

Common ecology quantifies human insurgency

Juan Camilo Bohorquez¹, Sean Gourley², Alexander R. Dixon³, Michael Spagat⁴ & Neil F. Johnson²

¹*Department of Industrial Engineering and CEIBA Complex Systems Research Center, Universidad de Los Andes, Bogota, Colombia*

²*Complex Systems Group, Physics Department, University of Miami, FL 33126, U.S.A.*

³*Cavendish Laboratory, Cambridge University, Cambridge CB3 0HE, U.K.*

⁴*Department of Economics, Royal Holloway College, TW20 0EX U.K.*

Many collective human activities, including violence, have been shown to exhibit universal patterns¹⁻¹⁹. The size distributions of casualties in both world wars 1816-1980 and terrorist attacks have separately been shown to follow approximate power-law distributions^{6, 7, 9, 10}. However, the possibility of universal patterns ranging across wars in the size distribution or timing of within-conflict events has barely been explored. Here we show that the sizes and timing of violent events within different insurgent conflicts exhibit remarkable similarities. We propose a unified model of human insurgency which reproduces these commonalities, and explains conflict-specific variations quantitatively in terms of underlying rules-of-engagement. Our model treats each insurgent population as an ecology of dynamically evolving, self-organized groups following common decision-making processes. Our model is consistent with several recent hypotheses concerning modern insurgency¹⁸⁻²⁰, is robust to many generalizations²¹, and establishes a quantitative connection between human insurgency, global terrorism¹⁰ and ecology^{13-17, 22, 23}. Its similarity to financial market models²⁴⁻²⁶ provides a surprising link between violent and non-violent forms of human behaviour.

The political scientist Spirling²⁷ and others^{9, 10} have correctly warned that finding common statistical distributions (e.g. power-laws) in sociological data is not the same as understanding their origin. Possible political, ideological, cultural, historical and geographical influences make conflict arguably one the ‘messiest’ of all human activities to analyse. Mindful of these challenges, yet inspired by recent studies of human dynamics^{1–11, 17, 28}, we analyze the size and timing of 54,679 violent events reported within nine diverse insurgent conflicts, placing equal emphasis on both finding and modelling common patterns. Such insurgencies typify the future wars and threats faced by society^{18, 19}.

Our data sources are real-time media databases, official (government and nongovernmental organization) reports, and academic studies. Supplementary Information (SI) provides details, plus dataset extracts. The event data from different conflicts was compiled by different researchers, often with cross-checking by independent research teams, thereby reducing systematic collection or recording biases. Comparison of event accounts across a wide range of sources^{12, 29} reduces potential media bias and mistaken aggregation (e.g. misreporting two events of size x_1 and x_2 as one event of size $x_3 = x_1 + x_2$) which would create significant errors in a tail-dependent estimate such as a power-law slope. We focus on measuring deaths, since injuries are harder to cross-check – but we check where possible that our conclusions are robust to the inclusion of injuries.

Figures 1 and 2 show our empirical findings for event size while Fig. 3 shows event timings. Our model (described later and shown schematically in Fig. 4) provides a quantitative explanation of these findings by treating the insurgent population as an ecology of dynamically evolving, decision-making groups, in line with several recent sociological hypotheses^{18–20}. In addition to explaining the ubiquity of approximate power-laws in the event size distribution and the apparent central role of the 2.5

exponent value (Fig. 1), it explains the conflict-dependent deviations *beyond* a power law (see green curves in Fig. 2). Furthermore, the same model framework also explains the common burstiness in the distribution of event timings that we observe across insurgent conflicts (see black curves in Fig. 3).

Figure 1 gives exponent values, obtained by applying Clauset et al's^{9,10} established methodology for estimating discrete power-law distributions, $p(x) \sim x^{-\alpha}$ for $x \geq x_{min}$ where x_{min} is estimated along with α . In all cases we cannot reject the hypothesis that the size distribution of the events follows a power law, but we can reject log-normality. Four detailed examples are shown in Fig. 2. Following our preliminary 2005 results for Iraq and Colombia, we had suggested¹² that other insurgent wars might be clustered around $\alpha = 2.5$. All the insurgent wars that we have analyzed support this hypothesis. By contrast, we find that the Spanish Civil War and the American Civil War – neither of which are considered insurgent – each give distributions where log-normal can *not* be rejected, and where even the best-fit α value is much smaller (near 1.7, which is the value for the aggregated sizes of conventional wars⁹). This finding provides quantitative support for claims circulating in the social science field^{18,19} that insurgent wars represent ‘open-source’¹⁸, ‘fourth-generation’¹⁹ warfare, with qualitatively different dynamics from traditional wars. Several trivial explanations of the data can be ruled out, such as proportionality to city size¹⁰.

