, relative to well-defined stationary models of interstate conflict.

The sizes or severities of wars, commonly measured in battle deaths, have been known since the mid-20th century to follow a right-skewed distribution with a heavy tail, in which the largest wars are many orders of magnitude larger than a "typical" war. In Richardson's original analysis of interstate wars 1820–1945 [32], he argued that war sizes followed a precise pattern, called a power-law distribution, in which the probability that a war kills x people is $\Pr(x) \propto x^{-\alpha}$, where $\alpha > 1$ is called the "scaling" parameter and $x \geq x_{\min} > 0$. He also argued that the timing of wars followed a simple Poisson process, implying a constant annual hazard rate and an exponential distribution for the time between wars [25, 26]. Although Richardson's analysis would not be considered statistically rigorous today, these patterns—a power-law distribution for war sizes and a Poisson process for their onsets—represent a simple stationary model for the statistics of interstate wars worldwide.

Here, this model is improved upon by first testing the statistical plausibility of its two key assumptions, then estimating their structure from empirical data, and finally combining them computationally to investigate the likelihood that the end of the Second World War and its subsequent long peace pattern, represents a change point in the observed statistics of interstate wars. Before examining the marginal distributions of war sizes and timing between war onsets, a brief overview is given of relevant mathematical and statistical issues for specifying these models and using them to distinguish trends from fluctuations in conflict time series.

A. Statistical Concerns

Power-law distributions have unusual mathematical properties [33, 34], which can require specialized statistical tools. (For a primer on power-law distributions in conflict, see Refs. [7, 35].) For instance, when observations are generated by a power law, time series of summary statistics like the mean or variance can exhibit large fluctuations that can resemble a trend. The largest fluctuations occur for $\alpha \in (1, 3)$, when one or both the mean and variance are mathematically undefined. In the context of interstate wars, this property can produce long transient patterns of low-severity or the absence of wars, making it difficult to distinguish a genuine trend toward peace from a transient fluctuation in a stationary process.

Identifying a power law in the distribution of an empirical quantity can indicate the presence of exotic underlying mechanisms, including nonlinearities, feedback loops, and network effects [33, 34], although not always [36], and power laws are believed to occur broadly in complex social, technological, and biological systems [37]. For instance, the intensities or sizes of many natural disasters, such as earthquakes, forest fires and floods [34, 38, 39], as well as many social disasters, like riots and terrorist attacks [35, 40], are well-described by power laws.

However, it can be difficult to accurately characterize the shape of a distribution that follows a power-law pattern [37]. Fluctuations in heavy-tailed data are greatest in the distribution's upper tail, which governs the frequency of the largest and rarest events. As a result, data tend to be sparsest precisely where the greatest precision in model estimates is desired.

Recent interest in heavy-tailed distributions has led to the development of more rigorous methods for identifying and estimating power-law distributions in empirical data [37, 41, 42], for comparing different models of the upper tail's shape [37], and for making principled statistical forecasts of future events [43]. This branch of statistical methodology is related to but distinct from the task of estimating the distribution of maxima within a sample [44, 45], and is more closely related to the peaks-over-threshold literature in seismology, forestry, hydrology, insurance, and finance [41, 42, 45–48].

Although Poisson processes pose fewer statistical concerns than power-law distributions, a similar statistical approach is used in the analysis here of both war sizes and years between war onsets. In particular, an ensemble approach is used [43], based on a standard nonparametric bootstrap procedure [49] that simulates the generative process of events to produce a series of synthetic data sets $\{Y\}$ with similar statistical structure as the empirical data X. Fitting a semi-parametric model $\Pr(y \mid \theta)$ to each Y yields an ensemble of models $\{\theta\}$ that incorporates the empirical data's inherent variability into a distribution of estimated parameters. This distribution is then used to weight models by their likelihood under the bootstrap distribution and to numerically estimate the likelihood of specific historical or future patterns [43].

Within the 1823–2003 time period, the end of the Second World War in 1945 is widely viewed as the most plausible change point in the underlying dynamics of the conflict generating process for wars and marks the beginning of the subsequent long peace pattern [11]. Determining whether 1945 marks a genuine a shift in the observed statistics of wars, and hence whether the long peace is plausibly a trend or a fluctuation, represents a broad test of the stationary hypothesis of war [24]. Evaluating other theoretically plausible change points in these data is left for future work.

Finally, some studies choose to limit or normalize war onset counts or war sizes (battle death counts) by a reference population. For instance, onset counts can be normalized by assuming that war is a dyadic event and that dyads independently generate conflicts [27], implying a normalization that grows quadratically with the number of nations. However, considerable evidence indicates that dyads do not independently generate conflicts [10, 17–22]. Similarly, limiting the analysis to conflicts among "major powers" introduces subjectivity in defining such a scope, and there is not a clear consensus about the details, e.g., when and whether to include China or the occupied European nations, or certain wars, like the Korean War [27]. War size can be normalized by assuming that individuals contribute independently to total violence, which implies a normalization that depends on either the population of the combatant nations (a variable sometimes called war "intensity") or of the world [4, 24]. However, there is little evidence for this assumption [4, 50], although such a per-capita variable may be useful for other reasons. In the analysis performed here, war variables are analyzed in their unnormalized forms and all recorded interstate wars are considered. The analysis is thus at the level of the entire world, and results are about absolute counts.

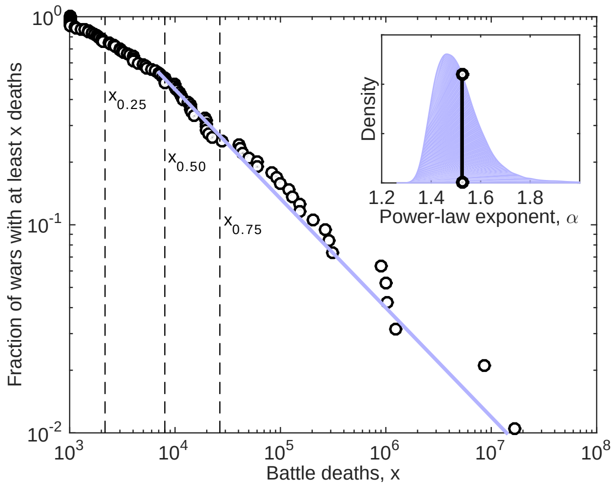

FIG. 2: **Interstate wars sizes, 1823–2003.** The maximum likelihood power-law model of the largest-severity wars (solid line, $\hat{\alpha} = 1.53 \pm 0.07$ for $x \geq \hat{x}_{\min} = 7061$) is a plausible data-generating process of the empirical severities (Monte Carlo, $p_{\mathrm{KS}} = 0.78 \pm 0.03$). For reference, distribution quartiles are marked by vertical dashed lines. Inset: bootstrap distribution of maximum likelihood parameters $\Pr(\hat{\alpha})$, with the empirical value (black line).

B. The Sizes of Wars

Considering the sizes of wars alone necessarily ignores other characteristics of conflicts, including their relative timing, which may contain independent signals about trends. A pattern in war sizes alone thus says little about changes in declared reasons for conflicts, the way they are fought, their settlements, aftermaths, or relationships to other conflicts past or future or the number of nations worldwide, among other factors. One benefit of ignoring such factors, at least at first, is that they may be irrelevant for identifying an overall trend in wars, and their relationship to a trend can be explored subsequently. Hence, focusing narrowly on war sizes simplifies the range of models to consider and may improve the ability to detect a subtle trend.

The Correlates of War dataset includes 95 interstate wars, the absolute sizes of which range from 1000, the minimum size by definition, to 16,634,907, the recorded battle deaths of the Second World War (Fig. 2). The estimated power-law model has two parameters, $x_{\min}$, which represents the smallest value above which the power-law pattern holds, and α, the scaling parameter. Standard techniques are used to estimate model parameters and model plausibility [37] (Appendix A).

The maximum likelihood power-law parameter is $\hat{\alpha} = 1.53 \pm 0.07$, for wars with severity $x \geq x_{\min} = 7061$ (Fig. 2, inset), and 95% of the bootstrap distribution of $\hat{\alpha}$ falls within the interval $[1.37, 1.76]$. These estimates, however, do not indicate that the observed data are a plausible iid draw from the fitted model. To quantitatively assess this aspect of the model, an appropriately-defined statistical hypothesis test was used [37], which indicates that a power-law distribution cannot be rejected as a data generating process ($p_{\mathrm{KS}} = 0.78 \pm 0.03$). That is, the observed data are as a group statistically indistinguishable from an iid draw from the fitted power-law model. This finding is consistent with past analyses of war intensities (population-normalized war size) over a similar period of time, which found a power law to be statistically plausible and at least as good a model as alternative heavy-tailed distributions, including a power-law distribution with an upper, exponential cutoff [37].

