Appendix 1. The Mars Definition of Conventional War

In this appendix we quote the key text defining conventional war for the Project Mars database (Lyall 2020).

“Conventional War: Armed combat between the military organizations of two or more belligerents engaged in direct battle that causes at least 500 battlefield fatalities over the duration of hostilities. A conventional war has several defining properties: (1) states field armies that are clearly demarcated (i.e. they are wearing uniforms); (2) these armies engage one another in direct battle with clearly recognizable front lines; (3) these armies exhibit basic levels of military specialization, including possessing infantry (foot soldiers equipped with firearms), cavalry (mobile units, typically on horse back or, later, utilizing vehicles with internal combustion engines), and artillery (military branch devoted to the use of projectile weapons for indirect fire).

Belligerent: A political entity that claims control over, and authority within, a defined territory and population, and that can field a conventional army. To enter the dataset as a combatant, a state must have the following traits: (1) a political capital; (2) the ability to control and tax a fixed population; and (3) be able to muster a military in the immediate aftermath of a declaration of hostilities if a standing army does not exist. Moreover, a state must suffer at least 1% of a war's (or campaign's) overall casualties to be included or deploy at least 5% of the total forces. This excludes states that did not participate in the fighting (i.e. only declared war officially) or that played a minor role in the war. Note that civil wars that are fought along conventional lines are also included. That is, we do not require that combatants be officially recognized by Britain or France (as COW does) to be included. It is the ability to fight a certain way, not diplomatic recognition, that governs inclusion in the dataset.

Armies: In addition, states must possess armies with certain characteristics to be included. Specifically, these militaries must (1) possess some degree of combined arms via military specialization (infantry/cavalry/artillery or modern equivalents); (2) be able to supply the majority of its soldiers with firearms; (3) seek concealment from the enemy via terrain, not by blending into the civilian population; and (4) are built along direct battle lines, that is, to engage and impose their will violently on the opponent's military machine."

Appendix 2. New Decline-of-War Critiques and their Associated Doomsday Predictions

We go into considerable detail on the weaknesses of the COW data, particularly the COW Inter-State data, in sections 2 and 4 of the paper. These issues pose a special challenge to the work of (Clauset 2018) and (Braumoeller 2019), which rely heavily on the COW Interstate data, so we took our critique of their work no further in the body of the paper. In this appendix, we round out our critiques by highlighting some under-appreciated problems with statistical approaches which apply even if we accept the COW data. Along the way we find that many fat-tailed distributions fit war-size distributions as well as power laws do, i.e., that there is no reason to privilege power laws in our decline-of-war discussions.

The case that (Clauset 2018) and (Braumoeller 2019) make against the decline-of-war thesis is, to a large extent, based on non-rejection of a hypothesis that a single power law fits the distribution of Inter-State war sizes for both 1816-2007 and 1946-2007.[1] The first limitation of what we will call the "unchanging power-law argument" is that the evidence for a power law in COW war sizes is weak; many other distributions are just as plausible as a power law.[2] For example, (Clauset 2018) fits a power law with an exponent of -1.53 that applies above a minimum war size of 7,061. We reran this computation and nearly replicated the result, obtaining an exponent of -1.51 that applies above a minimum war size of 6,525, using both the poweRlaw package (Gillespie 2020) and plfit, which separately implement the methods of (Clauset, Shalizi, and Newman 2009). Broadly consistent with the findings of (Clauset 2018), we estimate a p value of 0.63 for the Kolmogorov-

Smirnov test of the hypothesis that the data was generated by the estimated power law, i.e, the estimated power law is not rejected by the data. However, other fat-tailed distributions can easily be mistaken for power laws (Stumpf and Porter 2012) so (Clauset, Shalizi, and Newman 2009) recommend testing power-law estimates against alternative fat-tailed distributions such as lognormals and exponentials. We follow this advice and find that these tests do not give pride of place to the power law. Indeed, the test of (Vuong 1989), as implemented in the poweRlaw package, finds the power law and the lognormal to be equally plausible (Z value of -0.33) although the power law does come out weakly preferred to the exponential (Z value of 1.6).