Figure 3 demonstrates a common burstiness in the distribution for the number of events per day, n , irrespective of size. As explained in Methods and the SI, we compare the distributions over daily event counts for different epochs within the four modern conflicts for which we have such data, against control distributions (‘random war’) obtained by randomizing event occurrences within each epoch. The data for each conflict (green circles) deviates from its random war (dashed curve) in a similar way: the real war exhibits an over-abundance of light days (i.e. days with few attacks) and of

heavy days (i.e. days with many attacks), but a ‘lack’ of medium days as compared to the random war (see lower panel). By considering subsets of days, we have determined that these features are not just an artefact of a variation in attack volume across days of the week (e.g. Fridays, see SI). Interestingly, this burstiness has become more pronounced over time for the wars in both Iraq and Colombia, suggesting that they have become less random as they have evolved. These findings are insensitive to the precise specification of the epochs within a given conflict.

Our model framework in Fig. 4 incorporates two key features: (1) ongoing group dynamics within the insurgent population (e.g. as a result of internal interactions and/or the presence of an opposing entity such as a state army); (2) group decision-making about when to attack based on competition for media attention. Within this framework, we find that mechanism (1) dictates the features observed in Figs. 1 and 2 while mechanism (2) dictates those in Fig. 3. Mechanism (1) is consistent with recent work on human group dynamics in everyday environments²⁸, and with current views of modern insurgencies as fragmented, transient, and evolving¹⁸. Mechanism (2) is consistent with comments by former U.S. Senior Counterinsurgency Adviser David Kilcullen, who noted²⁰ that when insurgents ambush an American convoy in Iraq: “...they’re not doing that because they want to reduce the number of Humvees we have in Iraq by one. They’re doing it because they want spectacular media footage of a burning Humvee.” We consider the insurgent population as having an overall strength N made up from human combatants, information, resources and weapons – though for simplicity, one can think of N humans. N is continually being re-partitioned through coalescence and fragmentation processes, thereby producing an ecology of groups. A group’s strength at timestep t determines the number of human casualties x it would produce if it decided to engage in an event at that timestep. We take N to be approximately constant over time, though our main conclusions are unchanged if N evolves slowly with small fluctuations.

These two co-existing dynamical mechanisms generate rich time-series which can explain the numbers of events of different sizes at each timestep. However since the data in Figs. 1 and 2 is time-aggregated while that in Fig. 3 is size-aggregated, we can provide far more insightful explanations using simplified versions which treat the respective non-dominant mechanism in an averaged way. Consider first the simple situation in which the group coalescence and fragmentation processes in the insurgent population are represented by probabilities¹⁸. The fragmentation probability v_{frag} is taken to be small ($\sim 1\%$) to mimic the infrequent scenario in which a group member suddenly senses imminent danger and the entire group scatters. If fragmentation does not occur, the group may coalesce with another group with probability v_{coal} . This mimics the scenario in which two individuals initiate a communications link between them of arbitrary range (e.g. mobile phone call), and hence their respective groups of strength s_1 and s_2 act in a coordinated way with strength $(s_1 + s_2)$. Since these two processes can be triggered by any particular constituent group member at any time, the probability that it affects a specific group should be proportional to s ²⁴⁻²⁶. Treating mechanism (2) in an averaged way, we assume that all groups are equally likely to be involved in an event over time. This is consistent with the time-averaged behaviour of the full decision-making model (see later). The time-averaged distribution of group strengths s therefore acts like the distribution of event sizes x ¹², resulting in a steady-state approximate power-law distribution whose analytic solution $\alpha = 2.5$ ^{25,26} is within the empirical bounds of Clauset et al.'s total value 2.48 ± 0.07 for global terrorism¹⁰. This analytically-obtained theoretical value^{25,26} 2.5 is robust to many model generalizations^{21,25,26} (e.g. coalescence of multiple groups, fragmentation into groups larger than one) thereby offering an explanation for the observed bunching of the empirical values around 2.5 in Fig. 1.

Invoking a more realistic mechanism for grouping dynamics than simple probabilities (see SI), we find that our model framework can explain not only the