The statistical plausibility of a power-law distribution here provides only circumstantial evidence for the stationary hypothesis, as this analysis does not consider the exchangeability of the sequence of wars, or whether the Second World War is a plausible change point in the observed statistics of wars. The size of the Second World War is not statistically anomalous given the historical distribution of war sizes (Appendix B). Furthermore, the distribution of war sizes in the post-war period, defined as onsets that occurred during 1940–2003, is not statistically distinguishable from the distribution in the preceding period, 1823–1939 (Appendix C).

The relatively small sample size of the dataset necessarily reduces the statistical power of such tests, and thus a trend may still exist within these events and be obscured by the large fluctuations that power laws naturally produce. Additionally, the above distributional analysis only models the relative frequencies of the 51 largest conflicts ($x \geq 7061$, the upper 54% of the distribution), and nearly half of all interstate wars fall outside this domain. Small wars could thus follow a different pattern or trend than large wars. In the subsequent model of the joint distribution of war sizes and timing, a semi-parametric approach is used to capture the full distribution of war sizes (Appendix A) and to investigate the statistical power of the stationary hypothesis over time.

C. The Time Between Wars

A trend toward peace could also manifest as a trend in the time between new wars. That is, the sizes of wars may not be changing, but the time between consecutive wars overall or those of at least a certain size may be lengthening.

Delays between consecutive wars in the dataset range from 0 years, representing wars that began in the same year, to 18 years, the delay between the First Russo-Turkish war in 1828 and the Mexican-American war in 1846. However, long delays are uncommon, and in the post-war period no delay exceeded the 7 years between the onsets of the Franco-Thai war in 1940 and the First Kashmir war in 1947. Overall, wars have occurred at a relatively steady pace since 1823, with an average time of 1.91 years between consecutive war onsets. In only 14 of the 181 years (8%) were there multiple new war onsets

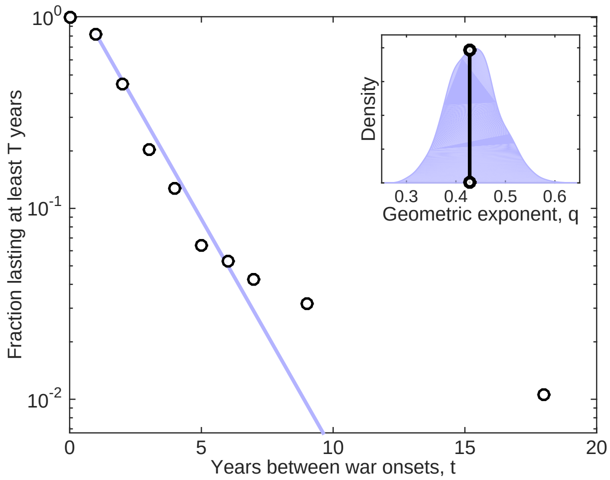

FIG. 3: **Times between interstate war onsets, 1823–2003.** The maximum likelihood geometric model (solid line, $\hat{q} = 0.428 \pm 0.002$ for $t \geq 1$) is a plausible data-generating process of the empirical delays (Monte Carlo, $p_{\mathrm{KS}} = 0.13 \pm 0.01$), implying that the apparent discontinuity at $t = 5$ is a statistical artifact. Inset: the bootstrap distribution of maximum likelihood parameters $\Pr(\hat{q})$, with the empirical estimate (black line).

in the same year, and all other years had either 0 or 1. Most wars ended no more than 2 years after their onset (79%), and hence the temporal analysis here focuses on the distribution of times between war onsets.

If the generation of war onsets follows a Poisson process, the distribution of years between onsets is given by a geometric distribution, since onsets are binned by year. For interstate conflicts, the maximum likelihood parameter for a geometric model of time between wars is $\hat{q} = 0.428 \pm 0.002$ for delays of at least $t \geq t_{\min} = 1$ year (Fig. 3, inset), and 95% of the bootstrap distribution of $\hat{q}$ falls within the interval $[0.354, 0.520]$.

To assess whether the observed data are plausibly an iid draw from this model, an appropriately-defined statistical hypothesis test [37] was used, with a geometric distribution as the null rather than a power law. This test indicates that a geometric model cannot be rejected as a data generating process ($p_{\mathrm{KS}} = 0.13 \pm 0.01$), implying that the observed delays are statistically indistinguishable from an iid draw from the fitted geometric model. Hence, although visually there appears to be a discontinuity in the empirical distribution around $t = 5$ years, this pattern is not statistically significant.

As with war sizes, the statistical plausibility of a simple geometric model for years between war onsets provides circumstantial evidence for a stationary process, as it does not test for exchangeability. However, the distribution of time between wars in the post-war period, 1940–2003 is also not statistically distinguishable from that of the preceding period, 1823–1939 (Appendix D). This finding supports the basic realist argument for a lack of any trend in the timing of wars, and it agrees with Richardson's original hypothesis that war onsets are well described by a simple a Poisson process [25, 26].

II. A STATIONARY MODEL OF CONFLICT

The liberalism thesis argues that after the Second World War, large wars became relatively less common than they were before it. In other words, the joint distribution of war sizes and their timing changed at a particular time. This idea is evaluated quantitatively here by combining the results on war size and timing to develop stationary models that can generate synthetic histories, or futures, of war onset times and corresponding sizes. Under such a model, a genuine trend like the long peace should be statistically unlikely. If it is not, then the stationary process's dynamics may be extrapolated into the future in order to estimate how long a long peace pattern must endure before it becomes statistically distinguishable from a long but transient excursion under stationarity. In this way, the statistical power of these tests may be investigated.

The long peace pattern is commonly described as a change in the frequency of large wars, and here a "large" war is defined as one with severity in the upper quartile of the historical distribution, i.e., $x \geq x_{0.75} = 26,625$ battle deaths. Over the initial 1823–1939 period, there were 19 large wars, with a new large war occurring on average every 6.2 years. The "great violence" pattern of 1914–1939, which spans the onsets of the First and Second World Wars, was especially violent, with 10 large wars, or about one every 2.7 years. In contrast, the long peace of the 1940–2003 post-war period contains only 5 large wars, or about one every 12.8 years, a dramatic reduction in the most severe conflicts relative to earlier periods. These patterns are represented quantitatively using an accumulation curve, which gives the cumulative count over time of wars whose size exceeds some threshold (Fig. 4).

A. Distinguishing trends from fluctuations

The likelihood of the post-war pattern can now be evaluated under three stationary models of war size and timing. Each model first chooses a sequence of onset years and then independently assigns a war size to each. Hence, these models differ only in their variability in generating the timings or sizes of wars, and each set of simulated wars is an exchangeable sequence of random variables.

For war onsets, Models 1 and 2 use the empirical onset dates as observed, producing 95 wars in each simulation. Model 3 generates a new war in each year according to a Bernoulli process with the empirically observed production rate (on average, a new war every 1.91 years) [65], in agreement with the fitted geometric distribution of delays. For war sizes, Model 1 draws a size value iid from

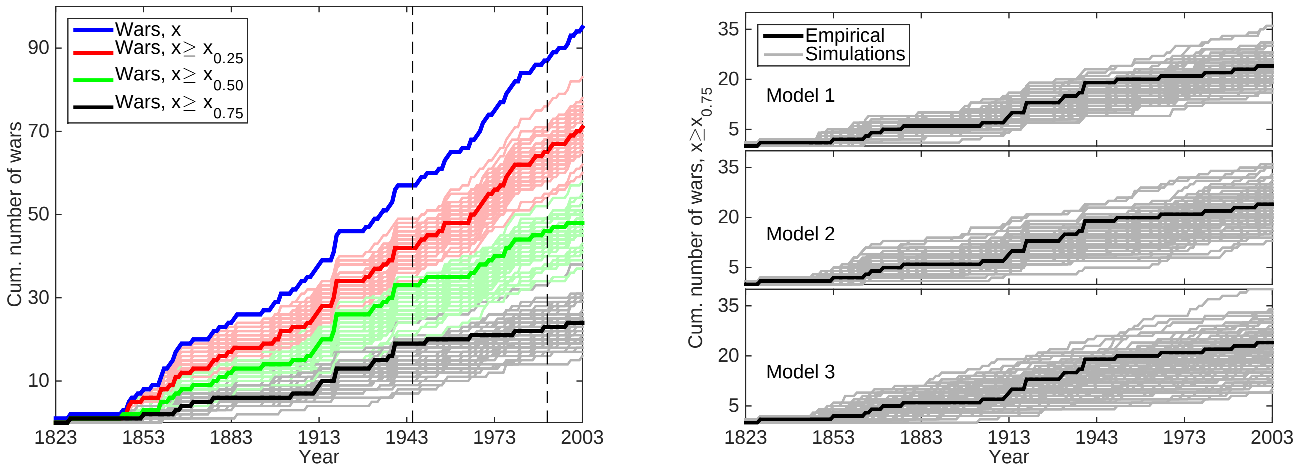

FIG. 4: **Cumulative counts of wars by general severity.** (**a**) Empirical counts of wars of different sizes (dark lines) over time against ensembles of simulated counts from a stationary model, in which empirical severities are replaced iid with a bootstrap draw from the empirical severity distribution (Model 1). For reference, dashed lines mark the end of the Second World War and the end of the Cold War. (**b**) For the largest-severity wars alone, empirical and simulated counts for three models of stationarity, which incorporate progressively more variability in the underlying data generating process (see text).