These distributions are but three out of many types of fat-tailed distributions so we used the GAMLSS package of (Stasinopoulos et al. 2017) to explore a wide range of possibilities.[3] A generalized gamma distribution turned out to be the best fit to the Inter-state COW data with an AIC statistic of the 2,140 while the power law and lognormal ranked 11th and 13th respectively, with AIC scores of 2,174 and 2,188, respectively.[4] These comparisons show, at a minimum, that there is no statistical reason to treat a power law as sacrosanct over other distributions. Indeed, the advice of (Burnham, Anderson, and Huyvaert 2011) is to dismiss the power law on the basis that its AIC exceeds that of the generalized gamma by more than 14. Panel a of figure 1 visually demonstrates this comparison; the power law (tail) does not stand out as, somehow, the “correct” fit.[5] To summarize, the argument for an unchanging power law, pre- versus post-World War 2, is undermined to the extent that COW war sizes may not even be well modelled by a power law in the first place.

The findings in the previous paragraph led us to discover that evidence singling out a power law in war sizes is also weak in the Mars data. The Vuong test renders the power law and lognormal to be equally plausible for both the low and high estimates of KIA counts, with Z values of 0.16 and 0.38 respectively. The power law is weakly preferred to the exponential with Z values of 1.50 for both the low and the high estimates of KIA. The GAMLSS package does place the power-laws fits for the low and high KIA estimates at ranks 6 and 5, respectively. The AIC scores for these power laws exceed the AIC minimizing fits by 4.2 and 7.6 for the low and high estimates of KIA, respectively, rendering the power-law fits to be plausible according to (Burnham, Anderson, and Huyvaert 2011). Nevertheless, several alternatives are at least as plausible as these power laws.

Figure 1A

A second limitation of the unchanging power law argument is its assumption that this one curve governs the distribution of war sizes far larger than all post-1945 data. Panel b of figure 1 shows the same power law, lognormal and generalized gamma distributions that are displayed in panel a but with only the post-1945 COW Inter-State data points. The three extrapolations beyond the last post-1945 data point differ greatly from one another. The range beyond the data accounts for an important piece of the story of panel b yet the logged X axis obscures the sheer boldness of this extrapolation. Panel c displays the same three curves in natural units on the X axis together with the same post-1945 COW battle death counts. Again, the three curves look like plausible fits to the post-1945 data points. But their extensions into the vast space beyond the data now look almost arbitrary. Moreover, the panels of figure 1 show merely three possible ways to describe a relationship about which we have no actual post-1945 data. Many slower and faster approaches toward 0 are roughly as plausible as the ones displayed in the graph, at least in the absence of some further analysis beyond simple curve fitting. In short, it could be useful for some purposes to model the data as fitting a single power law within some range (e.g., to measure the heaviness of the tail) just as one might use an exponential distribution if the mean is believed to be positive and well specified, but it is an unjustified leap to assume that this curve has predictive value through an order of magnitude beyond the range of the post-1945 data.

Third, even if we assume that CoW war sizes do follow a power law an underappreciated problem remains; the standard estimation methods can easily fail to detect very substantial changes in the power law exponent. To illuminate this issue, we generated 200 samples of 50 draws from power law distributions with exponents of 1.5, 1.6, 1.7, 1.8 and 1.9 (so 50,000 draws), all with xmin set to 1,000. The sample sizes of 50 are generous to the CoW-based post-WW2 estimates since there are 47 COW data points above 1,000 and only 17 of these are above the estimated xmin. While interpreting the results presented in table 1A, bear in mind that (Clauset 2018) estimated a power law exponent of 1.53 with a 95% confidence interval running from 1.37 to 1.76. Note, for example, that when the true power law exponent is 1.7 our bootstrapped probability of getting an estimated exponent below 1.7 or 1.6 is 0.41 and 0.15, respectively. Most researchers would interpret either outcome, especially the latter, as demonstrating that this estimated exponent is statistically indistinguishable from 1.53. But, in the tails, the difference between exponents of 1.53 and 1.7 is massive. Table 1A shows that draws above 100,000, 1 million and 10 million are, respectively roughly 2.4, 4 and 8.7 times as frequent when the exponent is 1.5 than it is when the exponent is 1.7. Worse, power laws with exponents of 1.5 and 1.8 can easily be declared to be statistically indistinguishable from one another even though the relative frequency of draws above 100,000, 1 million and 10 million, respectively, are 3.2, 6 and 18.9 times more frequent in the former case compared to the latter. In short, the unchanging power law argument would fail even if it were safe to assume that war sizes follow power laws because post-WW2 sample sizes are too small to offer a reasonable chance of detecting even very big changes.

