CONTEXTUALISING VS. FORECASTING: DEFINING THE ROLE OF DATABASES IN THE EMPIRICAL STUDY OF TERRORISM

Alexander Hull is a M. Phil candidate in Modern South Asian Studies at the University of Cambridge, as well as the South Asian editor at the Extremis Project. His academic interests include Talibanisation in the Tribal Areas of Pakistan, Islamic terrorism and radicalisation, and South Asian nuclear issues.

ABSTRACT

In recent years the application of quantitative research to the study of terrorism has been one of the hallmark progressions of the field. While this is indicative of the modernisation of terrorism studies, counter-terrorism practitioners must be clear in defining the role that quantitative research should play in the empirical study of terrorism. This paper will attempt to provide an argument for limiting the ends of quantitative databases to the contextualisation of terrorist attacks, rather than an attempt to use databases for forecasting attacks.

RÉSUMÉ

Au cours des dernières années, l'application des méthodes de recherche quantitative lors de l'étude du terrorisme a été un des progrès marquant de ce domaine. Alors que ceci est un indicateur de la modernisation de l'approche face à l'étude du terrorisme, les individus œuvrant dans la lutte contre le terrorisme devraient avoir une idée claire du rôle que la recherche quantitative devrait jouer dans l'étude empirique du terrorisme. Cet article propose des arguments afin de limiter les bases de données quantitatives à la contextualisation des attaques terroristes, plutôt que d'utiliser celles-ci afin de tenter de prévoir d'éventuelles attaques.

INTRODUCTION

The quantification and modelling of terrorism is in part a by-product of the broader trend within physics and economics of applying statistical models, more commonly associated with the "hard sciences," to the social sciences. Sociophysics, as this discipline is frequently called, began to take root within the physics and economics communities in the 1970s as an attempt to apply models used within statistical physics to social phenomena (Galam, 2004). The greatest advocate of the utility of sociophysics is the French physicist Serge Galam, and his models have been applied to social events such as voting patterns, the fragmentation or coalition of political parties, opinion dynamics, and decision making (Galam, N.D). This trend within the hard sciences converged with the explosion of literature and research on another social phenomena, terrorism, following the September 2001 attacks on the United States.

Following the 2001 attacks, practitioners and academics realized the serious dearth of field data that existed within the empirical study of terrorism (Arce *et al.*, 2011), and in recent years there have been several attempts to quantify terrorism, with actual terrorist attacks being the indices used. The University of Maryland's Study of Terrorism and Responses to Terrorism (START) currently operates the most comprehensive event database of terrorism. Its Global Terrorism Database (GTD) has logged over 100,000 terrorist attacks globally between 1970 and 2012. While databases such as the GTD are significant insofar as the aforementioned lack of field data can begin to be rectified, there is little discussion within the empirical literature about what quantitative databases are actually useful for. As Jenkins and Bond have noted, there has been a desire for an empirically-based early warning system that has the ability to forecast political crises long before 9/11 (Jenkins & Bond, 2001).

While the political crises of the 1980s and 90s were marked by instable regimes and humanitarian crises, the 21st century thus far has seen Western regimes' primary political instabilities stemming from transnational and domestic terrorism. It is perhaps a natural process for the

increasing data available on terrorism to be used for such an early warning system. The remainder of this paper will discuss what the ends of such empirical tools should be for researchers, with the assertion that quantitative data is useful for contextualising past events, but that policy makers and academics should avoid the temptation to turn quantitative datasets to "game" outcomes and attempt to develop an algorithmic early warning system.

