Process Modeling for the Study of Non-State Political Violence

Olivier L. Georgeon
Jonathan H. Morgan
The College of Information Sciences and Technology
The Pennsylvania State University, University Park, PA 16802
olg1@psu.edu, jhm5001@psu.edu

John Horgan
Kurt Braddock
The International Center for the Study of Terrorism
The Pennsylvania State University, Univesity Park, PA 16802
jgh11@psu.edu, khb125@psu.edu

Keywords:
Knowledge representation, Pattern recognition, Socio-cultural modeling, Terrorism

ABSTRACT: *Terrorism studies have and continue to face conceptual and analytic challenges that stem from the assumption that terrorism can be understood outside of its social and political context, as essentially a 'state' of being and/or set of personal qualities specific to the terrorist (Sageman, 2004; Taylor & Horgan, 2006). An under-explored alternative to this view is to see involvement in terrorism, at least in psychological terms, as a process rather than a state. One consequence of this is that we shift the focus away from individuals and their presumed psychological or moral qualities to an examination of process variables. These, by their nature, are more susceptible to change and thus form the basis of developing interventions. Interpreting these variables, such as changes in operational context or relationships between temporal events and individuals, requires tools capable of capturing time-sensitive semantic content. To date, there are few process-oriented tools and fewer analyses of terrorism data using these tools. In this paper, we present such a tool and offer an initial application for expanding and formalizing computationally our understanding of terrorism.*

1. Introduction

A major obstacle to greater conceptual development in the study of terrorism has been the assumption that we can understand terrorism outside of its social and political context. This has given rise to the view that terrorist acts essentially can be understood as stemming from an identifiable 'state' of being that can be analyzed to make predictions. Though popular, this assumption and the emphasis on static qualities that is implied by such an approach has proven ineffective, particularly in the development of meaningful counterterrorism initiatives (Horgan, 2009). Alternatively, it may be more valuable to consider involvement in terrorism (and political violence more broadly) as reflecting a complex process rather than a state.

Studying terrorism as a process makes us shift our focus from the individual and their presumed psychological or moral qualities to process variables. We can then begin to ask how changes in operational context, or how the relationships existing between events and the individual affects behavior (Taylor & Horgan, 2006). This is particularly important when considering how we might formulate strategies for managing and controlling the extent of terrorist events (Horgan, 2009).

In addition, as Taylor and Horgan (2006) note, considering terrorism as a process would be consistent with the way we tend to study other forms of illegal behavior such as criminality. A further benefit that follows from this is that our attention transitions from addressing the qualities of individuals (i.e., personality or "evil traits") that draw on intangible mentalistic concepts (that are, by definition, resistant to change and not visible) to identification of essentially tangible, practicable, and alterable matters. Moving our level of explanation away from properties to processes seems to offer tangible rewards beyond mere conceptual adequacy, and may offer a different approach, for example, to the development of more practical and efficient counterterrorism initiatives.

What then does assessing "terrorism as a process'' imply? In this paper, we use the definition of process developed by Taylor and Horgan (2006) in that we are essentially describing a sequence of events, involving steps or operations that are usually ordered and/or interdependent. We therefore seek to understand terrorist activity as a set of actions and reactions, often expressed in a reciprocal relationship in both an immediate and long-term sense between various actors. These actors can include but are not limited to: governments, terrorists, the media, the police and security services, politicians, and the civilians in general. As Taylor and Horgan explain, "the nature of that reciprocity may be expressed in a variety of ways, but it is important to note, however, that specifying or identifying the elements of the process does not necessarily imply a simple deterministic account, despite the ease with which such accounts may follow from post hoc analyses of events" (Taylor & Horgan, 2006, p 585).

In addition, describing activities as indicative of a process allows us to consider modeling events and their relationships. As Taylor and Horgan (2006) explain: "Modeling can take a variety of forms, and perhaps a continuum can be expressed between identifying and expressing mathematical or statistical probabilities about the relationships between events, and conceptual models of their relationships expressed as hypothetical constructs and intervening variables."

