ON (THE LIKELIHOOD OF) WAR

S G Coulson
CCW Research Fellow
June 2022
(N.B. this paper was originally produced in 2018)

Abstract

The Breke (1999) Conflict Catalogue is used to analyse trends in conflict. It is found that contrary to earlier research, the likelihood of war occurring has not decreased and that the outbreak of conflict is not random. There is no evidence to support claims that the world is getting more peaceful. The outbreak of large scale war is statistically possible with a frequency of greater than one per ninety years and the likelihood of smaller conflicts breaking out is much greater with at least one likely to occur every year. The implications of this study are that there is no evidence to support claims that the world is getting more peaceful. The outbreak of large scale war is statistically possible with a frequency of greater than one per ninety years and that the likelihood of smaller conflicts is much greater, with an expected occurrence of at least one per year.

Introduction

This paper considers the following three questions:

a. What is the likelihood of the outbreak of war?[1]

b. Does the occurrence of one war affect the likelihood of the outbreak of subsequent wars?

c. What is the interval between the outbreaks of war?

Despite the seemingly endless number of wars throughout the history of mankind, the outbreak of war is a relatively rare event. Overall deaths from warfare account for around 1% off all deaths (Hayes 2002); small compared with deaths from accidents and suicides. While the likely impact on an individual may be insignificant, the outcome of war is often existentialist at the social or state-level. Hence war occupies a much greater importance in the minds of social decision-makers than its consequences for overall life-expectancy merits.

Recent researchers (Pinker 2011, Freedman 2016) have considered that the probability of war occurring may have reduced since 1945 or is reducing as part of an of even-longer running trend extending over hundreds of years. Gaddis (1986) coined the term *long peace* to describe an apparent absence of warfare between major powers.

[1] Here the likelihood of war is taken to be the number of wars breaking out per year.

Conversely other researchers (Crillo & Taleb 2016) using data sets covering over two thousand years of conflict conclude that there is no evidence to support the theory that the likelihood of war is in decline. These and other authors (Clauset 2018) consider that the current 70-year long peace is statistically plausible and should not be taken as evidence that the outbreak of war is becoming less likely.

Disagreement between researchers is unsurprising given that discerning trends in conflict is complicated by differing definitions of conflict, different measures of conflict duration/intensity and incomplete datasets. The next section reviews these factors and identifies some of the sensitivities that can lead to very different conclusions in the direction of trends.

Conflict definitions and datasets

There is no overall agreement as to what counts as conflict other than a general consensus that conflict includes interstate war; some researchers include civil wars and insurgencies within their data. In the case of civil wars consistency in deciding what should be counted as conflict and what is internal unrest is problematic. For example, excluding all civil wars would remove the US Civil War (1861-1865) one of the major conflicts of the 19th Century from the study. Similarly, there are a number of cases in history where one state has not resisted when attacked by another such as the German invasion of Denmark in 1940 (Davis *et al.* 1978). Both civil wars and undisputed attacks can confuse the definition of what should be considered an outbreak of conflict and can make comparison between different datasets difficult. Richardson (1960), in one of the earliest studies into the statistics of conflict, used an exceptionally wide definition of conflict which included homicides and events with a small number of deaths that would not ordinarily be described as a war. Other studies consider only *large* conflicts, defined by imposing some arbitrary lower bound on the number of fatalities such as 26,625 (Clauset 2018).

The use of conflict related deaths is itself complicated by the decision to include non-combatant fatalities. Fischoff *et al.* (2007) note the problem posed by the use of mercenaries or private military contractors whose combatant status is ambiguous.

Similarly, determining over what period latent deaths can be ascribed to a conflict is equally problematic. For example, post traumatic stress disorder related deaths or chronic illnesses resulting from a particular conflict may cause deaths some years after the actual event. The UK Government recommends that not all deaths should be considered equally but gives greater weight to deaths of young or healthy people (HM Treasury 2005); however, this method has not yet been used to measure fatalities in conflict.
In practical terms the number of possible deaths from a given conflict is bounded above by the size of the population over the duration of that conflict. This has led some researchers to normalise the number of deaths in a particular conflict by some reference population such as that of the combatant nation or the global population (Crillo & Taleb 2016) or by considering dyads (Davis *et al.* 1978). The use of such normalisation implicitly assumes that the size of the population is a factor in determining the lethality of war (Clauset 2018). To avoid this difficulty, unless state otherwise, the absolute number of combat fatalities (military and civilian) are considered in this study.