THE UTILITY OF QUANTITATIVE DATABASES

What questions should academics and policy makers be asking when observing empirical studies of terrorism such as START's Global Terrorism Database? For 2011, START's dataset allows us to break down the incidences of terrorism globally, and trends can be elicited. For example, Iraq in 2011 had the greatest share of any nation of worldwide terrorist attacks (25%), while South Asia as a region saw the greatest proportion (48%) (START Consortium, 2012). This data presents two important utilities for counter-terrorism practitioners. Firstly, the quantification of attack indices allows policy makers to rationalize counter-terrorism resource allocation. For instance, despite the insistence by some that al-Qaeda Central is the primary transnational threat (Hoffman *et al.*, 2011), the empirical analysis shows the increased profile of the groups' affiliates, and the decline of the central groups attack profile. Quantitative databases are therefore useful for policy makers and counter-terrorism practitioners to speak about the current threat environment with more confidence and credibility, and move beyond the at times sensationalistic narrative about which areas or groups are, or are not, a serious threat.

Further, terrorism is unfortunately an increasingly apparent phenomenon, and quantitative databases allow policy makers and practitioners the ability to elicit trends from a very large dataset. As Sandler notes (2011), the GTD is particularly useful for recording terrorists' responses *during* ongoing terror incidents, such as the sequencing of the release of hostages. The issue going forward however is that policy makers will always attempt to design and implement a system that extracts as many correct signals as possible and therefore serves as an early warning system (Bussiere & Fratzscher, 2008). While there are significant obstacles to creating such an early warning system, Koenraad Van Brabant (2012) has identified a particular scenario whereby incident statistics may be useful in forecasting terrorist attacks. He notes that it may be possible to forecast as to the nature of one emerging threat, or group, in one region if it develops conditions that mirror a similar threat in another region. For example, the dataset of attacks perpetrated by al-Qaeda in Iraq may allow policy makers to speculate as to the likely impact of the growing al-Qaeda movement in northern Africa. However, regional dynamics, differing government responses to terrorism, and other factors make it necessary for a strong qualitative caveat when attempting to use one dataset to forecast impact across regions.

THE LIMITS OF QUANTITATIVE DATABASES

The limits of a quantitative approach to terrorism in general, and quantitative databases in particular, are many. It is for this reason that the ends of quantitative databases need to be clearly limited to the contextualisation of the current and/or past threat environment, and not be used as a tool to forecast terrorist attacks.

Firstly, quantitative databases have inherent methodological flaws. While this is an issue across subjects, the nature of terrorism makes this concern particularly acute. As Hellerstein (2008) notes, data entry is often corrupted by "spurious integrity" issues, whereby the user does not have data for a certain field and thus makes an estimate as to the most-logical entry. This issue is apparent within the GTD, where of the 417 incidents that took place within December 2012, approximately 50% do not contain a confirmed entry in the "perpetrator" index. The fact that many groups, the Afghan and Pakistani Taliban in particular, often claim responsibility for attacks carried out by unaffiliated groups compounds the problem of spurious integrity, as well as creates an impression from a quantitative perspective of an exaggerated profile. This issue should give pause to those who are quick to use quantitative databases in order to forecast future events.

While the inherent issues of data collection and entry issues of quantitative databases present limitations as to their utility, perhaps the largest obstacle that the nature of terrorism presents to quantitative forecasting is what a Chatham House report has termed "high-impact, low-probability events," or HILPs. Drawing our attention to terrorism as an HILP, the report states that "Unforeseen shocks, such as the 9/11 terrorist attacks raise questions of how we build capacity to respond to an event that has not been conceived in advance as a realistic threat or whose frequency is unknown" (Lee & Preston, 2012). While terrorist attacks are in and of themselves unfortunately not HILPs, there are particular types of terrorist attacks which *are* HILPs. These events, such as 9/11 or the Anders Breivik shooting, are the type of attacks that pose the greatest risk to society, and quantitative databases' ability to mitigate risk therefore must be measured against HILPs.