In this paper, we introduce a tool and initial trace modeling approach for expanding and computationally formalizing our knowledge of terrorism processes. We first introduce trace-modeling approaches as means of addressing the growing data/knowledge gap found in the social sciences. We then discuss the limitations of classic activity analysis. We move on to discuss process modeling using trace-modeling methods, providing a brief specification and offering a process-oriented trace-modeling tool, ABSTRACT, to support the modeling of terrorist activities. We follow this discussion with a description and analysis of an example trace developed from the Global Terrorism Database[i] (GTD). We then conclude with a brief discussion and review, noting challenges and implications of this modeling approach.

2. Addressing the data/knowledge gap

The data that may potentially inform us about terrorist processes is diverse. It can range from established sources such as intelligence reports and field work, case studies, and centralized logs of terrorism activity like the GTD to emerging media types such as chatroom logs, tweets, and other life streaming sources. For data, however, to inform us about a process, it must entail chronological information. Such data constitutes what we call a chronological activity trace. A chronological activity trace can be seen as a timeline of concrete or abstract events in which the analyst can find relations of causality between events, by referring to possible explanative theories.

Finding this network of abstract events and causal relations is challenging. This challenge raises a problem that we refer to as the data/knowledge gap. In essence, this challenge arises from an epistemological issue—the fact that to understand data we need previous knowledge, but to have previous knowledge we need to understand data. This is a general problem that is often related to Popper's (1972) evolutionist theory of knowledge. In this article, we limit our focus to addressing two dimensions of this issue: a) the gulf between disciplines (primarily between toolmakers and tool-users), and b) the conceptual gap in our understanding of terrorism.

On one hand, we have high-level descriptions of terrorist activity formulated over multiple decades and drawing primarily upon interviews, court transcripts, and case studies coming from the direct experiences of researchers. These theories continue to offer insights, but their dependence upon a relatively small set of retrospective accounts limits their predictive power. From these sometimes inscrutable and always evolving accounts (a snapshot view), researchers attempt to identify the dynamics of a fluid, time-sensitive, and frequently reflexive set of processes.

On the other hand, there is a growing store of low-level granular data of multiple types. Finding patterns or processes in this kind of low-level data continues to be a challenging research area, as examples in other domains of human activity show, e.g, car driving activity (Georgeon, 2008). Though this data potentially offers a means of evaluating and refining our theories, constructing a useful interpretation of this data is not only a difficult challenge for the social sciences but also for the information sciences—a challenge neither community can surmount in isolation. Social scientists will require tools to interpret data; information scientists require the expertise of social scientist to ensure both the relevance and applicability of those tools and data.

Furthermore, the success of such tools is likely to vary in relation to the tractability of the process or sub-process we are studying. While online recruitment by terrorists generates large volumes of data, we are much less likely to fully capture the influence of idiosyncratic or contingent factors, or formulate a complete picture of processes whose participants systematically destroy or distort the data necessary to understand that process. For example, collected data seldom entails information about underlying social mechanisms. Consequently, social scientists must hypothesize, based upon incomplete information, the existence, relative significance, and operation of these processes (Hedström, 2005). We, therefore, must be realistic about our ability to predict terrorism, and rather confine ourselves to attempting understand and potentially predict certain terrorist activities and processes.

We address the data/knowledge gap by using an iterative and reciprocal top-down/bottom-up approach, drawing downwards from models proposed by experts and upwards from granular data. This approach can also be seen as a process of modeling activity traces by applying *abductive reasoning*, i.e. searching for hypothetical causes to explain observed consequences. In our case, the observed consequences are the events recorded in the data. The hypothetical causes can be either events already recorded in the data or abstract events that the expert adds to the trace. In both cases, the expert asserts the causal link based on his models or expertise. Notably, logicians consider abductive reasoning both as a non-logically-valid method, and as the only method of logical inference that can yield new knowledge. Once formed, the hypothetical causes and explanations need to be recorded in the trace. Then, the system should help the analyst ensure formal consistency and evaluate these hypotheses in terms of usefulness for making predictions. We call this process *expert-driven trace modeling*.

We will discuss one approach for conducting this trace modeling process throughout this paper. We start with a presentation of a top-down analysis in section 3. This presentation leads us to specify the requirements for an activity-trace modeling tool in section 4. We then present our prototype implementation of such a tool in section 5. We present our usage of this tool for expert-driven bottom-up modeling of field data in section 6. We then discuss how we imagine the two processes (top-down and bottom-up) could meet in the middle.