approximate power-law behaviour and central role of the 2.5 exponent (Fig. 1) but also the behaviours *beyond* power-law observed in Fig. 2. Accounting explicitly for an opposing population (e.g. state army) with total strength N_B , the coalescence and fragmentation is now caused by the interactions between groups. The casualties produced by clashes between opposing groups can then be used to obtain the event size distributions (green curves in Fig. 2). Full details are given in the Methods section and the SI, along with the four model parameters for each conflict (total insurgent strength N_A , total state strength N_B and casualty scales C_S and C_L). When two opposing groups meet they fight, with some members of both groups killed and the smaller group fragmenting. C_S (C_L) sets the scale for the number of the smaller (larger) groups members destroyed. As C_S is increased, the model deviates increasingly from a straight line at low x , suggesting that Afghanistan and Colombia share the following similarity: in a clash in which the insurgent group is the smaller group, this insurgent group takes heavier relative losses than for the wars in Iraq and Peru. The ratio between the two populations' strengths (N_A and N_B) tends to control the slope itself, with greater strength differences resulting in steeper slopes. This suggests that there might be a greater difference between the strengths of the army and insurgency in Colombia than in Iraq or Afghanistan. The total insurgent strength N_A controls the large x roll-off in Fig. 2. Afghanistan and Peru deviate substantially from power laws for large x , which our model interprets as relatively small insurgency strength. Colombia and Iraq hardly deviate from power laws for large x , implying greater insurgency strength.

Since Fig. 3 features data aggregated over size, we replace the detailed grouping dynamics (i.e. mechanism (1)) by a time-averaged number of groups. Given the resolution of our data and the typical numbers of observed daily attacks, we take one timestep as equivalent to one day. If a group launches an attack during a day with many other attacks, its media coverage will in general be reduced. If instead it launches an attack on a quiet day, its media coverage will increase²⁰. Each group receives daily

some common but limited information (e.g. public radio or newspaper announcements about previous attacks, opposition troop movements, a specific religious holiday, even a shift in weather patterns). The actual content is unimportant provided that it becomes the primary input for the groups' decision-making process. (See Ref. 26 for a full description of an equivalent financial market version). Although the groups are heterogeneous in terms of their strategies, they tend to converge toward similar responses when fed the same information²⁶ thereby generating distributions (black curves) which are almost identical to those observed (green circles). Our model (see SI and Ref. 26) includes a confidence threshold which must be surpassed before any decision can be made, allowing us to interpret the increase in non-randomness over time for Iraq and Colombia as a decrease in this confidence threshold, i.e. the insurgent groups in both wars have become less cautious over time about whether to launch attacks. Reference 30 presents independent empirical evidence that groups of humans do indeed employ such generic decision-based mechanisms.

To our knowledge, our model provides the first unified explanation of high-frequency, intra-conflict data across human insurgencies. Other explanations of human insurgency are possible, though any competing theory would also need to replicate the results of Figs. 1, 2 and 3. Our model's specific mechanisms challenge traditional ideas of insurgency based on rigid hierarchies and networks, while its striking similarity to multi-agent financial market models²⁴⁻²⁶ hints at a possible link between collective human dynamics in violent and non-violent settings¹⁻¹⁹.

Methods Summary. For the event size distribution (Figs. 1, 2), we use Clauset et al's methodology^{9, 10} to estimate power-law exponents, and test power-law and lognormal hypotheses using the time-aggregated time-series of events. This methodology^{9, 10} is a widely accepted, published state-of-the-art statistical procedure for analyzing power-law-like distributions. For the event timing distribution (Fig. 3), we divide the time-

series for the number of events per day into epochs. These epochs are chosen such that there is no significant trend in the moving-average within each epoch. The precise specification of each epoch's time-window does not affect our main findings. We then generate 10,000 random wars by shuffling the date of the events within each section, averaging across the shuffles. Our model (Fig. 4) replicates the empirical size and timing patterns of Figs. 1, 2 and 3. Full details are given in Methods and SI.

METHODS

Power Law Fitting and Testing. We follow Clauset et al's^{9, 10} established methodology for estimating discrete power-law distributions. First, form the complementary cumulative distribution, $P(X \geq x)$, for event sizes within an individual conflict. For each possible choice of x_{min} perform a discrete maximum-likelihood estimation of the power-law exponent. Use Kolmogorov-Smirnov (KS) goodness-of-fit statistics to determine the best x_{min} . Estimate standard errors in estimated α 's with bootstrap resampling. Test the hypotheses that the data are generated by power laws using a goodness-of-fit-test. Check that a substantial fraction of deaths are within the power-law region $x \geq x_{min}$. The statistical significance of the power-law is always reasonable and x_{min} is reassuringly small (e.g. p -value = 0.74 and $x_{min} = 4$ for Colombia. See table 1 in SI).