Here we use Brecke's (1999) Conflict Catalogue, which records the number of violent global conflicts and casualties from 1400 AD to 2000. This data set was chosen as it offers the widest coverage of global conflict over the longest number of years in comparison with other data sources (eg. Richardson 1960, Uppsala Conflict Data Program (UCDP)).

Using the Catalogue, the likelihood of conflict, the number of conflicts that commenced per year are plotted for each year from 1400 to 2000 in Figure 1. There are 3,708 conflicts recorded during the 601 years covered by the Catalogue. Over this period the likelihood of conflict ranged between 0 and 23, the mean likelihood of conflict is 6.2.

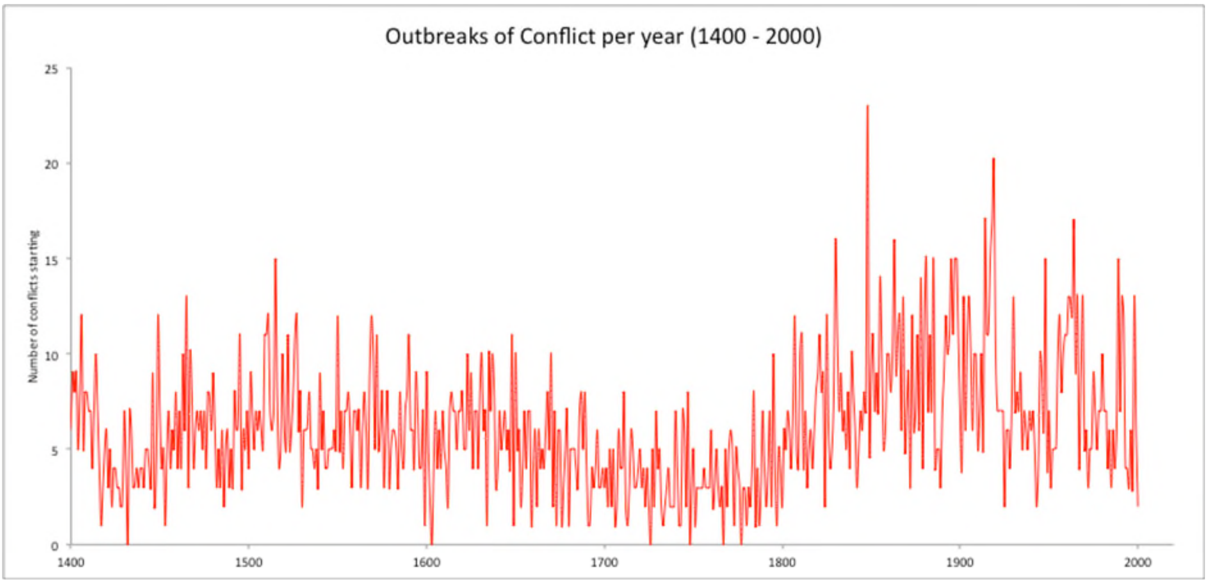

Figure 1: The number of conflicts starting per year for the years 1400-2003 using global conflict data taken from Brecke (1999).

If we consider arbitrary subsets of the database, say 200-year intervals, the mean likelihood of war varies from 4.4 (1600-1800) to 8.1 (1800-2000) (Table 1). This simple partition of the dataset suggests that the likelihood of war is increasing rather than decreasing; however, the results are sensitive to the size of the interval considered. For example, if instead we consider 100-year intervals (Table 2), the mean likelihood of war varies from 3.5 (1700-1800) to 8.2 (1800-1900). The changes in the 100-year interval mean suggests that the likelihood of war was increased from the 15th to the 16th Century and then deceased to a low in the 19th Century, a period which saw the 30 Years War and the Wars of Spanish Succession in Europe before increasing to a maximum in the 19th Century and actually decreased in the century of the First and Second World Wars.

Interval	**Mean likelihood of war**
1400-1600	6.0
1500-1700	5.9
1600-1800	4.4
1700-1900	5.9
1900-2000	8.1

Table 1: Mean likelihood of war for 200-year periods from 1400-2000 using data from Brecke (1999).