3. Top-down analysis

The literature provides us with diverse examples of top-level models of processes that lead to non-state political violence. Figure 1 depicts Horgan's (2009) description of the phases of involvement and engagement in terrorism. Critically, Horgan, as do other authors (e.g., Sageman, 2004), makes a distinction between radicalization and engagement in actual terrorist activity. In Figure 1, the circles represent conceptually discrete but often overlapping phases of activity. We can break these phases down into organizational sub-processes, as we do in Table 1 with the violent radicalization phase. Such break downs show the initial pathway to symbolic sequential modeling.

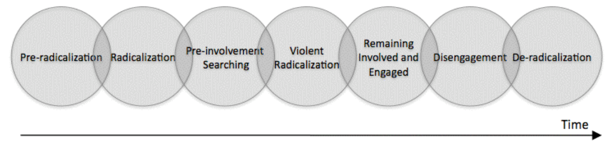

Figure 1: Pathway into, through, and out of terrorism (Horgan, 2009, p. 151).

From this break down, we have constructed a timeline representation of these different phases as shown in Figure 2. We have done so with an existing open-source visualization tool called Simile Timeline[ii].

Table 1: Hierarchy of sub-processes of violent engagement drawn from Horgan (2009).

(A) Decision and search activity - targeting and "pre-terrorism"
- Plan
- Have a leader
- Connect to an organization
- Search for suitable situations

(B) Preparation and "pre-terrorist" activity
- Target identification
- Identification and selection of appropriate personnel
- Training, general and specific to target
- Design and manufacturing related to device construction
- Device testing and preparation

(C) Event execution
- Bring device and manpower to the scene of the attack
- Maintenance, surveillance, security of the operation
- Dynamics of the event
- Securing of weapons after attack

(D) Post-event activity and strategic analysis
- Destruction of evidence
- Post-event evaluation

This modeling illustrates some of the limitations of available timeline visualization tools. Such tools require a precise timeline of events to represent events numerically—this proves unwieldy when modeling high level terrorist processes. As long as we do not know precisely at what timescales terrorist activities are operating (hours, days, weeks, months, years, or decades), we need to formalize the succession and relations between events as opposed to their real duration. Consequently, such a process model should be invariant through scale but should rather allow the analyst to express temporal relations such as sequentiality, concurrence, or overlap. In other words, we need a tool capable of supporting

Figure 2: High-level timeline drawn from Table 1.

pattern analysis on a more abstract level for datasets where few or no dates are available.

More broadly, we need tools that allow the analyst to represent events symbolically, identify symbolic patterns, and from those patterns develop new symbols that represent meaningful sequences of activity. These sequences, in turn, can indicate the emergence of processes over larger timescales. This approach addresses both dimensions of data/knowledge gap described above by (1) supporting the intelligible analysis of granular data that in turn can inform theory, and (2) by facilitating cooperation between knowledge-engineers and domain experts as they attempt to develop a meaningful trace.

4. Process modeling tools

Current software tools for activity analysis (such as NOLDUS [1], INTERACT[2] and MORAE[3]) do not meet our requirements in at least two ways (a review of such tools can be found in Hilbert and Redmiles (2000)). (a) Developed to analyze very detailed behavior data, such as a user interacting with a device, these tools typically only support sequential analyses spanning hours or days, as opposed to weeks, months, or years. (b) These tools also generally support data composed of low-level relatively simple events. They do not help the analyst manage the possibly evolving interpretation that he or she attributes to the events. Tools such as InfoScope[4], on the other hand, do provide high-level data visualization, but do not offer symbolic timeline analysis.

Concerning tools specifically developed to model trends in terrorist activity, we must cite the GTD Data Rivers tool developed by Lee (2008). The GTD Data Rivers is an interactive visual exploratory tool that allows analysts to investigate temporal trends in terrorism found in the GTD. The GTD Data Rivers aggregates important variables from the database and visualizes them as a comprehensible stack chart as shown in Figure 3.

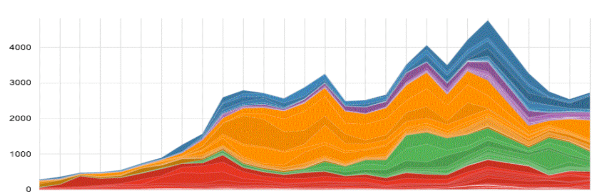

Figure 3: Number of events in the database differentiated by country (Lee, 2008).