Temporal Clustering Analysis. To test randomness in the timing of events we produce a 'random war' that looks like the real war but where the numbers of attacks on consecutive days are unrelated. We do this by shuffling the attacks and assigning them to random days. Since conflicts will often be non-stationary processes, we account for long-term trends by dividing the time-series into distinct epochs such that the moving-average within each epoch does not vary significantly. The choice of epochs does not affect our main results, and those shown in Fig. 3 are representative of a wide variety of such choices. Even within a given epoch, standard time series analysis gives weak and

noisy results for stationary measures such as event autocorrelation -- hence we do not employ these methods. The algorithm that we adopt is:

- 1) Take time-stamp of each attack
- 2) Plot observed number of attacks on each day of the conflict.
- 3) Split the time-series for the real war into smaller epochs to reduce effect of underlying long-term trends
- 4) Randomly reassign the events to new days within each epoch
- 5) Plot the shuffled number of attacks on each day to create a random war

This generates a random war that has the same total number of attacks and the same distribution of casualties as the real war. It also guarantees that any long-term (i.e. epoch to epoch) changes are the same for the random and real wars. We repeat this shuffling process to produce a large collection of random wars.

We then compare the frequency distributions over daily attack counts for the real wars against those for corresponding random wars. These are plotted in Fig. 3 for four wars. The horizontal axis represents the number of events per day or 'attack volume', n , and the vertical axis represents the frequency $p(n)$ of each n . The difference between the two is captured by the histogram directly under each panel. The bars in the histogram measure $\Delta(n)$, the difference between the real data and the random data divided by the standard deviation from the mean (see SI). Non-zero $\Delta(n)$ indicates short-term correlations between events.

Day-of-the-week effects. One possible candidate explanation for these short-term correlations may be that violence is dictated by the day of the week. Indeed, in Iraq on Fridays, the Muslim holy day, there are 15% fewer attacks. We therefore ran our

simulations against random wars both with and without specific days-of-the-week, and obtained almost identical results to Fig. 3. We conclude that Fig. 3's explanation is not a simple day-of-the-week effect.

Grouping Mechanism. Two separate populations of agents, of total size N_A and N_B , represent the insurgents and state army respectively. Each population is distributed in groups of various sizes. At each timestep, a group is selected at random out of the combined A and B population. This agent's group then interacts with a second group selected at random from the combined population (see SI). If the two groups are of the same population type they join together. If the two groups are from different populations, the smaller group experiences a random number of casualties up to parameter C_S times its size, while the larger group experiences a random number of casualties up to parameter C_L times the smaller group's size. The smaller group then fragments into randomly sized groups. If either population is below its target size (N_A or N_B) a new member is added to a randomly selected group in that population. The model evolves to a steady state. The number of casualties occurring in each interaction is recorded. This is averaged over 100,000 simulations to generate the mean distributions shown in Figure 2. The initial conditions of the model do not affect the steady state result, which depends only on N_A , N_B , C_S and C_L .

Attack Timing Mechanism. Suppose that the insurgency as a whole would like to create a constant level of daily attacks, which we call L . At each time-step (i.e. each day) a global signal (which is a proxy for the outcomes of previous days) is broadcast to all the groups. Each group reads the signal and takes the decision to attack or not, depending on the confidence of the group concerning their strategy and the success of this strategy as captured by p_r . If the confidence of the group is greater than the confidence threshold ($c_i \geq C$) then the groups decide to attack based on the probability p_r . The variable p_r is the probability that the group will react to the global information

in the same way it did last time that the signal was in a given state. Once the groups have decided their actions, the total number of groups that attacked is compared against the desired level L . Too much exposure (above L), and the groups wasted their attack. If few attacks are made, the groups receive plenty of attention and consider that they took the right decisions. The groups adapt their variables based on their success. The global signal is updated and the cycle is repeated.

1. Gabaix, X., Parameswaran G., Plerou, V. and Stanley, H.E. A Theory of Power Law Distributions in Financial Market Fluctuations. *Nature* **423** 267-270 (2003).
2. Lux, T. and Marchesi, M. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature* **397**, 498-500 (1999).
3. Dussutour, A., Fourcassie, V., Helbing, D. and Deneubourg, J.L. Optimal traffic organization in ants under crowded conditions. *Nature* **428**, 70-73 (2004).
4. Mantegna, R.N. and Stanley, H.E. Scaling behaviour in the dynamics of an economic index. *Nature* **376**, 46-49 (1995).
5. Barabasi, A.L. The origin of bursts and heavy tails in human dynamics. *Nature* **435**, 207-211 (2005).
6. Richardson, L.F. Variation of the Frequency of Fatal Quarrels with Magnitude. *J. Amer. Statistical Assoc.* **43**, 523-546 (1948).
7. Cederman, L-E. Modeling the Size of Wars: From Billiard Balls to Sandpiles. *Amer. Polit. Sci. Rev.* **97**, 135-150 (2003).
8. Lim, M., Metzler, R. and Bar-Yam, Y. Global Pattern Formation and Ethnic/Cultural Violence. *Science* **317**, 1540-1544 (2007).
9. Clauset, A., Shalizi, C.R., and Newman, M.E.J. Power-law distributions in empirical data. *SIAM Review*, in press (2009).