Interval	Mean likelihood of war
1400-1500	5.5
1500-1600	6.5
1600-1700	5.3
1700-1800	3.5
1800-1900	8.2
1900-2000	8.0

Table 2: Mean likelihood of war for 100-year periods from 1400-2000 using data from Brecke (1999).

The mean likelihood of war is also sensitive to where the intervals are taken from, so if we consider 100 periods from 1450 (Table 3), the mean varies from 4.1 (1650-1750) to 8.8 (1850-1950).

Interval	Mean likelihood of war
1450-1550	6.3
1550-1650	6.2
1650-1750	4.1
1750-1850	5.3
1850-1950	8.8

Table 3: Mean likelihood of war for 100-year periods from 1450-1950 using data from Brecke (1999).

Tables 1-3 illustrate some of the sensitivities in calculating the mean likelihood of war. No database of conflict can be entirely accurate in the extent of its coverage and the interval of years considered is by nature arbitrary. Depending on the intervals chosen very different conclusions can be drawn as to what constitutes a trend in warfare. This has caused different researchers to argue that the likelihood of conflict is increasing, decreasing or cyclic in nature. The differences in mean shown by Tables 1-3 strongly suggest that it would be premature to conclude that the likelihood of conflict is increasing or decreasing based analysis of what will always be a subset of all global conflicts. Further detailed statistical modelling is required to determine the time dependence on the mean likelihood of the outbreak of conflict.

From the Brecke dataset, the mean likelihood of war from 1400-2000 is 6.1 wars start every calendar year. This average is highly sensitive to dataset used to determine the onset of conflict, the interval over which conflicts are averaged and the size of the interval. For the single dataset considered here, the likelihood of the outbreak of war varied from 3.5 to 8.8 depending on which specific interval or width of interval of years was considered. Two conclusions can be drawn from this study, first that the likelihood of war is highly sensitive to the dataset, length of interval and specific interval chosen for analysis. Secondly, over all 601 years surveyed, there have only been six years when no conflict outbreaks have been recorded, so that the likelihood of at least one conflict breaking out in a given year over this period is greater than 99%. So, globally, over the last 601 years the likelihood of at least one war/conflict occurring for every calendar year is almost certain and there are no clear indications that the likelihood of war has either increased or decreased over this period.

These uncertainties associated with determining the mean likelihood of war applies equally to other conflict statistics. The value of mean likelihood chosen can significantly affect analysis of trends in conflict. An example of this is the question of the randomness of the outbreak of war, which is discussed in the next section.

The distribution of wars in time

Does the occurrence of one war affect the likelihood of the outbreak of subsequent wars? This is akin to asking how the distribution of the occurrence of war varies over time. Let $x,\ 0 \leq x < \infty$ be the number of conflicts breaking out per year and let $f(x)$ be the number of calendar years with x conflicts breaking out; the distribution of the outbreak of conflict is shape of the curve $f(x)$.

Richardson (1960) was the first to consider the distribution of wars in time. From analysis of a dataset of 59 conflicts from 1820-1929, Richardson concluded that the occurrence of the outbreak of war follows a Poisson distribution. If correct this would imply that the outbreak of war is random, so that the occurrence of one war does not influence the outbreak of subsequent wars.

Repeating Richardson's analysis using Brecke's (1999) Conflict Catalogue, the distribution of conflict is shown as a histogram (red bars) in Figure 2. The blue lines show the expected distribution for the likelihood of conflict if the distribution follows a Poisson distribution with the same mean as the likelihood of conflict from the Brecke data (ie. 6.1). Application of the chi-squared test statistic implies that the hypothesis that the distribution of conflict in time follows a Poisson distribution should be rejected.[2] The implications of this conclusion are that the distribution of war may not be random but obeys some other distribution will be discussed in the next section.