Table 1A

Finally, consider the following doomsday warning which provides a convenient opportunity to summarize much of the present paper so far:

> “When I sat down to write this conclusion, I briefly considered typing, ‘We’re all going to die,’ and leaving it at that. I chose to write more, not because that conclusion is too alarmist, but because it’s not specific enough….
>
> …If the parameters that govern the mechanism by which wars escalate hasn’t (sic) changed—and there’s no evidence to indicate that they have—it’s not at all unlikely that another war that would surpass the two World Wars in lethality

> will happen in your lifetime. And if it is bigger than the two World Wars, it could easily be a *lot* bigger."[6] (Braumoeller 2019)

This warning may be offered more as a spur to action than as a serious prediction, but it should not be a compelling prediction because it relies on the unreliable and Eurocentric COW Inter-State war data, assumes that war sizes are truly governed by a power law that continues to apply up to arbitrarily large sizes, including those more than an order of magnitude beyond all post-1945 data, and implicitly exaggerates the extent to which the methods used would detect relevant changes in power laws even if such power laws actually apply. Of course, all prediction about the future must extrapolate beyond known data and there is no bright line that separates reasonable from unreasonable extrapolation. We suggest, nevertheless, that the above doomsday prediction has crossed the line.

We close this appendix by reemphasizing the finding, important in its own right, that there is no special connection between power laws and the distribution of war sizes; many other distributions appear at least as plausible as power laws. For some analyses it could be useful to proceed as if war sizes obey a power law since the war size distribution does have a fat tail and power laws are both relatively simple to work with and aim to directly measure the tail. However, when one extrapolates far beyond existing data, as has been sometimes done in the decline-of-war debate, then the power-law assumption is not merely a convenient stylization of the data but, rather, the main driver of all predictions. Other plausible assumptions for extrapolating the data lead to dramatically different predictions.