10. Clauset, A., Young, M., and Gleditsch, K.S. On the Frequency of Severe Terrorist Events. *J. Conflict Resolution* **51**, 58-87 (2007).
11. Galam, S. and Mauger, A. On reducing terrorism power: a hint from physics. *Physica A* **323**, 695-704 (2003).
12. Johnson, N. *et al.*, Universal patterns underlying ongoing wars and terrorism.. arXiv:physics/0605035v1 [physics.soc-ph] <http://arxiv.org/abs/physics/0605035> (2006).
13. Johnson, D.D.P. and Madin, J.S., *Population models and counter-insurgency strategies*. In *Natural Security: A Darwinian Approach to a Dangerous World*, edited by R. Sagarin and T. Taylor. University of California Press, Berkeley, (2009).
14. Flack, J., Security in an uncertain world. *Nature* **453**, 451-452 (2008).
15. Adams, E.S. and Mesterton-Gibbons, M., Lanchester's attrition models and fights among social animals. *Behavioral Ecology* **14**, 719-723 (2003).
16. Wrangham, R.W., War in evolutionary perspective. *Emerging Synthesis in Science*, Pines Ed., New Mexico, Santa Fe Institute (1985).
17. Epstein, J.M., Nonlinear dynamics, mathematical biology, and social science. Addison-Wesley, Reading MA (1997).
18. Robb, J., Brave new war: the next stage of terrorism and the end of globalization. Wiley, New York (2007).
19. Hammes, T. The sling and the stone: on war in the 21st century Zenith Press, St. Paul (2004).
20. Packer, G., Can social scientists redefine the 'war on terror'? *The New Yorker*, December 18, (2006)
21. Rusczycki, B., Burnett, B., Zhao, Z., Johnson, N.F., Relating the microscopic rules in coalescence-fragmentation models to the macroscopic cluster size distributions which emerge. arXiv:0808.0032v2 [cond-mat.other] <http://arxiv.org/abs/0808.0032> (2009)

22. May, R.M. and McLean, A.R., *Theoretical Ecology Principles and Applications*. 3rd edition. Oxford University Press, Oxford (2007).
23. Couzin, I.D., Krause, J., Franks, N.R. and Levin, S.A., Effective leadership and decision making in animal groups on the move. *Nature* **433**, 513-516 (2005).
24. Eguiluz, V.M. and Zimmermann, M.G., Transmission of Information and Herd Behaviour: An Application to Financial Markets. *Physical Review Letters* **85**, 5659-5662 (2000).
25. DHulst, R. and Rodgers, G.J. Exact solution of a model for crowding and information transmission in financial markets. *International Journal of Theoretical and Applied Finance* **3**, 609-616 (2000)
26. Johnson, N.F., Jefferies, P. and Hui, P.M. *Financial Market Complexity*. Oxford University Press, Oxford (2003)
27. Spirling, A. The next big thing: Scale invariance in political science. Harvard University, <http://www.people.fas.harvard.edu/~spirling/documents/powerlawSend.pdf>
28. Palla, G., Barabási, A.L., and Vicsek, T., Quantifying social group evolution. *Nature* **446**, 664-667 (2007).
29. Hicks, M., M.D., Dardagan, H., Guerrero, G., Bagnall, P., Sloboda, J.A., and Spagat, M. The Weapons That Kill Civilians. *New England Journal of Medicine* **360**, 1585-1588 (2009).
30. Wang, W., Chen, Y., and Huang, J. Heterogeneous preferences, decision-making capacity, and phase transitions in a complex adaptive system. *Proceedings National Academy of Science USA* **106**, 8423-8428 (2009).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to J. Restrepo for his involvement in the data collection, and R. Zarama for discussions.

Author Contributions J.C.B., S.G., M.S. and N.J. worked on the data and data analysis. S.G., A.D. and N.J. worked on the model development. All authors participated in the write up and associated discussions, giving detailed feedback in all areas of the project.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions.

