BOEKEN

Small Wars, Big Data

The Information Revolution in Modern Conflict
Door Eli Berman, Joseph H. Felter and Jacob N. Shapiro
Princeton (Princeton University Press) 2018
408 blz.
ISBN 9780691177076
€ 27,-

The title of this book is a catchy misnomer. Small wars (i.e. asymmetric warfare) have killed hundreds of thousands of people, displaced millions more, and destroyed the physical infrastructure and social fabric of societies. By any definition, these wars are not small wars. Furthermore, the authors fail to address the heart of the concept of Big Data (i.e. automated data processing) as they are primarily interested in information retrieval. This criticism aside, the authors' efforts have resulted in a well-documented investigation into the role played by information in (counter) insurgency.

Small Wars, Big Data consists of ten chapters. Chapter 3 (*Information-Centric Insurgency and Counterinsurgency*) presents the theoretical model of the book. Chapters 4-8 examine conflict: past conflicts (Vietnam), and ongoing conflicts in Afghanistan, Iraq, and Syria, but also lesser-known conflicts in the Philippines and Nigeria. Chapter 4 details developments in information technology, e.g. – widespread use of mobile phone technology, and its effects on insurgent violence. A two-sided sword in the eyes of the authors: low-cost mobile telecommunication enables better intelligence on insurgent activities, but also enables insurgents to plan and coordinate attacks better. Chapter 5 discusses the role of governments in conflicts, whereas chapter 6 focuses on suppression, and how state actors suppress insurgent activity. Chapter 7 addresses civilian-rebel relationships, followed by an investigation of the hypothesis that poverty is the root cause of violence and insurgency in chapter 8. Chapters 9 and 10 summarize the (theoretical) work done and offer some principles/policy suggestions. The authors encourage large-scale, detailed information retrieval when dealing with armed conflicts. According to them, 'dramatic effects' can be achieved by making information-sharing safe for civilians. In their opinion this requires to 'invest in administrative data, then weigh the benefits of sharing against the risks' (pp. 320-323).

A new theory

The overarching thesis of *Small Wars, Big Data* is not new. The maxim that 'information flowing from non-combatants is the key resource in asymmetric conflicts' (p. 18) can be traced back to counter-insurgency (COIN) theorists stressing the need for detailed intelligence in combating insurgents.[1] But what sets Berman, Felter and Shapiro's book apart is the precise means by which they leverage that information, and the insights into various aspects of different insurgencies that such an approach yields. Their aim is ambitious: a new theory to better comprehend and deal with asymmetric warfare. Focusing on the triangular relationship between civilians, rebels and local government, the authors try to understand the complex relationships and interactions between these actors, which they consider paramount to the success of counterinsurgency operations (see Figure 3.1). How to achieve this? As they themselves put it, the accumulation of small facts can lead to bigger answers.

The three authors are researchers at the Empirical Studies of Conflict (ESOC) group of Princeton University.[2] What unites them, and gives ESOC its intellectual coherence, is a shared epistemology: faith in the power of data to reveal hidden truths about conflict zones, and thereby guide practitioners in the field. Their approach depends upon their relationship with the US military from which they have received significant funding. ESOC researchers have access to often classified datasets maintained by the military, e.g. the SIGACT-III database

---

1   See for example David Galula, *Counterinsurgency Warfare Theory and Practice* (Praeger, 1964); John A. Nagl, *Learning to Eat Soup with a Knife. Counterinsurgency Lessons from Malaya and Vietnam* (University of Chicago Press, 2005); David Kilcullen, *Three Pillars of Counterinsurgency* (US Government Counterinsurgency Initiative, 2006).
2   Chapter 2 (pp. 23-54) provides a detailed description of ESOC's motivation and approach.
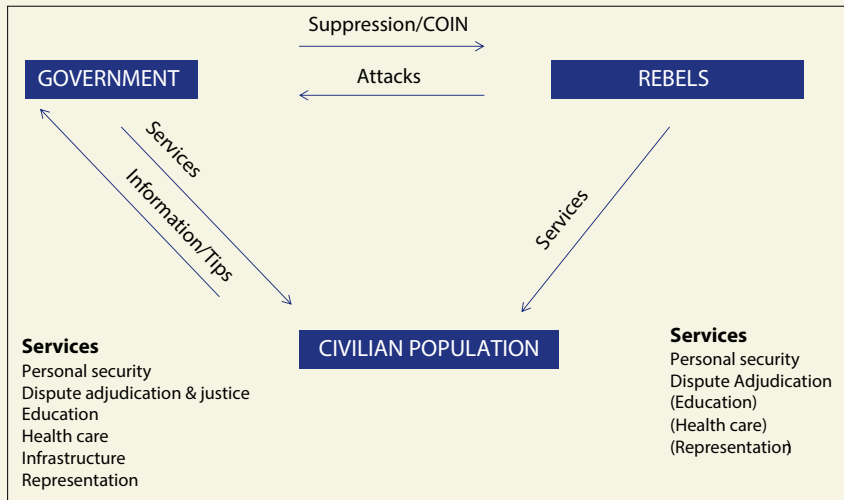
Figure 3.1. Asymmetric conflict modeled as a three-player game (source: Berman, Felter, Shapiro, Small Wars, Big Data, pp. 64-65)

