Are Modern Insurgencies Predictable? New Evidence from the Afghanistan and Iraq Wars

Andrew C. Shaver * and Austin L. Wright †

*Woodrow Wilson School of Public and International Affairs, Princeton University, Robertson Hall, Princeton, NJ 08540, and †Department of Politics, Princeton University, Corwin Hall, Princeton, NJ 08540

SOCIAL SCIENCES: Political Sciences

Substate conflict is the source tremendous human suffering and financial burden worldwide. Research recently published in Science claims to have uncovered predictable patterns in insurgent and terrorist attacks: like various unrelated physical phenomena, the rate of learning by insurgents (and thus rate of conflict escalation) follows a predictable power law process. Governments afflicted by insurgent violence might use such patterns to thwart some attacks. The authors do not, however, consider strategic interaction during conflict and derive their results using data that recent research indicates is likely biased. Using newly declassified, comprehensive, and methodically collected data on insurgent activities covering all years of the Afghanistan and Iraq wars supplied by the Pentagon, we demonstrate that these findings are an artifact of biased microdata. We further show that dynamic learning during conflict varies continually: at times, insurgents learn more quickly than their opponents; at others, they do not. Observable patterns of learning by insurgents in conflict environments are complex and reflect strategic imperatives. The results speak to the need for greater interdisciplinary collaboration. Parsimonious mathematical models used to describe physical phenomena are not clearly suited to modeling strategic political interactions. If physical or social scientists are to inform national security policies, they will likely need to do so collaboratively.

Significance statement

Substate conflict causes tremendous human suffering and financial burden worldwide. Research recently published in Science claims to have uncovered a predictable pattern in the escalation insurgent violence, which governments might use to thwart attacks. Using newly declassified, comprehensive data on insurgent activities during the Afghanistan and Iraq wars supplied by the Pentagon, we demonstrate that these findings are an artifact of biased micro-data and that dynamic learning varies continually: insurgents sometimes learn more quickly than their opponents; often they do not. Observable patterns of insurgent learning reflect strategic imperatives. Parsimonious models are not clearly suited for studying strategic interactions. If natural and social scientists are to inform national security policies, they will likely need to do so collaboratively.

Violent sub-state conflict is widespread. Rather than ushering in an era of peace and tranquility, as Francis Fukuyama famously predicted [7], the end of the Cold War has given way to a series of violent domestic conflicts fueled by secessionist, revolutionary, and related movements. In 2014, a majority of countries were targeted in violent attacks carried out by non-state actors. 78 states faced increased risk of internal violence over the previous year [2]. That same year, substate conflicts cost the world economy approximately 14.3 trillion dollars (13.4% of world GDP), a figure equivalent to the combined value of the economies of Brazil, Britain, Canada, France, Germany, and Spain [2].

Resulting human displacement alone has been overwhelming. One out of every 122 humans is now either a refugee, internally displaced, or seeking asylum. If this population were a nation, it would make up the 24th largest in the world [1]. While the well known cases of civil war in Syria and the reemergence of insurgency in Iraq have displaced 14 million people [3], conflicts from Colombia to the Democratic Republic of the Congo to Burundi have displaced many millions more [1].

This trend shows no signs of abating, and particular technological innovations raise the possibility that sub-state conflict will persist well into the 21st century. The proliferation of improvised explosive devices (IEDs), for instance, is one technological innovation favoring rebel organizations facing significantly stronger government competitors. As Singer (2012) explains, “[t]he IED has proved to be a cheap, relatively easy-to-use tool against both civilians and advanced militaries... [and] present[s] increasingly difficult balance-of-costs problems. It is unsustainable [for governments to invest] billions of dollars to fight a technology that costs the other side tens of dollars... The IED is not disappearing; rather it is proliferating...” [6].

In states where rebel organizations are active, what can government forces do to reestablish stability? According to Johnson et al., insurgents, including those who employ terrorist tactics, display predictable tendencies that may allow government forces to anticipate and prevent attacks [4]. As Yale political scientist and civil war scholar Jason Lyall notes, such a finding, if true, would represent “a major coup” [5].

Using highly detailed, comprehensive data on insurgent attacks from the recent conflicts in Afghanistan and Iraq supplied to us by the U.S. Defense Department for the purposes of this and related studies, we replicate [4]. We then study relative learning by insurgents and counterinsurgents. The only predictability we observe is that dynamic learning during conflict is fluid process that favors the insurgent at times and the counterinsurgent at others.