Empirical pattern	Formalization	Model 1	Model 2	Model 3
great violence	$\Pr(V \equiv n \geq 10$ large wars over $t \leq 27$ years)	0.107(1)	0.159(1)	0.121(1)
long peace	$\Pr(P \equiv n \leq 5$ large wars over $t \geq 64$ years)	0.622(2)	0.569(2)	0.681(2)
violence, then peace	$\Pr(V$ followed by $P)$	0.0030(2)	0.0029(2)	0.0055(2)

TABLE I: **Stationary likelihood of empirical conflict patterns.** Under three models of stationary conflict generation (see text), estimated likelihoods of observing one of three large-war patterns over the period 1823-2003: a "great violence," meaning 10 or more large war onsets ($x \geq x_{0.75}$) over a 27 year period (the empirical count of such onsets, 1914–1939); a "long peace," meaning 5 or fewer large war onsets over a 64 year period (the empirical count of such onsets, 1940–2003); or, a great violence followed by a long peace. Probabilities estimated by Monte Carlo. Parenthetical values indicate the standard error of the least significant digit.

the empirical distribution (a bootstrap), while Models 2 and 3 draw a size iid from a uniformly random member of the ensemble of semi-parametric power-law models obtained above. Intuitively, Models 1, 2, and 3 represent a sequence of increasing variance in the posterior distribution of war sizes and delays, with Model 1 producing the smallest variance and Model 3 producing the largest.

Within the empirical accumulation curve for large wars, the long peace is a visible pattern, in which the rate of production (slope of the accumulation curve) is substantially more flat than in the preceding great violence period (Fig. 4**a**; Appendix E). However, under all three models, the long peace pattern falls comfortably within the distribution of simulated curves (Fig. 4**b**), implying that the observed pattern is not statistically distinguishable from a long transient of the heavy-tailed distribution of historical war sizes.

In fact, most war sequences (57–68%) produced by the three stationary models contain a period of peace at least as long in years and at least a peaceful in terms of large wars as the long peace (Table I). These results show that long periods of relatively few large wars are downright common even when the hazard rate of a large war is constant and unchanging. Hence, observing a long period of peace is not necessarily evidence of a changing likelihood for large wars.

In contrast, few war sequences (11-16%) contain a period of violence at least as large and over no more time that the great violence, implying that this period was relatively unusual, although still not statistically rare, in the degree of clustering in time for large war onsets. Moreover, the joint probability of a period of great violence immediately followed by a long period of peace is genuinely rare, occurring in less than 1% of simulated sequences across models. This estimate is consistent with the liberalism hypothesis that some learning or adaptation resulting from the World Wars [15] may have changed the subsequent conflict generating process for subsequent wars. However, the magnitude of this statistic should be interpreted cautiously, as all sufficiently specific sequences of events under a stationary process have small likelihoods.

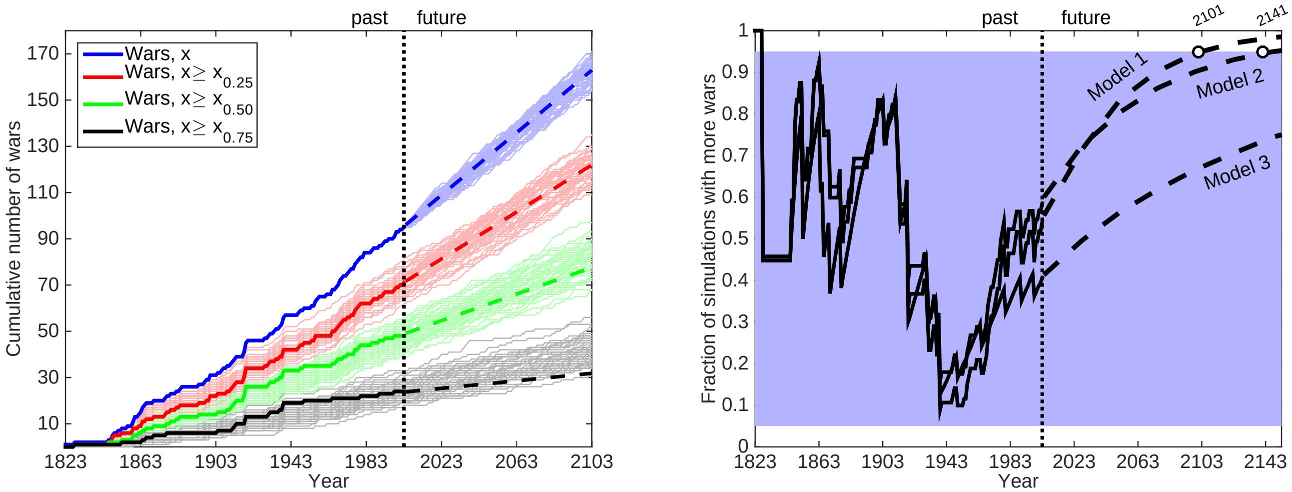

FIG. 5: **How long must the peace last?** (**a**) Simulated accumulation curves for wars of different sizes under a simple stationary model (Model 1; see text), overlaid by the empirical curves up to 2003 (dark lines) and linear extrapolations of the empirical post-war trends (the long peace) for the next 100 years (dashed lines). Quartile thresholds are derived from empirical severity data. (**b**) The fraction of simulated conflict time series that contain more large wars ($x \geq x_{0.75}$) than observed in the past, or than expected in the future relative to a linear extrapolation of the post-war tend. Years at which the post-war trend (the long peace) becomes statistically unlikely under a stationary model, relative to 95% of simulated time series, are marked with open circles.

B. Peering into the future

If the post-war pattern of relatively fewer large wars were permanent, at what future date could we reasonably conclude that this pattern is a trend, i.e., a genuine change in the statistics of large wars, and not a fluctuation? This question can be answered by extrapolating the simulated war sequences into the future. The variable of interest is then the fraction of simulations with a greater accumulation of large wars than either observed in the past or expected in the future, under a linear extrapolation of the long peace pattern, in which a new large war occurs every 12.8 years on average. This fraction's trajectory describes the evolution of the statistical likelihood of the empirical accumulation pattern of large wars over time.

To generate war sequences that cover both the historical 1823–2003 period and a future period of chosen length, Models 1 and 2 are modified to generate a new war in each year beyond 2003 according to the same Bernoulli process of Model 3. Otherwise, all features of all models remain the same. The result is a distribution of accumulation curves of any length required, which quantify the natural variability of the accumulation of large wars under a stationary process (Fig. 5**a**). And, as before, Model 1 produces the smallest variance in the posterior distribution and Model 3 produces the largest.

Over the historical period, the observed accumulation of large wars fluctuates throughout the middle 90% of simulated trajectories, confirming the above results that the historical record of large wars is not itself statistically unusual under a stationary process (Fig. 5**b**). The great violence pattern of 1914–1939, however, resulted in a dramatic shift in the significance of the observed number of large wars, moving from a point where most stationary sequences had more large wars than were observed empirically to a point where most had fewer. The extent of this shift, however, was not large enough to make the overall accumulation pattern statistically significant relative to the stationary hypothesis. The subsequent long peace pattern then moved the relative significance back to the middle of the null distribution.

In other words, the sparsity of large wars in the post-war 1940–2003 period thus served to counterbalance the large density of such wars in the preceding 1914–1939 period. Hence, in a purely statistical accounting sense, the long peace has simply balanced the books relative to the great violence. Had the great violence contained fewer large wars or were the long peace substantially longer, the recent empirical pattern of relatively few large wars would appear marginally more significant.