Figure 1A - Three Fits to CoW Interstate Battle Death Counts

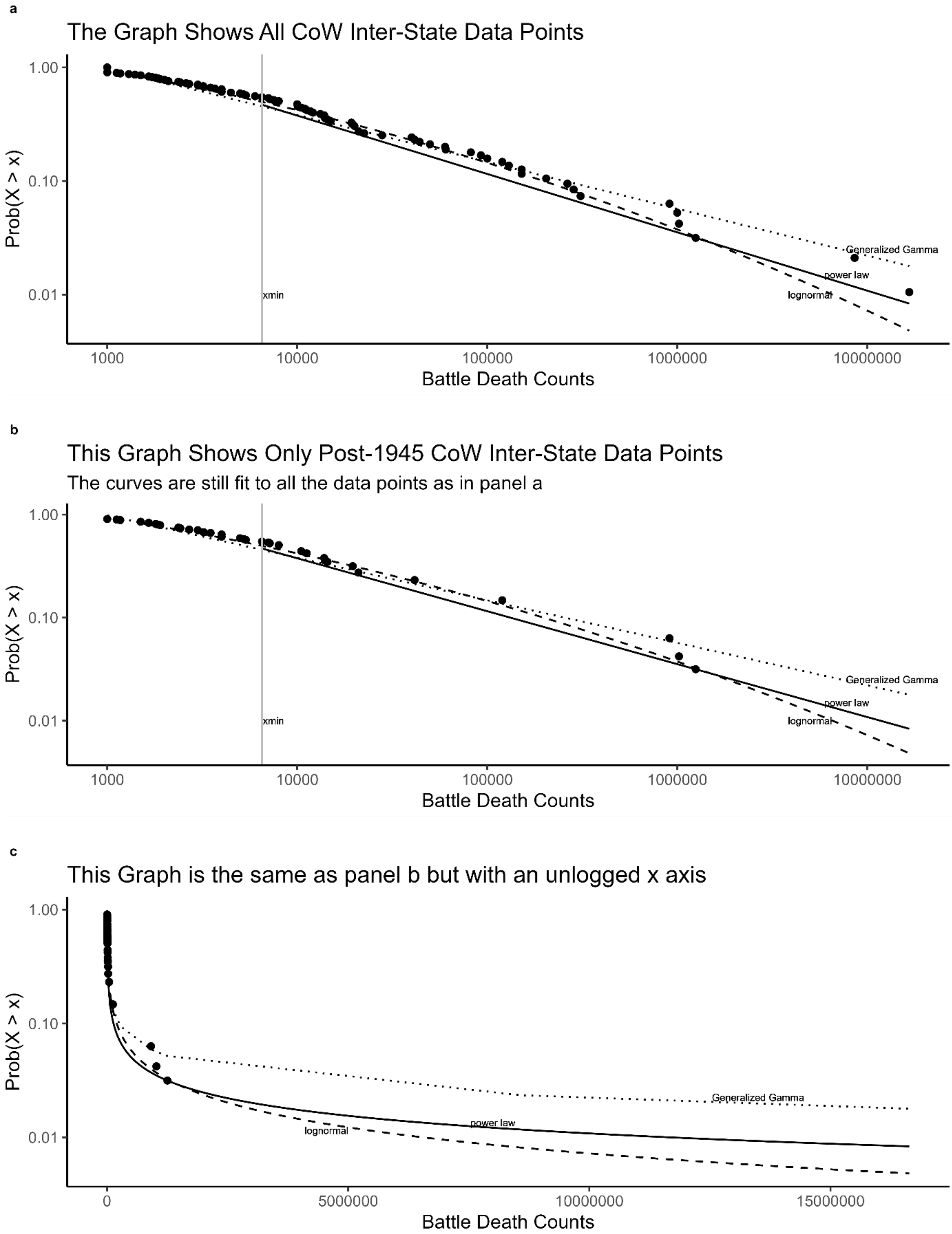

Table 1A - Power Law Estimates can Easily Overestimate Tail Probabilities Substantially

	True Alphas			
	1.6	1.7	1.8	1.9
Probabilities of Underestimating Alpha				
P(alpha estimate < 1.6)	0.41	0.15	0.06	0.00
P(alpha estimate < 1.7)	0.76	0.41	0.21	0.04
Relative Tail Probabilities				
P(x > 100,000|alpha = 1.5)/P(x > 100,000)	1.57	2.38	3.18	6.01
P(x > 1,000,000|alpha = 1.5)/P(x > 1,000,000)	2.06	3.92	6.02	14.13
P(x > 10,000,000|alpha = 1.5)/P(x > 10,000,000)	4.19	8.69	18.83	22.60

Simulated Data

Appendix 3. R Code for the Paper, including the Appendix

There are five R files underpinning the paper and appendix.

"Data Prep.R" moulds the original data, Mars, COW, some simulations and world population figures, into files that are ready for analysis and for the production of figures and graphs. The output of the data prep file is several .csv files that are then used in the other R files underpinning the paper. Some of the preparation is computationally intensive and we do not recommend running all of it without substantial computational power. We do, however, provide all the .csv files created by "Data Prep.R" so one can run the rest of the code without running the data prep script. Of course, it's also useful to examine the code even without running it.

Graphs and Tables.R is the code for all graphs and tables in both the body of the paper and the appendix except for figures 6 and 7. It uses the processed data produced by Data Prep.R. The remaining two figures are in figures6&7.R.

Calculations.R also uses the processed data and generates all calculations in the paper that are not embedded in a table or graph.

Finally, app.R is the code underlying the Shiny application that extends the Bayesian analysis in the paper. It uses the (slim version of the) Statistical Rethinking package (McElreath 2021) in order to showcase what happens when the priors and data analysed change. It was also used to generate Figure 6 in the main text and inspired Figure 7. The application is posted at:

https://erniethecat.shinyapps.io/After_the_Hemoclysim_app/

Bibliography for all the Appendices

Braumoeller, Bear F. 2019. *Only the Dead: The Persistence of War in the Modern Age*. New York, NY: OUP USA.

Burnham, Kenneth P., David R. Anderson, and Kathryn P. Huyvaert. 2011. 'AIC Model Selection and Multimodel Inference in Behavioral Ecology: Some Background, Observations, and Comparisons'. *Behavioral Ecology and Sociobiology* 65 (1): 23–35. https://doi.org/10.1007/s00265-010-1029-6.