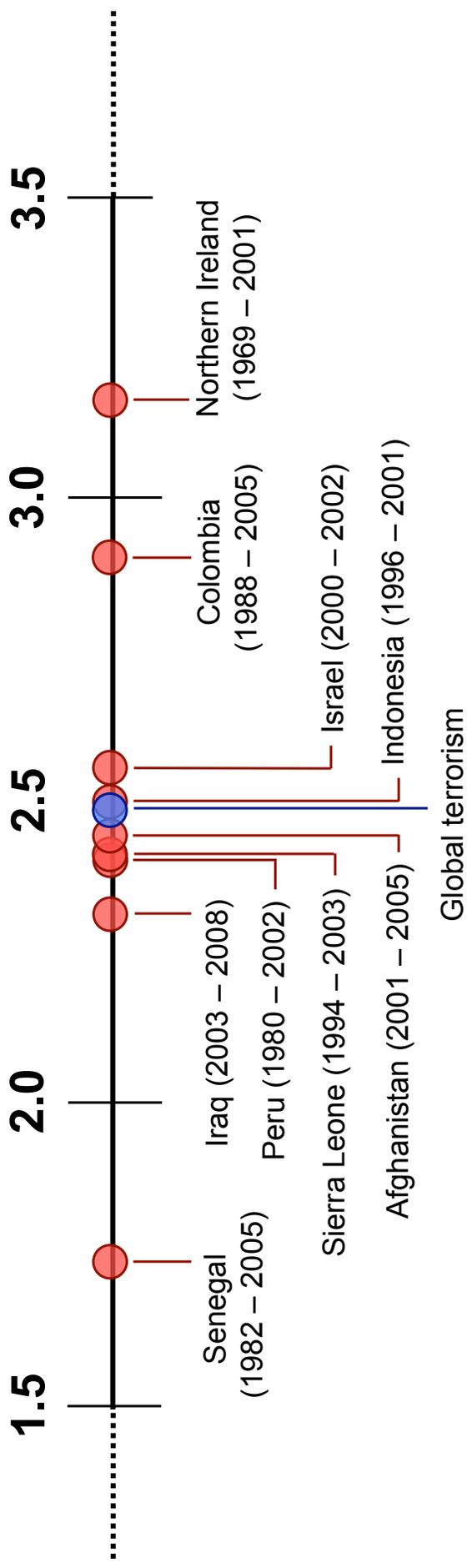
Correspondence and requests for materials should be addressed to N.F.J. (njohnson@physics.miami.edu).

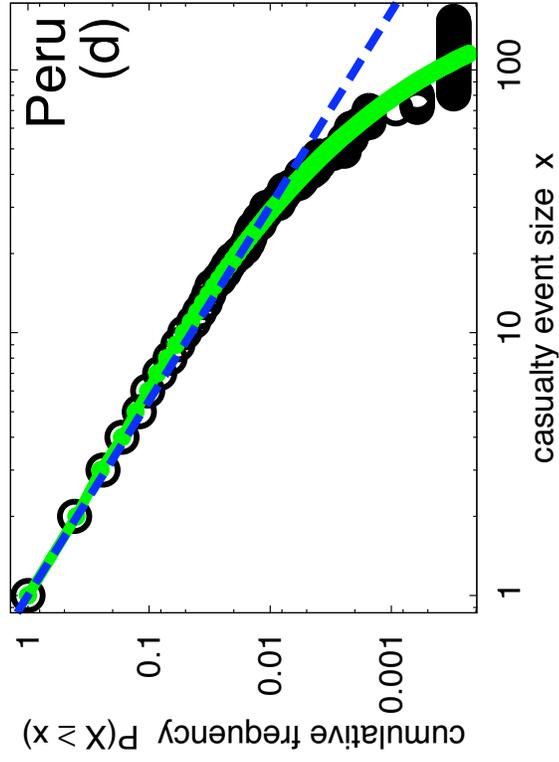
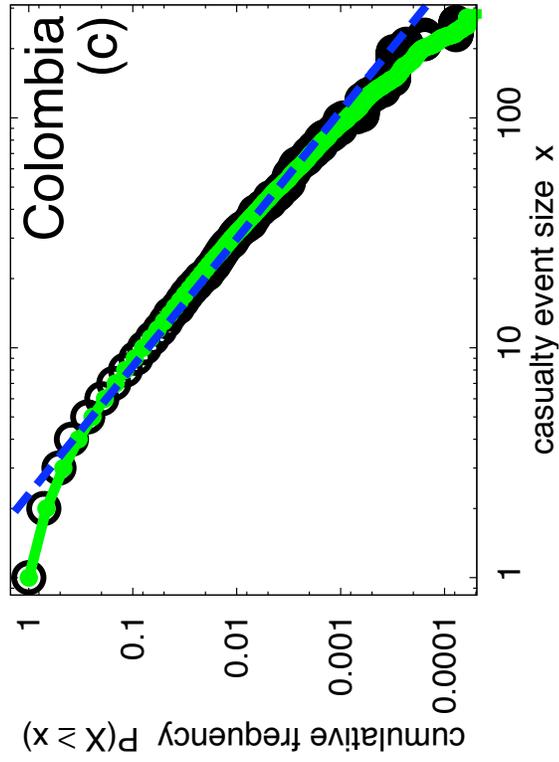
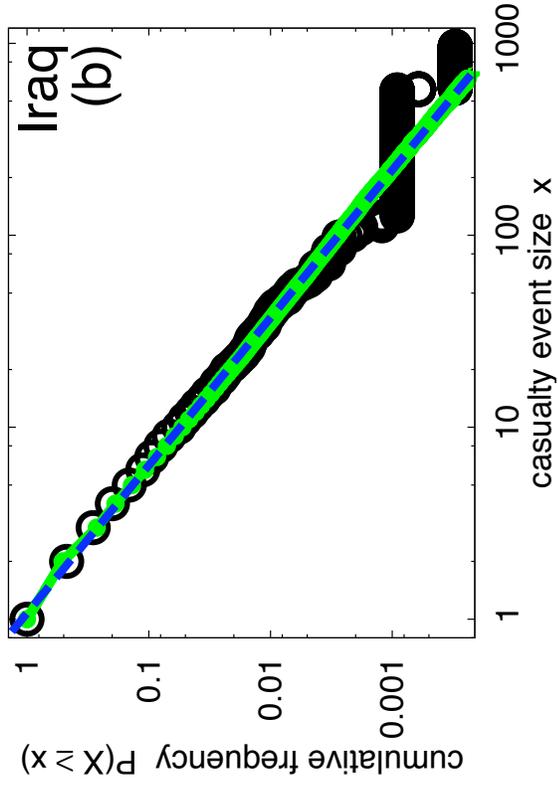
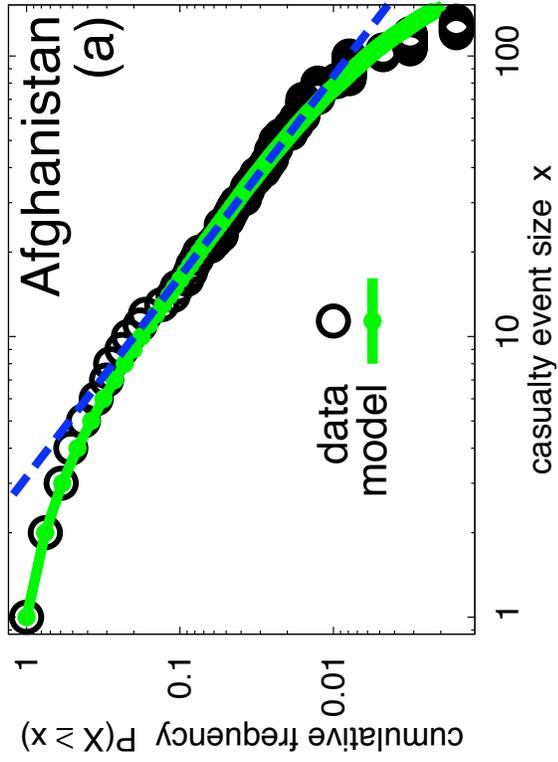
Figure 1 Power-law exponents. Value α for power-law $p(x) \sim x^{-\alpha}$ deduced from the empirical distributions of event size x (i.e. number of casualties) for insurgent conflicts. Statistical procedures follow Refs. 9 and 10. Blue dot shows the value 2.48 for distribution of total size of global terrorist events, from Clauset et al.¹⁰. The years in parentheses describe the empirical dataset range used to deduce α , not the actual conflict duration.