[1] http://www.noldus.com/
[2] http://www.mangold-international.com/en/products/interact.html
[3] http://www.techsmith.com/morae.asp
[4] http://www.macrofocus.com/public/products/infoscope/

Figure 3 illustrates the rise and fall in the frequency of terrorist attacks for the years 1970 to 1996; the bands in this case represent targeted countries within six regions: Europe, Asia, South America, North America, Africa, and the Middle East. This tool enables us to analyze large chronological trends but it only supports numerical value visualizations, and does not support symbolic process modeling.

This review of tools helped us identify the need for a trace-modeling tool. These are summarized in Table 2.

Table 2: A specification for process-oriented trace-modeling tools.

Modeling specifications	Sub-requirements
Model past activities (produce a representation of an activity that has occurred about which we have information)	Display symbolically what we know about particular events across multiple levels of abstraction including: location, time, actors involved, unique characteristics, etc.
Modeling current ongoing activities (produce a representation of an ongoing activity that we hope to control and/or predict)	Enable analysts to dynamically identify new events, meaningful sequences of events, and relations between events in order to find signatures of sequences that may lead to predictions.
Support the development of counter-factual scenarios from "abstractions" of real events	From these scenarios, develop inferences that inform the prediction of future events and suggest preventative courses of action.

5. A tool for terrorism process modeling: ABSTRACT

To fulfill the requirements expressed in sections 3 and 4, we modified ABSTRACT [5], a trace-modeling tool that we have designed in previous work (Georgeon, Henning, Bellet, & Mille, 2007). ABSTRACT enables the analyst to define transformation rules to process raw qualitative or quantitative data streams into abstract activity traces. These abstract activity traces are based upon symbols that the analyst can define and organize in an ontology. Analysts can then visualize these traces and iteratively refine the ontology, the transformation rules, and the visualization format. This iterative process helps the analyst make sense of the initially overwhelming behavioral data. This process and tool have been used in a road safety study to find patterns of interest in data collected with an instrumented vehicle (Henning, Georgeon, & Krems, 2007). Figure 4 illustrates the aspects of this modeling process as they apply to the

[5] http://liris.cnrs.fr/abstract/

present study. This process involves 5 steps represented in blocks (1) through (5).

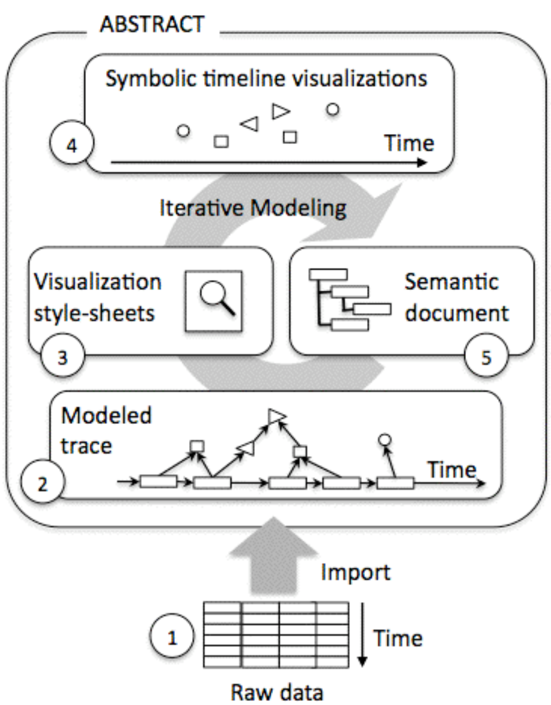

Figure 4: Process modeling with ABSTRACT.

(1): The raw data is usually stored in a spreadsheet where each line represents an event, and where the different properties of these events are recorded in columns.

(2): This data is imported into ABSTRACT under the form of a graph structure (RDF graph). In this graph, each event is a node. The analyst can add new events as new nodes during the modeling process. He or she can also add relations between nodes, including hypothetical causal relations that he or she asserts. In the figure, the geometrical shapes symbolize the events: rectangles, squares, circles, and triangles. The arrows represent the relations between events. Events also have properties attached to them as elements of the graph.