Breivik's attack, which killed seventy-seven people, draws attention to the weaknesses of the GTD database in two respects. Firstly, the GTD does not allow users to search attacks that have been carried out by individuals with no affiliation with a terrorist group. While this is a

simple data entry issue that can be easily resolved, more worrying is the fact that from a quantitative standpoint, Breivik's attack is a significant outlier. Norway is nearly immune from terrorist attacks, and before Breivik's 2011 attacks, only two people had been killed by terrorist attacks in the country (START Consortium, N.D). While this attack represents a quantitative anomaly, the qualitative approach can be useful. For instance, one month before the attack, Matthew Goodwin, one of the leading experts on right-wing extremism in Europe, cautioned about the possibility of a far right-wing "lone wolf" actor carrying out a serious attack (Arnot, 2011). While it is difficult to institutionalize expert opinion into any sort of formal early warning system, there are several cases where qualitative, subject matter experts have forewarned terrorist attacks,[1] and this advances perhaps the strongest case for restraint when using quantitative analysis to forecast terrorist attacks.

CONCLUSION

While the increasing availability of data in the counter-terrorism field is a positive indication of the maturation of the field, there has been little, if any, discussion as to the ends of quantitative databases. This paper has attempted to present a case for using quantitative databases such as the GTD for contextualising what is often a very large body of information and incidents. The case for using quantitative databases for forecasting terrorist attacks, however, is tentative at best. Not only do quantitative databases have inherent flaws in and of themselves, the nature of terrorism means that data collectors and analysts will often have to interpret information subjectively and within broader contexts, which quantitative databases will often fail to capture. Policy makers should therefore be clear in their expectations of quantitative databases, and rely upon them for contextualisation rather than forecasting.

NOTES

[1] Perhaps most famously, former National Coordinator for Security, Infrastructure Protection, and Counter-terrorism for the United States Richard Clarke's warnings about al-Qaeda's desire to attack the US homeland.

REFERENCES

Arce, D.G., Croson, R. & Eckel, C.C. [2011] "Terrorism experiments," *Journal of Peace Research*, 48(3).

Arnot, C. [June 28, 2011] "How can divisions between communities be closed?" *The Guardian*, http://www.guardian.co.uk/society/2011/jun/28/closing-community-divides-matthew-goodwin-interview (Accessed February 1, 2013).

Bussiere, M. & Fratzscher, M. [2008] "Low probability, high impact: Policy making and extreme events," *Journal of Policy Modeling* (30).

Galam, S. [May 1, 2004] "Sociophysics: a personal testimony," *Physica A*, 336(1-2).

Galam, S, "Sociophysics: A review of Galam models," N.D., Available: http://arxiv.org/pdf/0803.1800.pdf (Accessed January 30, 2013).

Hellerstein, J.M. [2008] "Quantitative Data Cleaning for Large Databases," *UC Berkeley.*

Hoffman, B., Habeck, M., Zelin, A. & Levitt, M. [2012] "Is al-Qaeda Central Still Relevant?" *The Washington Institute*, http://www.washingtoninstitute.org/policy-analysis/view/is-al-qaeda-central-still-relevant.

Jenkins, J.C. & Bond, D. [2001] "Conflict-Carrying Capacity, Political Crisis, and Reconstruction: A Framework for the Early Warning of Political System Vulnerability," *Journal of Conflict Resolution*, 45(1).

Lee, B. & Preston, F. [2012] "Preparing for High-impact, Low-probability Events: Lessons from Eyjafjallajökull," *Chatham House.*

START Consortium. [October 16, 2012] "New data reveals al-Qaida-linked groups among most active terrorist groups in the world," http://www.start.umd.edu/start/announcements/announcement.asp?id=424 (Accessed January 30, 2012).

START Consortium, "Terrorist Incidents in Norway Over Time," http://www.start.umd.edu/gtd/search/Results.aspx?expanded=no&casualties_type=&casualties_max=&success=yes&country=151&ob=GTDID&od=desc&page=1&count=100#results-table (Accessed January 30, 2013).

Van Brabant, K. [2012] "Incident Statistics in Aid Worker Safety and Security Management: Using and Producing Them," *European Interagency Security Forum.*