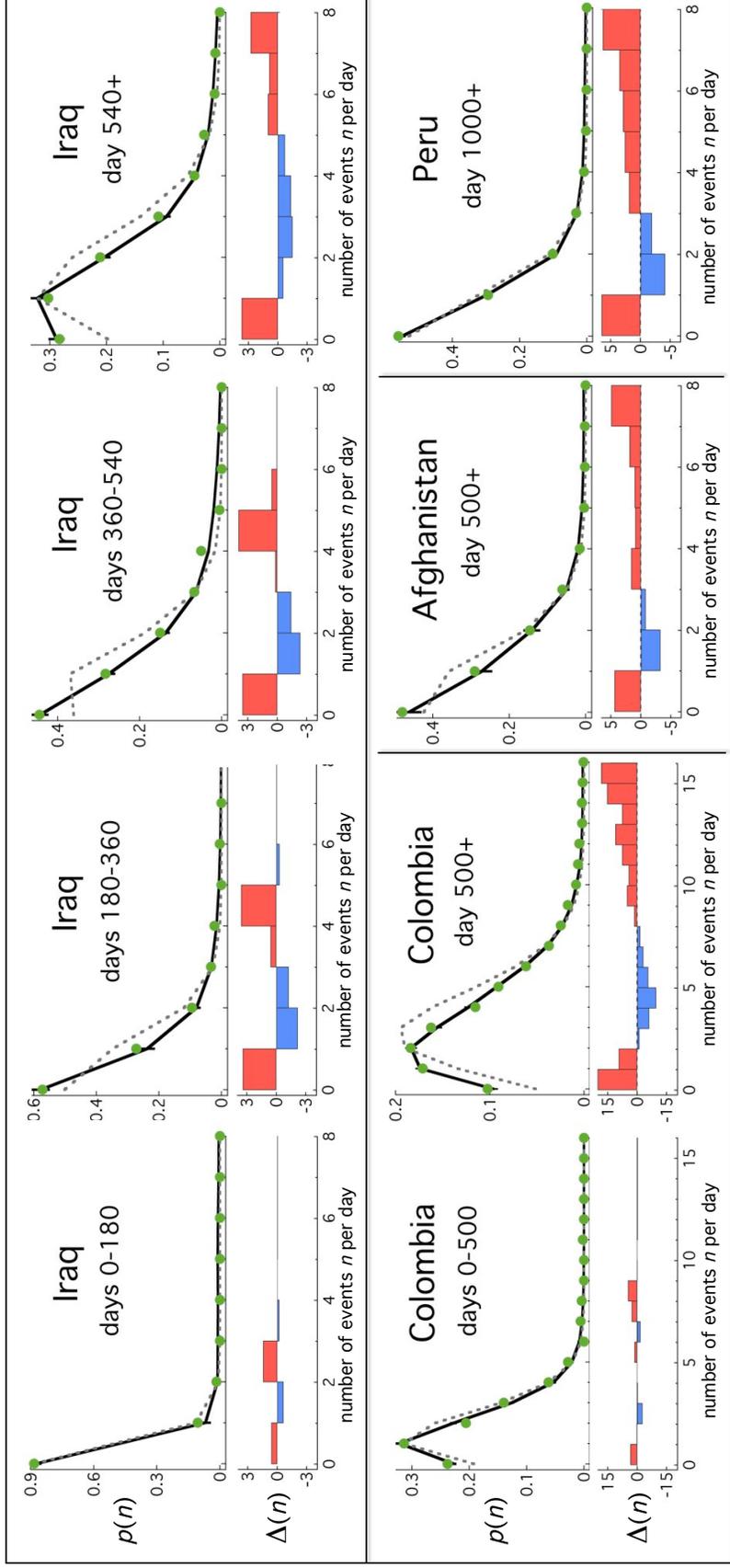
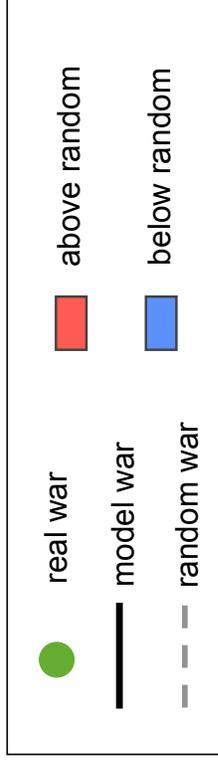
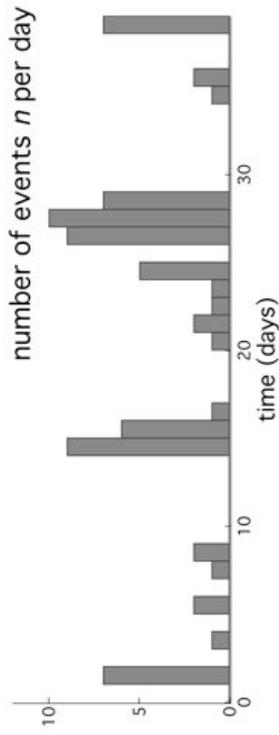
Figure 2 Size of events. Log-log plot of the complementary cumulative distribution of event size $P(X \geq x)$ (i.e. probability of an event of size greater than or equal to x) for four conflicts from Fig. 1. Horizontal axis shows event size x , i.e. number of casualties. Solid green curves show the results from our model. Blue dashed line is simply a straight line guide-to-the-eye, not a power-law fit.

Figure 3 Timing of events. Schematic shows a time-series with $n = 0, 1, 2, \dots$ events per day. Green circles show distribution $p(n)$ for number of days with n events in actual conflict. Dashed lines represent average values for random wars. Solid lines denote average distributions calculated from 10,000 realizations of our model (Fig. 4). Histograms below represent differences $\Delta(n)$ between real and random wars, in units of standard deviations from mean. Error bars for random wars, i.e. one standard deviation from mean of 10,000 shufflings, are shown but are small. Error bars for model wars demonstrate a small spread in run outcomes.

Figure 4 Model framework for insurgency. Insurgent population comprises an overall strength N , distributed into groups with diverse strengths at each timestep t . This distribution changes over time as groups join and break up. Dark shadows indicate strength, and hence casualties that can be inflicted in an event involving that group. Figs. 1 and 2 are derived from the number of events of size x aggregated over time. Figure 3 is derived from the number of events at a given timestep aggregated over size.







MODEL MECHANISM AT TIMESTEP t