Power Law Processes and Insurgent Violence

A variety of physical and social phenomena can be modeled as power law functions. From “infrastructure networks such as those of airlines” [15] to the formation of alliance networks

Reserved for Publication Footnotes
in the U.S. biotechnology industry to protein-interaction network of cells [11], exponential growth and decay processes persist in numerous unrelated contexts. In particular, learning by individuals and organizations has been found to follow a power law process - “[f]or a wide range of human activities, the time taken to complete a given challenging task decreases with successive repetitions, following an approximate power law progress curve” [4]. Johnson et al. theorize that learning by insurgents and terrorists follows a power law progress curve. If these authors are correct, the rate at which non-state actors carry out attacks increases exponentially as conflicts progress.

[4] are not the first to theorize that particular patterns of organized violence adheres to a power law process. [12] find that “new attacks [by terrorist organizations] lead to organizational growth and [that] the corresponding increase in size leads to faster production of new events because a larger size means more terrorist cells are operating in parallel... The typical form of this relationship can be mathematically modeled by a power law function”. [14] find that the scale of insurgent attacks in a variety of conflict environments adheres to a power law process.

**Data Quality.** Our primary concern with the results reported by Johnson et al. is empirical and relates specifically to the data underlying their study. The authors use public datasets made available byicasualties.org, the Memorial Institute for the Prevention of Terrorism, and the University of Chicago’s Project on Security and Terrorism. Each of these datasets is based on media reports of insurgent and terrorism related activities.

Recent research demonstrates that conflict datasets assembled using media reports are susceptible to troubling patterns of non-random missingness and that their use in empirical research “can fundamentally alter the conclusion we draw” [22]. Specifically, he shows that during the Afghanistan conflict media reports of violence have correlated positively with cellphone coverage because “cellphone coverage makes the reporting of violence more likely” [22].

**Data Generating Process.** Relatedly, the observational data used by Johnson et al. likely do not approximate the data generating process during the initial years of the Afghanistan and Iraq conflicts. The insurgencies in both countries were not slow to develop. In Iraq, for instance, little more than one month after Coalition forces had invaded, insurgents began conducting regular attacks [25]. Yet, recognition that an insurgency was emerging was eclipsed by the widespread alternative view that such attacks were the work of “the remnants of the Baath regime and Fedayeen death squads [26] and “common criminals” [27].

Underreporting of insurgent attacks during earlier periods of both conflicts poses an acute problem for efforts to estimate whether the rate of conflict escalation by insurgents follows a power law process. Methods for identifying such relationships are heavily influenced by the frequency of attacks early in each conflict zone. Under [4]’s approach, the steep portion of the power law progress curve is estimated from the relatively small number of initially infrequent attacks. Yet, patterns of infrequent attack may simply reflect limited and inconsistent media coverage rather than slow insurgent learning.

**Absolute vs. Dynamic Learning.** Our final concern is theoretical. Previous findings that learning by individuals can be modeled as a power law process are not clearly applicable to actors learning during periods of strategic interaction. Specifically, learning by individuals frequently occurs in contexts in which conditions governing the rate of learning are static (with potential for random variation in such governing forces in some contexts). For instance, a child learning to ride a bicycle on a dry, summer day will almost certainly encounter the same bicycle design, gravitational forces, and rolling friction coefficient during each attempted ride. A separate class of learning takes place in dynamic settings where the behavior of one actor influences the subsequent behavior of another [18]. Insurgency is one such example [16]. Insurgents attempting to outmaneuver counterinsurgents may modify their behavior based on what they have learned about counterinsurgents; but, yet, attentive counterinsurgents may note such adaptation and change their own tactics in turn.

[4] model this process as a correlated walk, which is applicable to settings where the direction of adaptation remains constant and only the pace of movement shifts. Yet the escalation of insurgent violence is unlikely to continuously increase. Instead, as a number of historical cases reveal (e.g., Oman, Mozambique, Guinea-Bissau, Eritrea, and Tibet [23]), the intensity of fighting between rebels and counterinsurgents varies substantially within and across contested areas of the same conflict. At times, insurgents learn more quickly than government rivals; at others, rebels fail to adapt to changes in the strategic environment.

Regardless of insurgent learning, the number of deadly attacks perpetrated as running conflict time increases may remain fixed or decrease because counterinsurgents are learning at least as quickly how to thwart insurgent operations. Even in cases where insurgent learning outstrips counterinsurgent learning, the increase over time would be linear if, for instance, the ratio of insurgent and counterinsurgent learning were approximately fixed. Johnson and his co-authors appear to dismiss any such possibilities, arguing that their “broadbrush theory does not require knowledge of specific adaptation or counter-adaptation mechanisms, and hence bypasses issues such as changes in insurgent membership, technology, learning, or skill set, as well as a need to know the hearts and minds of local residents” [4].