In the extrapolated future, the post-war pattern of relatively few large wars becomes progressively more unlikely under a stationary hypothesis (Fig. 5**b**). The particular year at which the long peace pattern becomes significant differs by stationary model, with crossing points around 100–140 years in the future for Models 1 and 2. Model 3 yields a much longer estimate because its stochastic process for war onsets leads to substantially larger variance in the accumulation statistics over time. In general, however, the long peace would need to hold for at least another century to be statistically distinguishable from a large but random fluctuation in an otherwise stationary process for war sizes and onsets. The consistency of the historical record of wars with a stationary model places an implicit upper bound on the magnitude of change in the underlying conflict generating process since the end of the Second World War. Such a change in the production of wars cannot be ruled out by this statistical analysis, but if it exists, it is evidently not a dramatic shift. That is, these results can be consistent with evidence of genuine changes in the international system, but they constrain the extent to which such changes could have impacted the production of interstate wars.

III. THE LONG VIEW

If the long peace does not reflect a fundamental shift in the production of large wars [24], then in the years between now and when such a pattern becomes statistically significant, the hazard of a very large war would in fact remain constant. In this case, a stationary model may be used to estimate the likelihood of a very large war occurring over 100 years, one like the Second World War, which produced $x_* = 16,634,907$ battle deaths. Using the ensemble of semi-parametric models for the sizes of wars and assuming a new war onset every 1.91 years on average [43], the probability of observing at least one war with x_* or more deaths is $p_* = 0.43 \pm 0.01$ (Monte Carlo) and the expected number of such events over the next 100 years is 0.62 ± 0.01. Hence, under stationarity, the likelihood of a very large war over the next 100 years is not particularly small.

Under an even stronger assumption of stationarity, the model can estimate the waiting time for a war of truly spectacular size, such as one with $x = 1,000,000,000$ (one billion) battle deaths. A conflict this large would be globally catastrophic and would likely mark the end of modern civilization. It is also not outside the realm of possibility, if current nuclear weapons were deployed widely.

Using the ensemble of semi-parametric models of war sizes and a longer Monte Carlo simulation, the model estimates that the median forecasted waiting time for such an event is 1339 years. Reflecting the large fluctuations that are natural under the empirical war size distribution, the distribution of waiting times for such a catastrophic event is enormously variable, with the 5–95% quantiles ranging from 383 years to 11,489 years. A median delay of roughly 1300 years does not seem like a long time to wait for an event this enormous in magnitude, and humans have been waging war on each other, in one way or another, for substantially longer than that.

The plausibility of this prediction is likely unknowable. However, a genuinely stationary process would hold equally well for the past as for the future, and there is no evidence of such an event in the long history of human conflict [50]. Its absence suggests that some aspects of conflict generation are probably not stationary in the way they have been modeled here, and hypotheses are easy to enumerate. For example, changes in world population, technology, and political structures have likely all played some role in increasing or decreasing the sizes of wars over very long periods of time, but none of these processes are represented directly in observed war sizes [51]. Looking toward the future, however, stationarity may be more plausible [52], and hence the prospect of a civilization-ending conflict in the next 13 centuries is sobering.

IV. DISCUSSION

The absence of a large war between major powers and relatively few large wars of any kind since the end of the Second World War is an undeniable international achievement. Whether this peaceful pattern should be expected to endure, however, has been a central mystery in conflict research now for several decades. On the one hand, a substantial body of scholarship now presents a compelling argument that the post-war peace reflects a genuine trend, based on mechanisms that reduce the likelihood of war [10, 17, 21] and on statistical signatures of a broad and centuries-long decline in general violence [4, 8, 13, 14] or the improvement of other aspects of human welfare [16]. The result is not unreasonable support for the optimistic perspective espoused by liberalism, in spite of reality's frequent disregard for the direction of the trend.

However, focusing only on mechanisms that explain a trend toward peace commits a kind of scientific fallacy by selecting on the dependent variable. A full accounting of the likelihood that the long peace will endure must consider not only the mechanisms that reduce the likelihood of war but also the mechanisms that increase it (for example, Refs. [53–55]). War-promoting mechanisms certainly include the reverse of established peace-promoting mechanisms, e.g., the unraveling of alliances, the slide of democracies into autocracy, or the fraying of economic ties, but may also include completely novel mechanisms.

In the long run, some of the processes that promote interstate war may be intimately related to the ones that reduce it over the shorter term, through feedback loops, tradeoffs, or backlash effects. For example, the persistent appeal of nationalism, the spread of which can increase the risk of interstate wars [56], is not independent of deepening economic ties via globalization [57]. The investigation of such interactions will be a vital direction of future work in conflict research. But without a complete understanding of mechanisms that promote interstate war, especially large ones, it is unclear whether the post-war pattern of peace will continue, or whether the formation and eventual dissolution of periods of peace are part and parcel for a dynamical but ultimately stationary system.

Three key difficulties for evaluating the changing likelihood of war from a mechanistic viewpoint are (i) the incomplete accounting of mechanisms, i.e., it seems unlikely that every process has been identified, either peace-promoting or war-promoting, (ii) the lack of understanding of "meta-mechanisms," which govern the emergence and relevance of particular concrete mechanisms over very long periods of time, e.g., due to evolving political structures, and (iii) the difficulty of integrating these varied mechanisms into a single calculation. In contrast, a statistical analysis of the historical record, like the one performed here, does not require accounting for all possible mechanisms. Hence, it provides both an alternative approach to arrive at some kind of answer to a difficult question, and a quantitative means by which to identify unusual patterns in need of explanation.

The analysis here finds that the post-war pattern of peace is remarkably unsurprising (Table I). Similarly long periods of relative peace are common occurrences in the naturally variable statistics of interstate wars (a point recently made by Ref. [24]). The more unusual pattern in the past two centuries is not a long period of relative peace, but the dramatically violent period that preceded it (in which 42% of large wars occurred over 15% of the total time). This period was so violent over such a short period of time that the subsequent long peace simply balanced the statistical books (Fig. 5**b**), making it entirely plausible that the timing and sizes of interstate wars since 1823 were generated by a stationary process. Hence, the long peace is not evidence by itself of a change in the underlying mechanisms that generate conflict (Figs. 4**b**). In fact, the analysis here estimates that the post-war pattern of relative peace would need to endure in its current form for at least another 100 years before it would become statistically unusual enough to justify a claim that it represents a genuine trend (Fig. 5**b**).

A related finding is the remarkable stability of war onsets of any size since 1823 (Figs. 3 and 4**b**), whose statistics are consistent with a simple Poisson process, as originally proposed by Richardson [25, 26]. In other words, the annual hazard rate of a new interstate war is evidently also stationary, despite changes in many relevant factors [15], including the number of states, which increased by nearly an order of magnitude over this time period.

These results undermine the optimism of liberalism by implying that the enormous efforts after the Second World War to reduce the likelihood of large interstate wars have not yet changed the observed statistics enough to tell if they are working. This does not lessen the face-value achievement of the long peace, as the severity of a large war between major powers using modern military technology could be very large indeed, and there are real benefits beyond lives saved [16] that have come from increased economic ties, peace-time alliances, and the spread of democracy. However, it does highlight the continued relevance of the realist perspective and the appropriateness of a stationary process as the null hypothesis for patterns in interstate war. It also highlights the difficulty of understanding the role that human agency plays in driving trends and fluctuations in the statistics of interstate wars. Specifically, how can so much concerted effort, by so many individuals and organizations over so many decades of time, not be evidence of a genuine trend?

The answer may be simply a shift in perspective. Evidence from the study of complex social and biological systems [58] suggests that we often underestimate the importance of complexity and overestimate our ability to understand complicated causes of complicated effects, especially those that represent the aggregation of many inconsequential individual actions. Human agency certainly plays a critical role in shaping shorter-term dynamics and specific events in the history of interstate wars. But, the distributed and changing nature of the international system evidently moderates the impact that individuals or coalitions can have on longer-term and larger-scale system dynamics.

In this sense, the correct level of description for understanding trends in conflict may be the entire system, above the level of individual states, individual conflicts, or even individual peace- or war-promoting mechanisms. A pattern like the long peace could thus be real and understandable, produced by mechanisms that have genuinely reduced the likelihood of war over this period, and yet still be consistent with an overall stationary process running at a larger scale.

To illustrate this point, consider a professional basketball game and the ups, downs, and reversals in the lead size by one team over the other. As a spectator or player, one can readily explain why the lead size increased at one time or decreased at another. Each scoring event can be tied to specific actions and individuals within the game, and to individual or team strategies. At this level, scoring trends have interpretable causes that depend on actions at the same scale as the events themselves.

However, when thousands of individual games across teams and seasons are aggregated in order to consider basketball as a system, the relevance of such explanations blurs. Patterns at this scale cannot be attributed to specific actions or individuals, and instead emerge from subtle correlations within or constraints on collective behavior. Indeed, at this scale, the empirical statistics of lead sizes in basketball are nearly indistinguishable from those of a simple unbiased random walk [59, 60], and exciting trends within individual games are just statistical fluctuations at the system level.