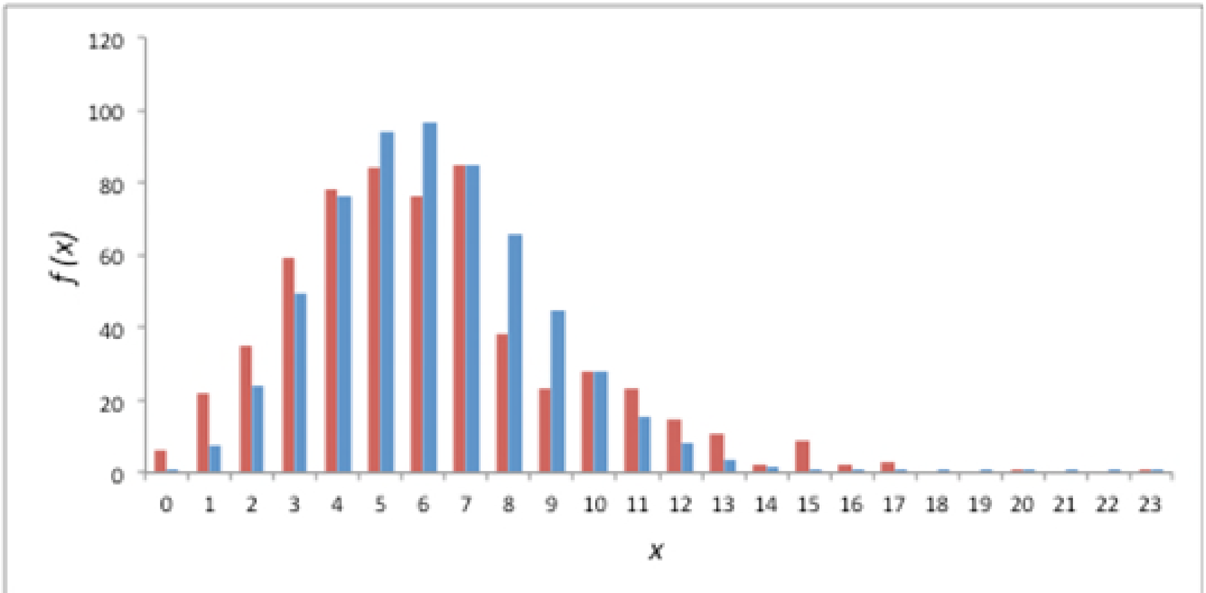

Figure 2: The red bars show the distribution of outbreaks of x conflict per calendar year, $f(x)$ versus the number of outbreaks of conflict (x) in a year for conflicts from 1400 – 2000 compared with a Poisson distribution with $\lambda = 6.1$.

[2] The conclusion is the same if a more statistically rigorous test, Fisher probability with the chi-squared value for $n-1$ degrees of freedom is used; ie., the Poisson distribution should be rejected.

Richardson's original analysis only considered conflicts that caused fatalities within a specified range 3,162-31,622[3]. The distribution in Figure 2 used all wars within the Brecke dataset not just those within Richardson's casualty range. Within the Brecke database there are 125 conflicts, with a mean likelihood of 1.1 that produced 3,162-31,622 casualties with in the years 1820-1929; more than double the number of conflicts identified by Richardson. The blue columns in the histogram in Figure 3 show the distribution of the number of outbreaks of x conflict per calendar year $f(x)$ versus the number of outbreaks of conflict (x) in a year for conflicts from 1820-1929 within the casualty range considered by Richardson.

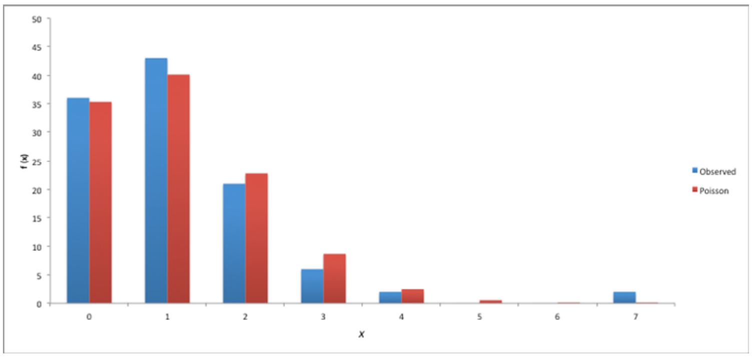

Figure 3: The blue columns show the distribution of the number of outbreaks of x conflict per calendar year, $f(x)$ versus the number of outbreaks of conflict (x) in a year for the 125 conflicts from 1820 – 1929 that produced between 3,162 & 31,622 casualties. The red columns show the expected Poisson distribution for 125 conflicts with $\lambda = 1.1$.