Conditions of strategic interaction make it unlikely that insurgent learning follows a power law process. For this reason, relative (or dynamic) learning may be a more theoretically appropriate concept for scholars interested in understanding patterns of insurgent violence.

**Results.** We test whether the escalation of insurgent violence in Afghanistan and Iraq approximate a power law dynamic. For a substantial number of Afghan and Iraqi provinces, our results reveal that the rate of attack across time decreased rather than increased. Following Johnson et al. in their interpretation of model results, in these areas, insurgent performance worsened with time, implying the absence of (or diminished) learning amongst insurgents in such areas. The source(s) of these negative relationships is uncertain. However, one explanation is that Coalition forces and their host-government partners learned more rapidly than insurgents in these areas, progressively slowing their rate of attack as both conflicts progressed.

Among provinces for which the estimated escalation coefficients are negative, most of the corresponding conflict processes are best described by a linear model. For many of the remaining provinces, the power law function fits the observed
data less well than an exponential model. Finally, the magnitude of the escalation coefficient and/or the model coefficient of determination of all remaining possible province cases fall well outside the range of values to which power law processes correspond.

These results are concurrently presented in Figure 1. Figures 1a and 1c correspond to the Afghanistan and Iraq wars, respectively. In each, the predicted escalation coefficient for incidents of insurgent attacks using IEDs for each province is plotted against the respective model coefficient of determination ($R^2$). Orange points denote those provinces in which, under ordinary least squares regression, a linear model accounts for a greater proportion of total outcome variation. Conflict processes in these provinces can be immediately rejected as following a power law dynamic. Next, for the set of provinces in which the power law model (log-log) accounts for a greater proportion of total outcome variation than the linear model, observations better explained with an exponential fit are plotted as red dots. These too can safely be rejected as following a power law process.

Remaining points are indicated in blue and include only seven of Afghanistan and Iraq's 52 combined provinces. These points represent possible though unlikely cases in which the rate of conflict escalation follows a power law dynamic. To identify province cases for which a power law process cannot be rejected, we construct an “acceptable range” area on the plot. In cases of definitive power law processes, associated model $R^2$ values must exceed 0.763 [9]. A value of 1, the maximum theoretically feasible value of $R^2$, is set as the upper bound. The range of escalation coefficient values are taken directly from [4]. If [4]'s claim is valid, most if not all province points should be observed in blue points within this designated area. No such cases are observed.

These results persist across administrative units. When the district is adopted as the unit of analysis, the same approximate pattern emerges (Figures 1b and 1d). At this level of aggregation, which we believe would be more likely to produce estimates consistent with a power law process if [4]'s claim is correct, not a single case of Iraq and Afghanistan’s combined 592 districts is consistent with the authors’ claim that the frequency of insurgent attacks follow a power law process. Although a power law under least squares regression initially appears to fit some of the data better than a linear model, the power law function is never a definitively better fit of the observed data than an exponential function while meeting the $R^2$ threshold discussed above.

In **Supporting Information**, we introduce alternative measures of insurgent violence. The main result is consistent when the frequency of direct fire, indirect fire and an aggregate measure of all attacks are analyzed as a function of running conflict time, at either the province or district level (Figures SI.1 and SI.2).

When the primary statistical test is replicated using Worldwide Incidents Tracking System (WITS) data so that only severe (fatal) attacks are included in the analysis, results are unchanged (Figure SI.3). Finally, our primary model results are unchanged when the specific date cutoffs and sampling rules adopted by [4] are imposed on our datasets and the replication is repeated (Figures SI.4 and SI.5).

In contrast, we find support for the proposition that learning during periods of insurgency is a dynamic process. Specifically, we find that, throughout both wars, the rate at which counterinsurgents cleared IEDs in conflict zones both increased and decreased with running conflict time. In Iraq, the clear rate falls steadily in favor of insurgents during the early years of the conflict, reaching its lowest point around early 2006 (Figure SI.6b). Counterinsurgents then steadily increase the clear rate until it stabilizes around the beginning of 2008. In Afghanistan, this trend is nearly reversed (Figure SI.6a). In both theaters, the rate of learning never attained a steady state.