Counter-intuitively, the stationary pattern of scoring in basketball appears despite the strategic efforts of highly-skilled players to make it otherwise. In fact, it may be precisely these independent efforts of skilled individuals in competition with each other that produces the observed stationary statistics [60]. This analogy suggests a new direction in conflict research, one aimed at identifying and testing mechanisms that could cause stationary statistics in the long term to emerge out of non-stationary dynamics at smaller temporal and spatial scales. A related direction would seek to understand what actions are necessary in order to genuinely alter these mechanisms and to change the characteristics of the stationary process.

Without a more clear understanding of the underlying mechanisms that drive the production of conflicts over long periods of time, or access to sufficiently broad and reliable data by which to identify them, a satisfying answer to the debate over trends in war may remain out of reach. Progress may be further complicated by the fact that interstate wars are only one conflict variable among many [5], which are surely interdependent. The analysis here indicates that none of the variations in the frequency and severity of wars since 1823 are statistically plausible trends. However, this finding may occur in part because there are compensatory trends in other variables that mask a subtle underlying change in the conflict generating processes. Proponents of the trend toward peace cite patterns across multiple conflict variables, or focus on patterns among developed nations, as evidence of a broad shift toward less violence [4, 8, 13, 14]. But, not all conflict variables support this conclusion, and some, such as military disputes and the frequency of terrorism, exhibit the opposite pattern [23, 43]. Untangling the interactions of various conflict variables, and characterizing both their trends and their differences across different groups of nations, is a valuable line of future work.

For instance, a stationary process for interstate wars is not inconsistent with an overall decline in per capita violence [4], because the human population has grown dramatically over the same period [24, 28]. The nonstationarity dynamics in human population, in the number of recognized states, in commerce, communication, public health, and technology, and even in the modes of war itself make it all the more puzzling that the hazard of interstate war in general has remained evidently so constant.

If the statistics of interstate wars are genuinely stationary, the risk over the next century of a very large war is uncomfortably high. The results here thus highlight the importance of continued efforts to ensure that the long peace endures and to prevent fragile peace-promoting institutions or systems from falling in the face of stable or contingent processes that drive the production of war. Much of this work must be done on the policy side. In the long run, however, research will play a crucial role by developing and evaluating mechanistic explanations, ideally at the system scale, of the likelihood of war [61, 62], which will help shed new light on what policies at what scales will promote peace.

Acknowledgements: The author thanks Kristian Skrede Gleditsch, Scott Atran, Lindsay Heger, Conor Seyle, Nils Petter Gledistch, Michael Spagat, Lars-Erik Cederman, and Bailey Fosdick for helpful conversations, and Isaac Asimov, John C. Wright, Alastair Reynolds, and Iain M. Banks for inspiration. A preliminary version of these results was presented at a Santa Fe Institute event at the 2012 Aspen Ideas Festival, and the current version was supported in part by the One Earth Future Foundation.

Appendix A: Distribution models

A power-law distribution for a continuous-valued random variable x has the form

$$\Pr(x \,|\, \alpha, x_{\min}) = \left(\frac{\alpha-1}{x_{\min}}\right)\left(\frac{x}{x_{\min}}\right)^{-\alpha} \quad , \tag{A1}$$

where $\alpha > 1$ and $x \geq x_{\min} > 0$. The choice of $x_{\min}$ indicates the smallest value above which the power-law behavior holds, and hence Eq. (A1) is a model of the war size distribution's upper tail. This specification replaces the difficult problem of modeling both the distribution's body and tail [42, 44, 45] with the less difficult problem of identifying a value $x_{\min}$ above which a model of the tail alone fits well.

For a particular choice of $x_{\min}$, the maximum likelihood estimator for the scaling parameter α given observed values $\{x_i\}$ is

$$\hat{\alpha} = 1 + n \left/ \sum_{i=1}^{n} \ln(x_i/x_{\min}) \right. \quad . \tag{A2}$$

When observed values are discrete but large ($x \geq 1000$), as in the case for interstate war sizes, Eq. (A1) is a statistically close approximation of the discrete power-law distribution [37].

The choice of $x_{\min}$ is obtained using the Kolmogorov-Smirnov goodness-of-fit statistic minimization technique (KS-minimization) [35, 37]. This method falls in the general class of distance minimization methods for selecting the size of the tail [45], and is widely used in the estimation of power-law tail models in empirical data.

The KS statistic [63] is the maximum distance between the CDFs of the data and the fitted model:

$$D = \max_{x \geq x_{\min}} |S(x) - P(x)| \quad , \tag{A3}$$

where $S(x)$ is the CDF of the data for the observations with value at least $x_{\min}$, and $P(x)$ is the CDF of the maximum-likelihood power-law model for the region $x \geq x_{\min}$. The estimate $\hat{x}_{\min}$ is then the value of $x_{\min}$ that minimizes D. In the event of a tie between several choices for $x_{\min}$, the smaller value is choosen, which improves the statistical power of subsequent analyses by choosing the larger effective sample size.

A geometric distribution for a discrete-valued random variable t has the form

$$\Pr(t \,|\, q, t_{\min}) = (1-q)^{-t_{\min}}\,(1-q)^{t} q \quad , \tag{A4}$$

where $q \in [0, 1]$ and $t \geq t_{\min} \geq 0$. For a particular choice of $t_{\min}$, the maximum likelihood estimator of the rate parameter q given observed values $\{t_i\}$ is

$$\hat{q} = \left(1 - t_{\min} + \frac{1}{n}\sum_{i=1}^{n} t_i\right)^{-1} \quad . \tag{A5}$$

In the limiting case of $t_{\min} = 0$, a conventional geometric distribution is recovered. In the analysis, the value of $t_{\min} = 1$ is fixed to represent a simple Bernoulli process for war onsets, as $t = 0$ represents the case where multiple onsets occur in a given year.

The statistical plausibility p of a tail model can be estimated via Monte Carlo under a semi-parametric null model in which the fitted model governs the frequencies above $x_{\min}$ and a nonparametric bootstrap governs the frequencies below. To generate synthetic data with the correct form in order to estimate the null distribution of the test statistic D, the following approach is used, as described in Ref. [37].

Let $\Pr(x \mid \theta, x_{\min})$ denote a particular tail model with parameters θ. Suppose that the observed data set $\{x_i\}$ has n_{tail} observations $x \geq x_{\min}$ and n observations in total. A new data set is then generated with n observations as follows. With probability n_{tail}/n, a random deviate y_i is drawn from $\Pr(x \mid \theta, x_{\min})$. Otherwise, with probability $1 - n_{\text{tail}}/n$, the variable y_i is set equal to an element chosen uniformly at random from among the observed values $\{x_i < x_{\min}\}$.

Repeating the process for all $i = 1 \ldots n$ generates a complete synthetic data set that follows $\Pr(x \mid \theta, x_{\min})$ above $x_{\min}$ and follows the empirical (non-tail) distribution below. The tail model is then fitted to $\{y_i\}$ by estimating $\hat{\alpha}$ and $\hat{x}_{\min}$, calculate the KS statistic for this model and these data, repeating this process n_{sample} times, construct a null distribution by which to evaluate the statistical significant of the empirical KS statistic.

In the joint model of war size and timing, the same semi-parametric procedure is used to generate synthetic deviates that follow a specified power-law distribution above $x_{\min}$ and the empirical data below. This approach allows us to leverage all of the available data to generate realistic synthetic conflict time series, which can be analyzed for various purposes.

Appendix B: The likelihood of a large interstate war

A simple test of the stationarity hypothesis for the sizes of wars is to consider whether the size of the Second World War, the largest observed event, is statistically anomalous relative to the overall distribution of war sizes. Here, the likelihood of such an event is estimated numerically, using a bootstrapping procedure on the empirical data to fit an ensemble of statistical models from which a probabilistic estimate of the likelihood of a large war is then derived [43]. If the likelihood of an event with severity $x_* = 16,634,907$ battle deaths or larger is nontrivial, then the size of the Second World War is not, in fact, statistically anomalous [43].

After removing the x_* event from the complete data set, an ensemble of power-law models is then constructed, each fitted independently to a bootstrap sample of the remaining 94 war sizes. Figure S1a shows such an ensemble of fitted tail models, along with the size distribution for the 94 war sizes, which fall well within the "cloud" of model lines. Similarly, the distribution of estimated model parameters shifts only slightly toward larger values, meaning slightly less heavy-tailed distributions, as a result of removing the largest event (Fig. S1a inset).