Statistical analysis leads us to reject the hypothesis that the distribution of the outbreak of conflict within Richardson's casualty range is consistent with a Poisson distribution.

Similarly, if we ignore the upper bound on Richardson's casualty range and consider the 654 conflicts that produced more than 3,162 casualties over the full period of Brecke's dataset (1400-2000), and if we follow Richardson's criteria for classifying major conflicts as those which resulted in greater than 3,162 casualties and consider the time between outbreaks of these major conflicts, Brecke's dataset identifies 654 wars that fall into this category between 1400 and 2000. For this subset the mean liklihood of the outbreak of major conflict is ≈ 1.1. A hystogram of the distribtuion of the x conflict per calendar year, $f(x)$ versus the number of outbreaks of conflict (x) in a year for the 654 conflicts compared with a Poisson distribution is shown in Figure 4.

[3] Richardson (1960) smoothed his data by expressing conflict fatalities on a logarithmic scale, so that 3,162 -31,622 fatalities corresponds to $\log_{10} 3.5 – \log_{10} 4.5$.

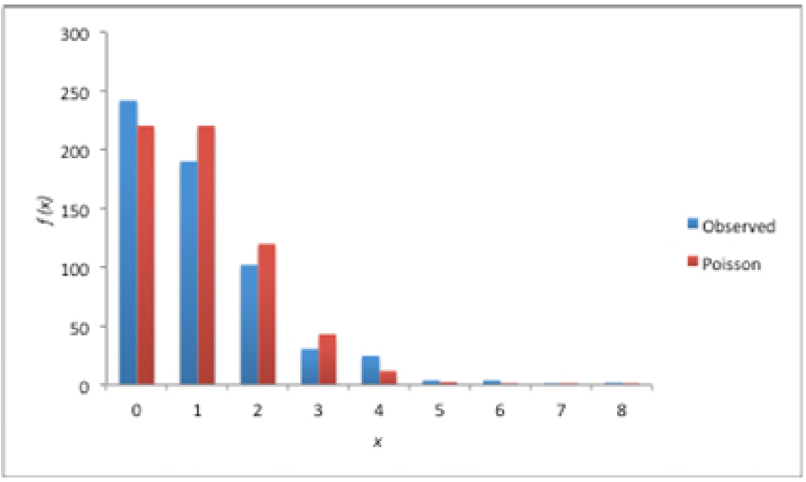

Figure 4: The blue columns show the distribution of the number of outbreaks of x conflict per calendar year, $f(x)$ versus the number of outbreaks of conflict (x) in a year for the 654 conflicts from 1400 - 2000 that produced greater than 3,162 fatalities. The red columns show the expected Poisson distribution for 654 conflicts with $\lambda = 1.1$.

Statistical analysis leads us to reject the hypothesis that the distribution of the outbreak of conflict of statistical analysis rejects the Poisson distribution hypothesis.

In both cases the main cause of the deviation from the Poisson distribution is the small number of years that experienced a high number of outbreaks of conflict. The use of alternative, heavy-tailed distributions such as the exponential distribution or the Lorentzian distribution may provide better representations of the outbreak of conflict than the Poisson distribution. If so, this would suggest that the likelihood of a given number conflicts starting within a calendar year is not a random event, independent of the number of conflicts starting in the previous calendar year. This contradicts the findings of Richardson (1960) and others.

The analysis in this section assumed that the mean likelihood of conflict was stationary (ie. time independent) over each time interval considered. The use of time dependent means can be estimated using the different means given in Tables 1-3, using these means does not lead to a Poisson type distribution for the outbreak of conflict.

This suggests that, as in the above section, the distribution of the outbreak of major conflict is not a random event. If the distribution of the outbreak of wars in time is not random, what about the interval between wars? The next section investigates the distribution of the number of years of peace between conflicts.

The interval between conflicts

Instead of counting the outbreaks of war in a given calendar year, let us consider the interval between major conflicts; ie. the number of years between major conflicts starting. Of the 654 major conflicts in the Brecke dataset the interval between their outbreaks varied between 0 and 11 years. Over the 601 years covered by the dataset, the mean interval between outbreaks of major conflict was ≈ 0.92. That is, on average between 1400 and 2000, a major war started every 336 days.