**Discussion**

The findings reported by Johnson et al. are bold and provide cause for considerable optimism. Though humankind has, over thousands of years, succeeded in producing various institutions reducing the overall level of violence to which a given individual is exposed [28], violent conflict remains a perennial and consistent feature of human interaction. The possibility that real-time conflict data can be introduced to parsimonious mathematical models to effectively predict the timing of future attacks is compelling. Yet, claims to this effect appear to be wholly the artifact of incomplete conflict data and improper methodological techniques. Instead, the only regularity we observe amongst combatants is that dynamic learning never attains a steady state: at times, insurgents learn more quickly than their opponents; at others, they learn more slowly. Adaptation by counterinsurgents is, of course, possible. However, insurgents are likely to adjust their tactics in response. Conflict scholars have made a number of recent discoveries relating to patterns of insurgent violence ([10], [16], [19]). However, the discovery of an empirical law of insurgency with which governments might predict and prevent political violence remains elusive.

**Methods and material**

To determine whether the empirical concerns described affect [4]'s inferences, we replicate their analysis using detailed georeferenced insurgent attack data compiled by the U.S. Defense Department throughout the Afghanistan and Iraq wars. The full conflict dataset from the Afghanistan war was until recently classified and, until the publication of this article, unavailable to the public. That dataset is being released with this article. The full Iraq war dataset was collected and released by [17].

These datasets represent the most complete collection of information on the insurgency in Afghanistan and Iraq available. U.S. and partner military forces routinely reported information on insurgent activities as they operated in both theaters. Equipped with global positioning system equipment, these forces logged both the georeferenced location and date of each insurgent attack in which they were involved. As a consequence, these data are not subject to the constraints that bias media-based data. [22] confirms this using a subset of the Afghanistan data.

The data also classify the method of each attack. Primary attack types include IED explosion, direct fire, and indirect fire. The Afghanistan dataset includes all recorded insurgent attacks perpetrated between January 2002 and February of 2015 and consists of nearly half of one million total observations. For Iraq, the period of coverage is from January 2004 through December 2011 and includes approximately three hundred thousand observations.

**Analysis.** To test whether the results reported by [4] are robust to the use of alternative data, we replicate the authors’ empirical strategy using the U.S. Defense Department datasets. Like [4], we focus on the rate of IED attacks as running conflict time increases. For our primary analysis, we impose no restriction on the dates of coverage.
Formally, [4] hypothesize that the relationship between the period of time between insurgent attacks and the running conflict is given by the scale-invariant function $\tau_n = \tau_1 n^{-\gamma}$, where $n$ represents conflict days. They estimate values of $\tau_1$ and $b$ with least squares regression. We follow [4] in using least squares regression to estimate the following quantities of interest:

$$\log(\tau_{1,p}) = \sum_{i=1}^{n} \log(i_p) + \sum_{i=1}^{n} \log(i_p)(-b) \quad [1]$$

$$b_p = \frac{n \sum_{i=1}^{n} \log(i_p) \log(\tau_{1,p}) - \sum_{i=1}^{n} \log(i_p) \sum_{i=1}^{n} \log(\tau_{1,p})}{-n \sum_{i=1}^{n} (\log(i_p))^2 + (\sum_{i=1}^{n} \log(i_p))^2} \quad [2]$$

by testing the equation:

$$\log(\tau_{n,p}) = \log(\tau_{1,p}) - b \log(n_p) \quad [3]$$

where $n$ and $\tau$ denote conflict days $\{1, \ldots, p\}$ and the time between any two recorded insurgent attacks, respectively. These quantities are estimated for each of Afghanistan and Iraq’s provinces $p$. Each regression produces an $R^2$ value. Like [4], we assess the power law fit relative to a linear model by comparing all such $R^2$ values to those produced with a linear model:

$$\zeta_{n,p} = \zeta_{1,p} - b^* n_p \quad [4]$$

and an exponential model:

$$\log(\gamma_{n,p}) = \gamma_{1,p} - b^* n_p \quad [5]$$

where $\zeta$ and $\gamma$ denote the temporal lag between attacks (i.e., $\tau$). Following [4], for provinces in which the rate of insurgent conflict escalation is power law distributed, the estimated escalation coefficient $b \in [0.5, 2.5]$. Furthermore, extending [9], the coefficient of determination ($R^2$) should, at minimum, exceed the threshold value of .763.

**Robustness Checks.** To determine whether the results of the primary analysis are consistent with other model specifications, we replicate the analysis above, adopting the district as the unit of spatial analysis. Although [4] adopt the province in their statistical tests, the province is less likely to reveal power law processes in the data because there are so few province-day observations during both the Afghanistan and Iraq conflicts in which no IED attack took place. At the district-day level of analysis, there is considerably greater variation in the binary measure to estimate the rate of attack escalation.