Across the ensemble of models, the average probability of observing at least one event out of 95 with severity at least x_* is $q = 0.525 \pm 0.002$, and the marginal probability that any particular war (drawn iid) is at least as large as x_* is $p_* = 0.0092355319$ (a uniform hazard rate).

Hence, the expected number of wars needed to observe one event of size x_* or larger is $1/p_* = 108$. With 95 observed wars over 181 years of time, war onsets have occurred, on average, every 1.91 years. Thus, the average recurrence time of an event of size x_* or larger is about 205 years, and the size of the Second World War cannot reasonably be considered an outlier with respect to the overall severity distribution of wars since 1823.

That said, the quantity q is only an average, and the expected variance in the arrival times of large wars will be large due to the heavy-tailed distribution of war sizes. As an analogy, consider the sizes of earthquakes, which also follow power-law statistics. There has been substantial scientific and popular interest in estimating the likelihood of another "great" quake in the San Francisco area. The last such event was a magnitude 7.8 quake in 1906, which led to the destruction of about 80% of the city and killed at least 3000 people. Simple estimates of the waiting time for another such earthquake vary from 80–200 years, and more sophisticated seismological models estimate an average waiting time of 101 years, with 68% of the density falling in the interval 40–162 years [64]. The development of comparable models for interstate conflicts would likely require significant advances in both our understanding of the conflict generating process and the uncertainty that underlies statistical forecasts.

Appendix C: War severities, before and after 1940

Here, a quantitative evaluation is made to assess whether the distribution of war sizes in the post-war period (1940–2003) differs from what would be predicted from the pattern of war severities in the pre-war period (1823–1939).

This question is considered in two ways. First, a forecast is made for the number of wars in the post-war period for different orders of magnitude, e.g., 10^3–10^4, 10^4–10^5, etc. and the differences between the post-war data and these predictions are examined. Second, the distribution of war severities after 1939 is compared to the ensemble of models fitted to the pre-war data. These two tests provide evidence for whether the observed frequencies of wars of different sizes in the post-war period differ from the pattern found in the pre-war period. If a substantial trend exists, then the observed post-war size distribution should clearly deviate from the ensemble of models fitted to pre-war data.

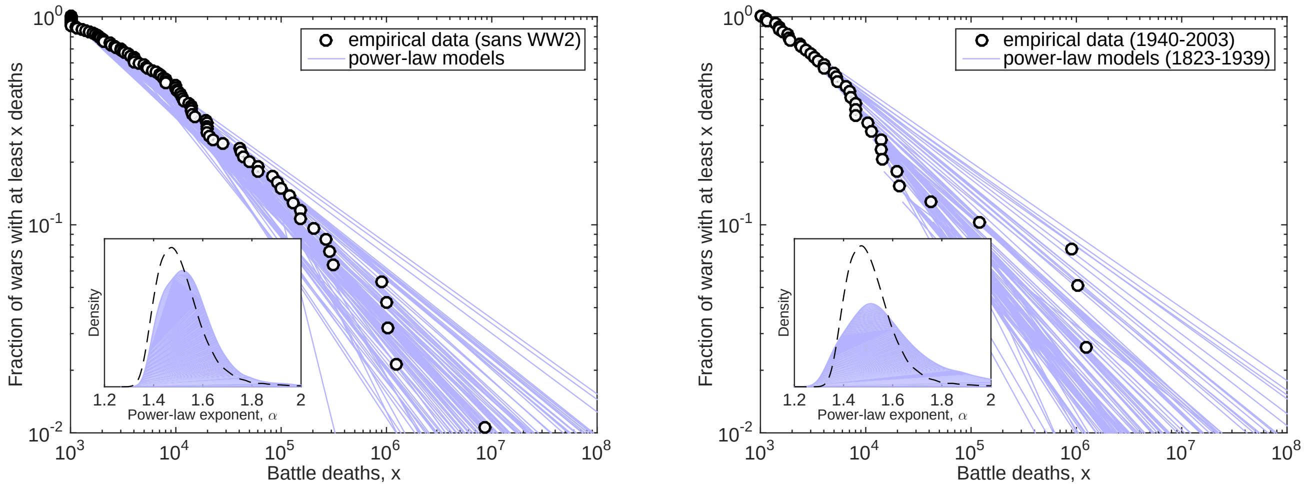

FIG. S1: Power-law tail models fitted to bootstrap samples of (**a**) all conflict sizes (1823–2003) but with the WW2 event removed, and (**b**) pre-war conflict sizes (1823–1939; events prior to WW2). In (**b**), the pre-war models (1823–1939) are compared with the post-war conflict sizes (1940–2003), showing good agreement. Insets plot the bootstrap distributions of $\hat{\alpha}$ for each set of tail models, showing little change relative to the distribution estimated from the full data set (dashed black line).

As before, an ensemble of power-law models is first fitted to bootstrap samples of the sizes of the 56 wars whose onset was in 1939 or earlier (the pre-war period). The average fraction of wars these models predict to occur is then computed by order of magnitude, and these model-driven forecasts are compared to the observed fractions of the 39 war severities in the post-war period (Table S1). The proportions of wars observed and predicted to fall within each severity class exhibit no simple relationship, and are relatively close in many places. This closeness suggests that the war sizes in the post-war period may not be generated by a dramatically different underlying process compared to those of the pre-war period.

However, some differences do occur, which may reflect an underlying trend toward peace. During the post-war period, that data contain roughly 7 times fewer moderate-sized wars (severity between 10^5–10^6) than predicted by the pre-war distribution, and no wars in the largest bins (more than 10^7 battle deaths). The data also contain 1.4 times more small wars (severity less than 10^4) and 1.6 times more larger wars (between 10^6–10^7) than predicted.

The lattermost statistic should be treated with particular caution as both the observed and predicted fractions at the upper end of the distribution are very small. Their small size tends to amplify the apparent degree of disagreement between a continuous prediction and an empirical distribution derived from a finite sample. The absolute differences (last column in Table S1) show that the observed and predicted numbers are in fairly close absolute agreement, except for the smallest and moderate-sized war bins. Deviations in the lower end of the distribution should be more reliable, suggesting that the sizes of some wars have genuinely been reduced relative to what the pre-war size distribution predicted.

Visually, however, the ensemble of models fitted to the pre-war data overlaps strongly with the empirical distribution of war sizes in the post-war period (Fig. S1**b**), suggesting that the deviations calculated in Table S1 may be transient fluctuations. In fact, using the maximum likelihood power-law fit on the pre-war sizes (1823–1939) as a reference model, a statistical hypothesis test indicates that this model cannot be rejected as a data-generative process for the post-war sizes (1940–2003; $p_{KS} = 0.92 \pm 0.03$). That is, the pre- and post-war size distributions are not statistically distinguishable.

Finally, the variation in the war-size distribution's parameter over time is considered. Dividing the 95 events into overlapping periods of time several decades in length, each containing 23 or 24 wars, a power-law model is fitted to the sizes of the wars whose onsets fall within each period. Comparing the time series of estimated exponents $\hat{\alpha}(t)$ with the stationary model, estimated from the full 95 events, shows interesting variation around the stationary

severity range	observed f_{obs}	expected f_{exp}	$f_{\mathrm{obs}}/f_{\mathrm{exp}}$	$f_{\mathrm{obs}} - f_{\mathrm{exp}}$
10^3–10^4	0.5385	0.3803	1.42	0.158
10^4–10^5	0.3590	0.3746	0.96	0.016
10^5–10^6	0.0256	0.1773	0.14	-0.152
10^6–10^7	0.0768	0.0470	1.64	0.030
10^7–10^8	0.0000	0.0139	—	-0.014
10^8–10^9	0.0000	0.0069	—	-0.007

TABLE S1: Observed and expected fractions of war sizes, by order of magnitude, for the post-war period (1940–2003).

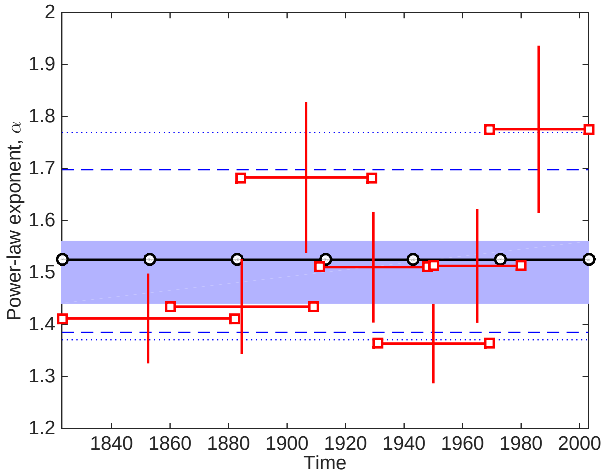

FIG. S2: Variability in the estimated scaling exponent $\hat{\alpha}$ (vertical lines show std. err. estimate) for events occurring within each of 7 overlapping periods 1823–2003, relative to the bootstrap distribution of exponents for all data together (solid blue indicates the middle 50% of $\Pr(\hat{\alpha})$, dashed lines mark the middle 90%, and dotted lines mark the middle 95%).

pattern (Fig. S2).