Clauset, Aaron. 2018. 'Trends and Fluctuations in the Severity of Interstate Wars'. *Science Advances* 4 (2): eaao3580. https://doi.org/10.1126/sciadv.aao3580.

Clauset, Aaron, Cosma Rohilla Shalizi, and M. E. J. Newman. 2009. 'Power-Law Distributions in Empirical Data'. *SIAM Review* 51 (4): 661–703. https://doi.org/10.1137/070710111.

Cunen, Céline, Nils Lid Hjort, and Håvard Mokleiv Nygård. 2020. 'Statistical Sightings of Better Angels: Analysing the Distribution of Battle-Deaths in Interstate Conflict over Time'. *Journal of Peace Research* 57 (2): 221–34. https://doi.org/10.1177/0022343319896843.

Fagan, Brennen T., Marina I. Knight, Niall J. MacKay, and A. Jamie Wood. 2020. 'Change Point Analysis of Historical Battle Deaths'. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 183 (3): 909–33. https://doi.org/10.1111/rssa.12578.

Gillespie, Colin. 2020. 'PoweRlaw: Analysis of Heavy Tailed Distributions'. https://CRAN.R-project.org/package=poweRlaw.

Lyall, Jason. 2020. 'Replication Data for: Divided Armies'. Harvard Dataverse. https://doi.org/10.7910/DVN/DUO7IE.

McElreath, Richard. 2021. *Rethinking: Statistical Rethinking Book Package*. https://github.com/rmcelreath/rethinking.

Shmueli, Galit. 2010. 'To Explain or to Predict?' *Statistical Science* 25 (3): 289–310. https://doi.org/10.1214/10-STS330.

Spagat, Michael, and Stijn van Weezel. 2020. 'The Decline of War Since 1950: New Evidence'. In *Lewis Fry Richardson: His Intellectual Legacy and Influence in the Social Sciences*, edited by Nils Petter Gleditsch, 129–42. Pioneers in Arts, Humanities, Science, Engineering, Practice. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-31589-4_11.

Stasinopoulos, Mikis D., Robert A. Rigby, Gillian Z. Heller, Vlasios Voudouris, and Fernanda De Bastiani. 2017. *Flexible Regression and Smoothing: Using GAMLSS in R*. 1st edition. Chapman and Hall/CRC.

Stumpf, Michael P. H., and Mason A. Porter. 2012. 'Critical Truths About Power Laws'. *Science* 335 (6069): 665–66. https://doi.org/10.1126/science.1216142.

Vuong, Quang H. 1989. 'Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses'. *Econometrica* 57 (2): 307–33. https://doi.org/10.2307/1912557.

Zwetsloot, Remco. 2018. 'Testing Richardson's Law: A (Cautionary) Note on Power Laws in Violence Data'. *Available at SSRN 3112804*.

[1](Clauset 2018) divides the data at 1940, rather than 1945, but this difference does not matter since COW Inter-State has only one war with just 1,400 battle deaths between 1940 and 1945. Note that (Braumoeller 2019) emphasizes a war-size measure he calls “intensity”, i.e, battle deaths divided by the country’s population whereas (Clauset 2018) works with the raw battle death numbers. In this section, we work with just raw battle deaths but the ideas here are applicable to both measures.

[2] (Zwetsloot 2018) was an early sceptical paper about the use of power laws in violence data.

[3] Note that the poweRlaw package fits a distribution to just the tail of the data whereas GAMLSS seeks a distribution that best describes the entirety of the data.

[4] The BIC and likelihood ratio tests are model selection alternatives to AIC minimization. However, for prediction AIC minimization is generally recommended (Shmueli 2010), making it especially relevant in the present context.

[5] As a robustness check, we dropped the wars with the 10 lowest battle death counts, all 1,000, and re-estimated. Again, the model with the minimum AIC was the generalized gamma. The power law was 12th, trailing more than 18 points behind, so still recommended for dismissal by (Burnham, Anderson, and Huyvaert 2011).

[6] Note that the failure of (Braumoeller 2019) to reject a hypothesis of no change in a power law coefficient does not imply that there is no evidence of a change in war escalation mechanisms and, indeed, (Cunen, Hjort, and Nygård 2020), (Fagan et al. 2020) and (Spagat and van Weezel 2020) all found such evidence.