NB: Rectangles – represent actors; arrows – stand for actions. At the top government and rebels opposing each other. According to the authors information-centric insurgency necessarily includes the bottom rectangle: civilians can choose to provide information to the government and might be influenced in doing so by services provided by the government, rebels, or both. Curiously, however, there is no arrow (no information flow) from the civilian population to rebels, which seems a rather implausible omission.

that 'recorded every "significant activity" involving U.S. forces with a precise time and location stamp' in Afghanistan and Iraq. In turn, the US military has gained detailed insight into when and where to deploy patrols, where military checkpoints prove most effective, and whether surge tactics in Iraq affected sectarian violence.

Despite this commitment to a data-driven approach, the book is not written by crude neo-positivists. The data approach depends upon qualitative research to reveal localized variables within conflict zones. Without that contextual awareness, data can prove misleading. To illustrate this, the authors provide monthly trends in combat incidents in the 24 most violent districts of Afghanistan (2005-2014). District variances reveal local differences that macro-data on the

conflict as a whole would neglect. Without sensitivity to local conditions, these data-sets are easy to misinterpret. Take the decline in violent incidents in the Kamdesh district between 2010-2012. 'If you went by the data alone', the authors write, 'you might be forgiven for thinking that ISAF forces began doing something right from 2010 onwards. In fact, however, what explains this decline is that ISAF withdrew almost entirely from Kamdesh in this period, leaving the Taliban there with no enemy to fight.' (p. 36).

This approach to data analysis rests on a number of core principles:
1. It focuses upon 'microdata' – the more 'micro' the better – that reveal distinct patterns in particular localities.
2. Analysis of those data must be sensitive to qualitative contextual knowledge.

3. The process of reaching conclusions is empirical and iterative, slowly building a body of small facts, to arrive at some larger conclusions.
4. Empirical and theoretical research engage in a dialogue with each other, as theoretical assumptions about causal relationships in conflict are tested through empirical analysis.

The results of this sensitive data approach, in terms of their implications for COIN practitioners, are impressive. The authors demonstrate how, when and where to distribute aid to local populations to achieve maximum impact in countering insurgents. They demonstrate that large-scale infrastructure projects often run into problems, and can even prove counter-productive, whereas smaller scale (50,000 dollars or less) and more targeted projects in relatively secure environments are more likely to yield positive results. They show how local conditions in the data drives the decision to engage an enemy, and how complex the relationship between economic conditions and violence is, and how understanding and responding to that complexity is likely to improve security.

These results, and the sophisticated manner in which they are achieved, make *Small Wars, Big Data* a vital addition to the bookshelf of anyone interested in COIN operations, especially practitioners in the field. Ultimately, this approach is likely to yield 'smarter' COIN-operations by helping commanders and aid workers recognise the specific conditions of local environments. But some of the wider claims and underlying assumptions of this impressive study might give more skeptical readers food for thought.

BOEKEN

The use of the SIGACTS database calls into question data integrity issues. What exactly constitutes 'significant data'? How accurate is this trove of data on geographical locations, time stamps and types of incidents: from IEDs, complex attacks, hit-and-run ambushes, casualties, to demonstrations? Data can reveal, but can also hide, disguise, and change appearances like a chameleon.

Such considerations are only taken note of in passing (ref 16, p. 333): 'The broad need to exercise caution when using administrative data from conflict zones is amply demonstrated by the many examples in Ben Connable's excellent *Embracing the Fog of War. Assessments and Metrics in Counterinsurgency*.'[3] This study lists many complaints with the SIGACTS database (pp. 161-166):

- 'SIGACTs are never accurate. The database isn't even complete. If you think that every incident that happens is entered […] you're kidding yourself.'
- 'I'd say 10 to 20 percent of attack reports […] had misinterpretations of category. When you […] read the narrative of the attack, you find that […] the wrong codes were plugged in to label the type of attack.'
- 'Sometimes, reports are changed or updated eight or ten times after they are initially recorded.'
- 'SIGACTS data provide limited support for empirical analyses owing to geographical or temporal gaps in coverage, and missing, incomplete and erroneous data. SIGACTs are limited in support of COIN-type security assessment.'