Two excursions of $\hat{\alpha}(t)$ above the stationary pattern indicate two periods of time during which conflicts were generally less severe than average. These excursions occur, not coincidentally, just prior to the First World War and after about 1970. These variations suggest that the war size distribution may fluctuate in more complicated ways than is assumed by a simple stationary model, and that changes in the distribution's shape may correspond to genuine periods of peace. The timing of the second period of lower-severity conflicts agrees with the popular hypothesis of a genuine but historically recent trend toward peace. However, the fact that a similar period of lower-severity conflicts occurred prior to the First World War, and was followed by a period of particularly severe conflicts, suggests that long periods of relatively more frequency lower-severity conflicts may simply be long transients in a fundamentally stable, but highly variable, process (see Discussion, main text).

Appendix D: War onsets, before and after 1940

Following the analysis of war severities, here the geometric model of peace durations is fitted to the 56 periods of peace that occurred prior to 1940 and use this model to make a forecast for the relative frequencies of periods of peace in the post-war period. If a substantial trend exists, then the observed post-war duration distribution should clearly deviate from the stationary prediction derived from the pre-war durations.

The observed and predicted proportions of delays of different lengths are relatively close to each other (Table S2), but like war severities, show no simple pattern. This closeness suggests that delays between onsets in the post-war period may not be generated by a dramatically different underlying process compared to those of the pre-war period, despite a substantially larger number of states worldwide. And, as with war sizes, some minor differences do appear. In particular, in the post-war period, the longest delay between new war onsets anywhere in the world is 7 years. Compared to the shape of the distribution over 1823–1939, the post-war period has a slight overabundance of 1-, 2- and 4-year periods of peace, but an under abundance of 3- and 5-year periods.

duration (years)	observed f_{obs}	expected f_{exp}	$f_{\text{obs}}/f_{\text{exp}}$	$f_{\text{obs}} - f_{\text{exp}}$
1	0.3947	0.3194	1.24	0.075
2	0.2632	0.1983	1.33	0.065
3	0.0789	0.1231	0.64	-0.044
4	0.0789	0.0764	1.03	0.003
5+	0.0263	0.1249	0.21	-0.099

TABLE S2: Observed and expected fractions of durations of peace (no new war onsets), for the post-war period.

Supporting this hypothesis of no clear trend in war onset delays between the pre- and post-war periods, a non-parametric two-sample Komogorov-Smirnov test finds that the deviation between the pre- and post-war delay distributions to be not statistically significant ($D^* = 0.0893$, $p_{\text{KS}} = 0.99$).

Appendix E: Variability in onset rate by war size

A useful way to represent the timing of war onsets of different sizes is the cumulative number of wars with severity at least x_q, where q denotes a quantile of the war size distribution. For example, $x_{0.25}$, $x_{0.50}$, and $x_{0.75}$ are the distribution's quartiles. The average production rate is simply the slope of this accumulation curve (Fig. S3). Over the 1823–2003 period, the production rate of wars of any severity remained fairly stable, with one new war beginning every 1.91 years, on average, with a standard deviation of 2.40 years.

Because larger wars are less frequent than smaller wars, the production rate for wars with severity $x \geq x_q$ must be lower than for some $x_{q'} < x_q$. In a stationary process, the variability around the average delay between events does not itself vary with war size. In contrast, if the relative variability increases as progressively larger conflicts are considered, then larger wars tend to cluster together in time, with longer periods of peace between them.

This possibility can be quantified by calculating the delay variable's coefficient of variation $c_v = \hat{\sigma}/\hat{\mu}$, which measures the sample variability of a quantity relative to its sample mean, for groups of wars with progressively greater severity. As a reference point, for all wars, the value is $c_v = 1.26$. As only those wars with severity above some level are considered, similar values of $c_v =$

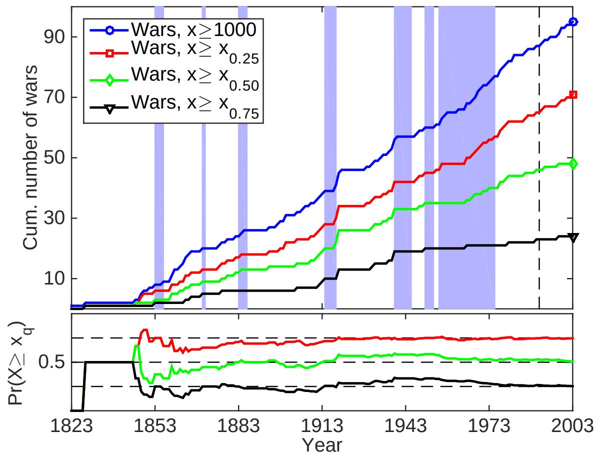

FIG. S3: (upper panel) The cumulative number of wars over the 1823–2003 period, for wars overall (blue), and for wars with severity above the top three quartile values (red, green, black). Shaded regions indicate the spans of a selection of significant wars, for visual reference. (lower panel) The cumulative fraction of wars over time, for the same thresholds in the upper panel, showing relatively stable proportions.

1.19, 1.11, and 1.03 are observed for the delays between war onsets for severity thresholds $x_{0.25}$, $x_{0.50}$, and $x_{0.75}$, respectively (Fig. S3).

The consistency in these coefficients of variation suggests that periods of peace tend to be distributed relatively evenly across time and across war severities, which is consistent with the stationarity hypothesis. However, these coefficients of variation represents calculations on a nested set of events. Hence, the slight by steady decrease in the calculated c_v as progressively higher quantile thresholds are considered is consistent with an underlying pattern in which temporal clustering tends to increase slightly among more severe conflicts (Fig. S3).

The most notable pattern appears in the accumulation curve for the most severe wars ($x \geq x_{0.75}$ in Fig. S3): here, the production rate in the post-war period appears substantially lower than those of the lower quartile accumulation curves. That is, the shape is consistent with a relative slowdown in the production of these most-severe conflicts during the post-war period, compared to either the production of smaller conflicts in any period, or the production of the same size of conflicts in the pre-war period. This type of pattern qualitatively agrees with the analysis of war sizes and the long peace pattern after the Second World War.

To quantify this decrease in the production rate of the most-severe conflicts after the Second World War, the data are split into two groups: conflicts with onsets in the 1823–1939 period and those with onsets in the 1940–2003 period. These conflicts are then further split into those with severity in the upper quartile (most severe) and those not in the upper quartile (all others). In the post-war period, a new conflict in the most severe group occurs on average every 12.8 years, while the average delay for all others is 1.88 years (or, roughly 6.8 smaller conflicts per large event). In contrast, in the pre-war period, the average delays are 6.2 years and 3.2 years, respectively (or, roughly 1.9 smaller conflicts per large event). That is, there are longer periods of peace between the most-severe conflicts in the post-war era, but shorter periods between smaller conflicts.

These statistics are skewed, to some extent, by the great violence period of 1914–1939 period. Excluding the conflicts in this period shows that the earlier 1823–1913 period, leading up to the great violence, is relatively closer to the 1940–2003 period that followed. Specifically, the average delays in earlier period are 10.1 years and 3.1 years, for the most severe and all other conflicts (or, roughly 3.3 smaller conflicts per large event). That is, the production rate of the largest events before and after the great violence were fairly close, with the main difference being in the production rate of all other conflicts, which was higher in the long peace that followed the great violence.

[1] M. R. Sarkees, F. Wayman, *Resort to War: 1816 – 2007* (CQ Press, Washington, DC, 2010).

[2] A. Gat, *War in Human Civilization* (Oxford University Press, Oxford, 2005).

[3] D. P. Fry, *Beyond War: The Human Potential for Peace.* (Oxford University Press, Oxford, 2006).

[4] S. Pinker, *Better Angels of Our Nature* (Penguin Books, 2012).

[5] K. S. Gleditsch, A. Clauset, *Handbook of International Security*, W. C. Wohlforth, A. Gheciu, eds. (Oxford University Press, 2018).

[6] J. S. Levy, W. R. Thompson, *Causes of War* (Wiley-Blackwell, West Sussex, UK, 2010).

[7] L.-E. Cederman, Modeling the size of wars: From billiard balls to sandpiles. *American Political Science Review* **97**, 135–150 (2003).