We now let (x) be the number of calendar years between the outbreak of conflicts that produced greater than 3,162 fatalities, ie. the peace interval. Figure 5 shows the distribution of $f(x)$, the number of conflicts with a peace interval of length (x) years. A quick comparison between the observed distribution and the expected Poisson distribution with a mean of 0.92 suggests that $f(x)$ is unlikely to be a Poisson distribution and this is confirmed by statistical analysis: so that the interval between the outbreaks of war is unlikely to be random.

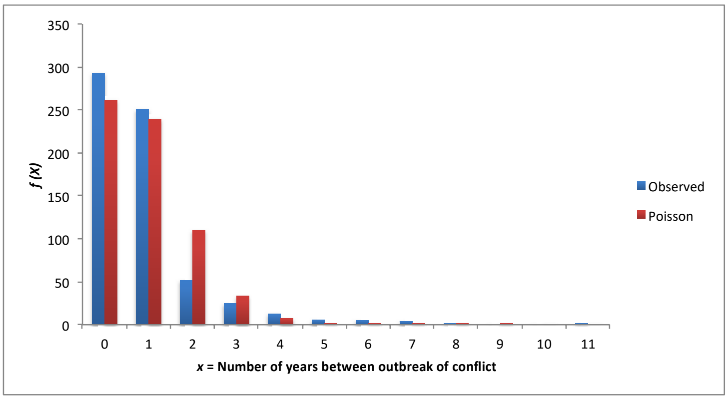

Figure 5: The blue columns show the distribution of the number of outbreaks of x conflict $f(x)$ versus the number of years between outbreaks of conflict (x) for the 652 of the 654 conflicts from 1400 - 2000 (ie. excluding the two conflicts that occurred in 1400 and 1999) that produced greater than 3,162 fatalities. The red columns show the expected Poisson distribution for 652 conflicts with $\lambda = 0.92$.

Where there is a departure from true randomness, we expect that clusters or bunching may occur, so that some periods of history may experience a greater frequency of major conflicts compared with others. If conflicts are not random and clustering may occur, is it still possible to tell if a long interval of peace is to due to statistical clustering rather than some underlying cause that means mankind is getting genuinely less bellicose? Queuing theory is a mathematical approach to analyse waiting time in systems and has applied to answer problems in clustering (Gunther 2010); one of the most famous being why buses come in threes (Eastaway & Wyndham 1998).

In Queuing theory the peace interval between conflicts is called the *Lifetime*. When trying to predict the time to the next conflict we do not know what the lifetime is, what we do know is the time between today and the previous conflict, which called the *Age*. The period between now and the outbreak of the next conflict is known as the *Residual Life*. So that the *Lifetime* between conflicts is equal to the sum of the *Age* and the *Residual Life*.

The *Residual Life* is a function of not just the mean lifetime between conflicts, but also the variance about the mean lifetime (Kleinrock 1975). The variance can be considered to be an average difference between each observed lifetime and the mean lifetime. Calculating the *Residual Life* allows us to estimate the most likely time to the next conflict.

For simplicity, let us consider the interval between the outbreaks of very large conflicts or mega wars, which we will define as those that cause greater than one million casualties. The Brecke database records eight examples of these mega wars between 1400 and 2000. The mean number of calendar years between the outbreak of one mega war and the outbreak of the next mega war is 43.6 years, with a variance of 75.1 years. This gives a Residual Life of around 86 years.

The last occurrence of mega war recorded by Brecke was Nigeria Civil War (Biafra War) from 1967-1970. A residual life of 86 years implies that the next mega war should be expected by 2053. Estimates of the total casualties in the Nigeria Civil War are highly uncertain, if we exclude this conflict from the list of mega wars, the residual lifetime reduces to around 67 years. The last outbreak of mega war in the Brecke dataset is the Vietnam War (1964 19975), which implies the next mega war would be expected by 2031. This suggests that the current absence of a major conflict is not statistically unusual and that the current *long peace* falls within the expected peace interval.

If we consider not just mega wars but all 654 conflicts in the Brecke dataset that produced greater than 3,162 fatalities, the mean lifetime between the outbreak of conflicts is 0.9 years with a variance of 2.0 years. This implies that the residual lifetime of conflicts producing greater than 3,162 fatalities is around 2.5 years. The average duration of conflicts producing greater than 3,162 fatalities is around 4.2 years. So that it is highly likely that there is always at least one conflict producing greater than 3,162 fatalities per year occurring at any given time.