Connable is fair in his treatment by stating that 'anecdotal comments are not sufficient to call into question the validity of the entire SIGACT data set.' According to him: 'Empirically disproving accuracy would require a […] very costly study that might ultimately prove to be inconclusive. But if it is difficult to conclusively disprove the accuracy of the SIGACT data, it is also difficult to prove their accuracy.' In his opinion there are many problems with the SIGACT data, including conceptual ones regarding violent incidents in COIN. For example, how to differentiate between insurgency-related violence, and common criminal activity, or intra-tribal and inter-ethnic violence. Without a method for tracking non-insurgency-related violence, the rise or fall of insurgency-related violence cannot be ascertained.

The fact that *Small Wars, Big Data* does not investigate data integrity more closely is striking. For their research into the Iraq arena the authors used 168,730 SIGACTS reported, or 87 percent out of a total of 193,264 (p. 123). This 'neglect' is all the more striking since they wish to identify causal relationships (see the paragraph *Establishing Causal Relationships through Iterated Research*, p. 43 onwards). They explicitly aim for a 'theory or research design for causal identification using Randomized Control Trials (RTC)'. Simultaneously admitting that 'while it would be enormously beneficial to deploy RCTs to analyze the determinants of violence and support for violent groups, is unfortunately often impossible.' (p. 47).

Besides, Big Data is geared towards finding similarities between seemingly unrelated datasets. The logic of correlation, however, only implies probability (not causation) and even strong correlations might be a coincidence. An analysis based on statistical probabilities will always produce false positives (targeting innocent people) and false negatives (unnoticed security risks). Ultimately, Big Data is ill-equipped to deal with causal mechanisms – the holy grail of this book.

'Information-centric model'
The authors' propose an 'information-centric' model that seems to imply that winning COIN-campaigns is a matter of having more accurate and more granular data. Their results demonstratethat you can improve COIN-operations with a sensitive understanding of complex datasets, but some asymmetric conflicts might simply prove unwinnable, even with the best data scientists present. In the Vietnam War, ideological and nationalist considerations helped the North Vietnamese/Vietcong continue to fight despite appalling casualty rates. However, the death of 1,1 million North Vietnamese/Vietcong soldiers against 200,000-250,000 South Vietnamese and over 58,000 American soldiers did secure rebel victory. The authors also fail to mention the successful application of brute force. They could have solidified their case by discussing the effectiveness of an information approach against tactics of mass incarceration/resettlement, and large-scale killing or outright genocide. Consider the two brutal wars fought by the Russians in Chechnya during the 1990s/2000s.

The authors' understanding of COIN depends upon a particular historio-

---

3 Ben Connable, *Embracing the Fog of War. Assessments and Metrics in Counterinsurgency* (RAND, 2012).

graphical interpretation of the Vietnam War, namely, that the 'heart and minds' and COIN-operations by the US military in Vietnam were largely successful (p. 199). Many historians would oppose this view, citing the intractable nature of the conflict, the problem of being regarded as an invading power, and declining support on the home front, to name but a few interpretations of the US defeat. However effective the 'hearts and minds' campaign in Vietnam might have been, clearly it – was not effective enough to win the war. Would more and better data have changed this outcome?

Much of the authors' assumptions about the possible response of civilians to COIN operations depend upon game theory. Critics argue that this approach simplifies the choices available to actors at a given moment, and the possible non-rational explanations for why a civilian might agree to cooperate with COIN forces. There is often a thin line between good science and scientism. This book contains a great deal of the former, but sometimes the authors' assertions about the supremacy of the data-driven approach runs the risk of slipping into the latter.

Finally, *Small Wars, Big Data* is a hefty read requiring some stamina from the reader. The authors try to improve readability by providing introductory (hypothetical) stories to 'pull' the reader into the chapters. It is up to the reader to determine whether they have succeeded or not, but it is an admirable attempt to sensitize those interested in the fact that dealing with the dilemmas of (counter)insurgency concerns real people living real lives. This certainly holds true for civilians caught as always between a rock (rebels) and a hard place (local government).

All concerns aside, *Small Wars, Big Data* is undoubtedly an innovative study that contains rich empirical depth that will likely inform both practitioners and academic debate on COIN and asymmetric conflict for decades to come. In short, an essential addition to the COIN bookshelf. ∎

Dr. A. Claver (Netherlands Ministry of Defence) and dr. S. Willmetts (Leiden University)