[8] J. L. Payne, *A History of Force: Exploring the Worldwide Movement against Habits of Coercion, Bloodshed, and Mayhem* (Lytton, Sandpoint, ID, 2004).

[9] S. J. Cranmer, B. A. Desmarais, E. J. Menninga, Complex dependencies in the alliance network. *Conflict Management and Peace Science* **29**, 279–313 (2012).

[10] M. O. Jackson, S. M. Nei, Networks of military alliances, wars, and international trade. *Proc. Natl. Acad. Sci. USA* **112**, 15277–15284 (2015).

[11] J. L. Gaddis, The long peace: elements of stability in the postwar international system. *International Security* **10**, 99–142 (1986).

[12] A. M. Saperstein, The "Long Peace" – result of a bipolar competitive world? *Journal of Conflict Resolution* **35**,

68–79 7 (1991).
[13] J. S. Goldstein, *Winning the War on War* (Dutton, Hialeah, FL, 2011).
[14] T. R. Gurr, Ethnic warfare on the wane. *Foreign Affairs* **79**, 52–64 (2000).
[15] L.-E. Cederman, Back to Kant: Reinterpreting the democratic peace as a macrohistorical learning process. *American Political Science Review* **95**, 15–31 (2001).
[16] M. Roser, E. O. Ospina, J. Mispy, Our World in Data (2017). `https://ourworldindata.org`.
[17] J. L. Ray, Does democracy cause peace? *Annual Review of Political Science* **1**, 27–46 (1998).
[18] M. D. Ward, R. M. Siverson, X. Cau, Disputes, democracies, and dependencies: A reexamination of the Kantian peace. *American Journal of Political Science* **51**, 583–601 (2007).
[19] J. S. Levy, Alliance formation and war behavior: An analysis of the great powers, 1495–1975. *Journal of Conflict Resolution* **25**, 581–613 (1981).
[20] B. A. Leeds, Do alliances deter aggression? The influence of military alliances on the initiation of militarized interstate disputes. *American Journal of Political Science* **47**, 427–439 (2003).
[21] W. Li, A. E. Bradshaw, C. B. Clary, S. J. Cranmer, A three-degree horizon of peace in the military alliance network. *Science Advances* **3**, e1601895 (2017).
[22] A. Alesina, E. Spolaore, On the number and size of nations. *Quarterly Journal of Economics* **112**, 1027–1056 (1997).
[23] M. Harrison, N. Wolf, The frequency of wars. *Economic History Review* **65**, 1055–1076 (2012).
[24] P. Cirillo, N. N. Taleb, On the statistical properties and tail risk of violent conflicts. *Physica A* **429**, 252–260 (2015).
[25] L. F. Richardson, The distribution of wars in time. *Journal of the Royal Statistical Society* **57**, 242–250 (1944).
[26] L. F. Richardson, *Statistics of Deadly Quarrels* (The Boxwood Press, Pittsburgh, 1960).
[27] R. M. Siverson, M. D. Ward, The long peace: A reconsideration. *International Organization* **56**, 679–691 (2002).
[28] B. F. Braumoeller, Is war disappearing? (2013). APSA Chicago 2013 Meeting. Available at SSRN: `https://ssrn.com/abstract=2317269`.
[29] S. Pinker, Fooled by belligerence: Comments on Nassim Taleb's "The Long Peace is a statistical illusion" (2012). Preprint, `http://bit.ly/1UD3ynq` (accessed 12 November 2012.
[30] M. Basseville, I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application* (Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993).
[31] M. Small, J. D. Singer, *Resort to Arms: International and Civil Wars, 1816–1980* (Sage Publications, Beverly Hills, 1982).
[32] L. F. Richardson, Variation of the frequency of fatal quarrels with magnitude. *Journal of the American Statistical Association* **43**, 523–546 (1948).
[33] M. Mitzenmacher, A brief history of generative models for power law and lognormal distributions. *Internet Mathematics* **1**, 226–251 (2004).
[34] M. E. J. Newman, Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* **46**, 323–351 (2005).
[35] A. Clauset, M. Young, K. S. Gleditsch, On the frequency of severe terrorist events. *Journal of Conflict Resolution* **51**, 58–87 (2007).
[36] W. J. Reed, B. D. Hughes, From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. *Physical Review E* **66**, 067103 (2002).
[37] A. Clauset, C. R. Shalizi, M. E. J. Newman, Power-law distributions in empirical data. *SIAM Review* **51**, 661–703 (2009).
[38] P. Bak, C. Tang, Earthquakes as a self-organized critical phenomena. *Journal of Geophysical Research* **94**, 15635–15637 (1989).
[39] B. D. Malamud, G. Morein, D. L. Turcotte, Forest fires: An example of self-organized critical behavior. *Science* **281**, 1840–1841 (1998).
[40] M. Biggs, Strikes as forest fires: Chicago and Paris in the late 19th century. *American Journal of Sociology* **111**, 1684–1714 (2005).
[41] R. J. Adler, R. E. Feldman, M. S. Tuqqu, eds., *A Practical Guide to Heavy Tails: Statistical Techniques and Applications* (Birkhäuser, Boston, 1998).
[42] S. I. Resnick, *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling* (Springer-Verlag, New York, 2006).
[43] A. Clauset, R. Woodard, Estimating the historical and future probabilities of large terrorist events. *Annals of Applied Statistics* **7**, 1838–1865 (2013).
[44] L. de Hann, A. Ferreira, *Extreme Value Theory: An Introduction* (Springer-Verlag, New York, 2006).
[45] R.-D. Reiss, M. Thomas, *Statistical Analysis of Extreme Values: with Applications to Insurance, Finance, Hydrology and Other Fields* (Birkhäuser, Basel, Switzerland, 2007).
[46] E. J. Gumbel, The return period of flood flows. *The Annals of Mathematical Statistics* **12**, 163–190 (1941).
[47] B. Gutenberg, C. F. Richter, Frequency of earthquakes in California. *Bulletin of the Seismological Society of America* **34**, 185–188 (1944).
[48] W. J. Reed, K. S. McKelvey, Power-law behavior and parametric models for the size-distribution of forest fires. *Ecological Modelling* **150**, 239–254 (2002).
[49] B. Efron, R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, New York, NY, 1993).
[50] S. Bowles, Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science* **324**, 1293–1298 (2009).
[51] M. Spagat, Is the risk of war declining? (2015). Sense about Science blog, `http://bit.ly/2oNGKtT`.
[52] L.-E. Cederman, T. C. Warren, D. Sornette, Testing Clausewitz: Nationalism, mass mobilization, and the severity of war. *International Organization* **65**, 605–638 (2011).
[53] S. A. Bremer, Dangerous dyads: Conditions affecting the likelihood of interstate war, 1816-1965. *Journal of Conflict Research* **36**, 309—341 (1992).
[54] E. D. Mansfield, J. Snyder, Democratization and the danger of war. *International Security* **20**, 5–38 (1995).
[55] K. Barbieri, Economic interdependence: A path to peace or a source of interstate conflict? *Journal of Peace Research* **33**, 29–49 (1996).
[56] G. Schrock-Jacobson, The violent consequences of the nation: Nationalism and the initiation of interstate war. *Journal of Conflict Resolution* **56**, 825-852 (2012).
[57] A. D. Smith, National identity and the idea of European unity. *International Affairs* **68**, 55–76 (1992).
[58] M. Mitchell, *Complexity: A Guided Tour* (Oxford University Press, 2011).
[59] A. Gabel, S. Redner, Random walk picture of basketball

scoring. *Journal of Quantitative Analysis in Sports* **8**, 8 (2012).

[60] A. Clauset, M. Kogan, S. Redner, Safe leads and lead changes in competitive team sports. *Physical Review E* **91**, 062815 (2015).

[61] L.-E. Cederman, N. B. Weidmann, Predicting armed conflict: Time to adjust our expectations? *Science* **355**, 474–476 (2017).

[62] J. M. Hofman, A. Sharma, D. J. Watts, Prediction and explanation in social systems. *Science* **355**, 486–488 (2017).

[63] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, Cambridge, UK, 1992).

[64] J. B. Rundle, P. B. Rundle, A. Donnellan, D. L. Turcotte, R. Shcherbakov, P. Li, B. D. Malamud, L. B. Grant, G. C. Fox, D. McLeod, G. Yakovlev, J. Parker, W. Klein, K. F. Tiampo, A simulation-based approach to forecasting the next great San Francisco earthquake. *Proc. Natl. Acad. Sci. USA* **102**, 15363–15367 (2005).

[65] Because the production rate of wars before 1846 is essentially zero, the times of the two wars 1823–1845 are fixed, and use the Bernoulli process thereafter. This centers the empirical accumulation curve within the simulated curves and constructs a more reasonable test.