Conclusions

Determining trends on conflict is highly sensitive to the dataset used, the definition of conflict, the length of interval and specific interval chosen for analysis. Bearing in mind these caveats, analysis of the Brecke dataset has enabled the investigation of the three questions posed at the start of this paper: the likelihood of the outbreak of war; how the occurrence of the outbreak of war varies over time; and, if the length of the interval between outbreaks of war.

Over the 601 years covered by the Breke dataset, the likelihood of at least one conflict breaking out somewhere in the world in any given year over this period is greater than 99%. There are no clear indications that the likelihood of war has either increased or decreased over this period
It was found that the distribution of wars; the variation of the outbreak of war over time does not fit a Poisson distribution for all conflicts recorded in the Breke data set. This contradicts the findings of Richardson (1960) and others. If we restrict our analysis to the same criteria used by Richardson, conflicts that produced between 3,162 and 31,622 casualties and occurred during the years 1820 and 1929, the Breke dataset suggests that these conflicts do not follow a Poisson distribution. Similar conclusions are obtained if we consider all conflicts that produced above 3,162 casualties over the 602 years surveyed by Breke. In all cases there is a small number of years where a large number of conflicts break out giving a tail to the distribution that varies from Poisson behaviour.

If the likelihood of a given number conflicts starting within a calendar year is not a random event, this implies that is not independent of the number of conflicts starting in the previous calendar year, so that the outbreak of conflict may well be correlated with pervious conflicts. There are numerous historical cases studies where at least one of the causes outbreak of one conflict has been shown to be an earlier conflict. The First World War being a factor in the outbreak of the Second World War, and the end of the Hundred Years War influencing the War of the Roses, are just two examples of correlation of conflict.

This deviation from Poisson behaviour allows conflicts to be studies using Queuing Theory where the residual life, the time between the present and the next outbreak of war is a function not just of the mean lifetime between conflicts, but also the variance of the mean lifetime. Applying this to the Breke dataset, the residual lifetime of conflicts that produce greater than one million casualties, mega wars is found to be 86 years. This suggests that the current *long peace* with no mega wars occurring is not statistically significant. If we extend this analysis to cover all conflicts producing greater than 3,162 fatalities per year, it is highly likely that there is always at least one conflict occurring at any given time.

References:

Brecke, P., *The Conflict Catalogue, Centre for Global Economic History*; *Brecke, Peter. 1999. Violent conflicts 1400 A.D. to the present in different regions of the world*, http: / / www.inta.gatech.edu / peter / PSS99_paper.html, Accessed on 16 Mar 18.

Crillo, P., Taleb, N. N., On the statistical properties and tail risk of violent conflicts. Physica A: Statistical Mechanics and its Applications, 452, 29-45 (2016).

Eastaway, R., and Wyndham, J., *Why Do Buses Come in Threes?* Wiley, New York, 1998.

Fischhoff, B., Atran, S., Fischhoff, N., *Counting Casualties: A framework for respectful, useful records.* Journal of Risk Uncertainty (2007) 34 1-19.

Freedman, Lawrence, *The Long Peace and nuclear deterrence,* Perspectives from the Engelsberg Seminar 2015. Edited by Almqvist, K., & Linktaler, A., Stockholm 2016.

Gaddis, J. L., *The Long Peace: Elements of stability in the postwar international system*. Int. Secur. 10, 99 – 142 (1986).

Gunther, N. J., *Of Buses and Bunching: Strangeness in the Queue,* TeamQuest White Paper, 2010 TeamQuest Corporation.

Hayes, B., Statistics of Deadly Quarrels. American Scientist, Vol. 90, No. 1, January-February, 2002, pp. 10-15.Nate Silver, The Signal and the Noise, Penguin Books, 2012.

HM Treasury, *Managing Risks to the Public* (2005).

Kleinrock, L., *Queueing Systems*, Vol.I: Theory Wiley, New York, 1975.

Pinker, Steven, *The Better Angels of Our Nature*, Penguin Books, 2011.

Richardson, L. F., *Statistics of Deadly Quarrels.* (Quincy Wright and C. C. Lienau, eds.) Pittsburgh: Boxwood Press, 1960.