Next, we repeat this exercise, substituting, separately, direct fire and indirect fire attacks for IED attacks. Finally, estimates are generated using an “all attacks” variable, which tracks the occurrence of any insurgent attack type. This analysis is carried out to guard against the possibility that some unique feature of IED attacks leads to spurious positive or null findings. These tests are carried out at both the province and district levels.

[4] also test their theory using fatal attack data, and one possible rejoinder to the replication exercise is that our results only reflect differences in the severity of the attacks studied. We address this possibility by using data on fatal events drawn from Worldwide Incident Tracking System (WITS) dataset.

Finally, in the primary analysis, we make use of the full times series for both conflicts, which exceed the temporal coverage available to [4] in their analysis. To ensure that results are not specific to a particular period of the war, we replicate our primary analysis, adopting the date cutoffs and sampling rules that [4] employ.

**A Test of Dynamic Learning.** Escalation of insurgent violence may or may not reflect learning. Studying insurgent violence without measuring changes in counterinsurgent strategy yields an incomplete perspective on adaptation by non-state actors. To assess the alternative hypothesis that insurgent and counterinsurgent learning fluctuates throughout a conflict, we construct a variable that quantifies the relative rate of learning using IED detection. We calculate the reported number of IEDs Coalition and host-government forces in Afghanistan and Iraq discovered and cleared relative to the total number of exploded and cleared IEDs. This IED clear rate provides a measure of insurgent effectiveness in detonating their bombs. A clear rate that decreases over time indicates that insurgents are detonating an increasing proportion of their devices against targets. While counterinsurgent forces may be learning from insurgent behaviors during the period, insurgents are presumably learning more quickly. To isolate clear-rate trends across the two conflicts, we decompose weekly average clear-rate time series for the countries using local polynomial regression (LOESS) [21]. This process removes seasonal trends identified with weekly data.

**ACKNOWLEDGMENTS.** The authors are grateful to Jacob Shapiro for inspiring this project and thank Aaron Clauset for comments on this project. We gratefully acknowledge support for this project provided by the Lynde and Harry Bradley Foundation and the National Science Foundation (Grant No. 2012152169). Responsibility for errors remains solely with the authors. Data used in this article will be made available at scholar.princeton.edu/ashaver and www.austinwright.com.
Fig. 1: Power law diagnostics for IED events in Afghanistan and Iraq, province and district.

(a) Afghanistan, district, IED
(b) Afghanistan, province, IED
(c) Iraq, province, IED
(d) Iraq, district, IED
Fig. SI.1: Additional power law diagnostics for Afghanistan and Iraq, province.
Figure S12: Additional power law diagnostics for Afghanistan and Iraq, district level.

(a) Afghanistan, district level.
(b) Afghanistan, direct fire.
(c) Afghanistan, indirect fire.
(d) Afghanistan, all events.

(e) Iraq, district level.
(f) Iraq, direct fire.
(g) Iraq, indirect fire.
(h) Iraq, all events.

Coefficient of determination, R²

Acceptable range

False positive
No
Yes

Power law fit diagnostics: Afghanistan (district, direct fire)

Power law fit diagnostics: Afghanistan (district, indirect fire)

Power law fit diagnostics: Afghanistan (district, all events)

Power law fit diagnostics: Iraq (district, direct fire)

Power law fit diagnostics: Iraq (district, indirect fire)

Power law fit diagnostics: Iraq (district, all events)
(a) Afghanistan, province, casualties tracked by WITS system

(b) Afghanistan, district, casualties tracked by WITS system

Fig. SI.3: Power law diagnostics for fatal events in Afghanistan using WITS system, province and district
FIG. S1. Additional power law diagnostics for Afghanistan and Iraq provinces (h) Sampling rules

(a) Afghanistan, province, indirect fire
(b) Afghanistan, province, direct fire
(c) Afghanistan, province, IED
(d) Afghanistan, province, all events
(e) Iraq, province, indirect fire
(f) Iraq, province, direct fire
(g) Iraq, province, IED
(h) Iraq, province, all events
Fig. SI.5: Additional power law diagnostics for Afghanistan and Iraq, district ([4] sampling rules)
FIG. SI 6: Dynamic learning in strategic environments: evidence from time series decomposition of relative learning by insurgents and counterinsurgents.

(a) Afghanistan Time Series Decomposition

(b) Iraq Time Series Decomposition

Fig. SI 6: Dynamic learning in strategic environments: evidence from time series decomposition of relative learning by insurgents and counterinsurgents.