(3): The analyst defines style-sheets to render the modeled trace as *symbolic timeline visualizations*. These style-sheets are XSLT (eXtensible Stylesheet Language Transformation), a language for transforming XML documents into other XML documents.

(4): The timeline visualizations are SVG (Scalable Vector Graphics) documents that are displayed by any SVG compatible browser such as Firefox. We present an example of this visualization in Figure 5. ABSTRACT makes this visualization interactive—the user can both scroll the timeline, as well as click on events to show their properties and follow hypertext links to further documentation in a supporting wiki page.

(5): The analyst defines the types of events in the semantic documentation system. Within the system, he or she provides, on one hand, the textual documentation that explains each event category while on the other specifying the events' visualization properties, namely the geometrical shape, color, icon, and y position. Collectively, these event types form an event ontology that can appear in the traces. This ontology is exported as a RDFS graph (Resource Description Framework Schema). These graphs are then exploited by the style-sheets to render the visualization timeline.

To support the computational process modeling of terrorist activity, we modified ABSTRACT in two ways:

a): We implemented a server version that allows for concurrent modeling by multiple team members—typically a researcher in information sciences who focuses on tool and style-sheet development, and investigators in the domain of interest, in this case specialists in terrorism studies.

b): We have used a semantic wiki[iii] to implement ABSTRACT's ontologies and documentation system. Previous versions of ABSTRACT used Protégé[iv] as an ontology editor. Using semantic-media-wiki has several advantages. For one, the wiki principle offers a manageable and easy way for analysts to attach descriptions to event types. For another, wikis are sharable across the web and allow the construction of shared representations between different users. Finally, a semantic wiki supports the association of semantic properties to pages, in our case: a type/sub-type hierarchy and visualization properties.

6. Symbolic timeline representation of events collected from the field

Using ABSTRACT, we have obtained representations of terrorist activity like that shown in Figure 5. Figure 5 displays terrorist activity in the Republic of Ireland between 1970 and 2007 taken from 143 events. The upper half of this visualization represents a zoom consisting of a one hundred day interval, centered upon January 10, 1973. The lower half represents the entire (37 year) time-course. The interactive features of this representation are available online[v]. This visualization illustrates what we mean by symbolic timeline visualization and modeling. Unfortunately, this data does not include behind-the-scene information and does not inform us about the underlying processes that are happening. It is intended here as a demonstration of a method equally applicable to more detailed, and thus more illuminating, data.

In Figure 5, each event is represented by an icon and possibly a second icon appended to it. The first icon is associated with the field "WEAPON_TYPE". The three main weapon types are represented: "Firearms" (gun), "Explosive" (star) and "Incendiary" (flame). When the weapon type is unspecified, the event is represented as a gray circle. The second icon, representing a body outline, is appended when the "ATTACK" field is equal to "assassination".

The "y" position is associated with the field "PERPETRATOR". Meaning, the principal terrorist groups are each represented on a distinct line. Loyalist groups are represented above the central axis. Republican groups are represented below the central axis. Events whose affiliation is unknown are represented on the center axis.

The user can click on the event to show a tip window associated with it. The tip window displays the properties of the event. This tip window provides hypertext links to the definition of the different types in the semantic wiki. By following these links, the analyst can change the visualization properties as well as the textual explanations, before generate new timeline visualizations. The "GTD_ID" field gives a link to the GTD page that provides a comprehensive description of the event.

To illustrate the descriptive utility of this layout, let us consider the historical events associated with the Irish Troubles and how they are illustrated in Figure 5. For the group represented by the lower-most row on the y-axis (Group 11- the Irish republican Army), you'll notice that there are three sizeable lulls in activity toward the end of their campaign. After the second lull, there were two attacks that occurred in the first half of 1998. In April of 1998, several political parties (including Sinn Fein and its associated military force, the Provisional Irish Republican Army) came together to sign the Good Friday Agreement in an attempt to bring peace the Ireland, Northern Ireland, and the United Kingdom. Although Sinn Fein was a signatory to the Good Friday Agreement, it is possible that some individuals within the IRA were opposed to the peace process and engaged in activity contrary to its stipulations.

One limitation of the dataset employed here is the lack of

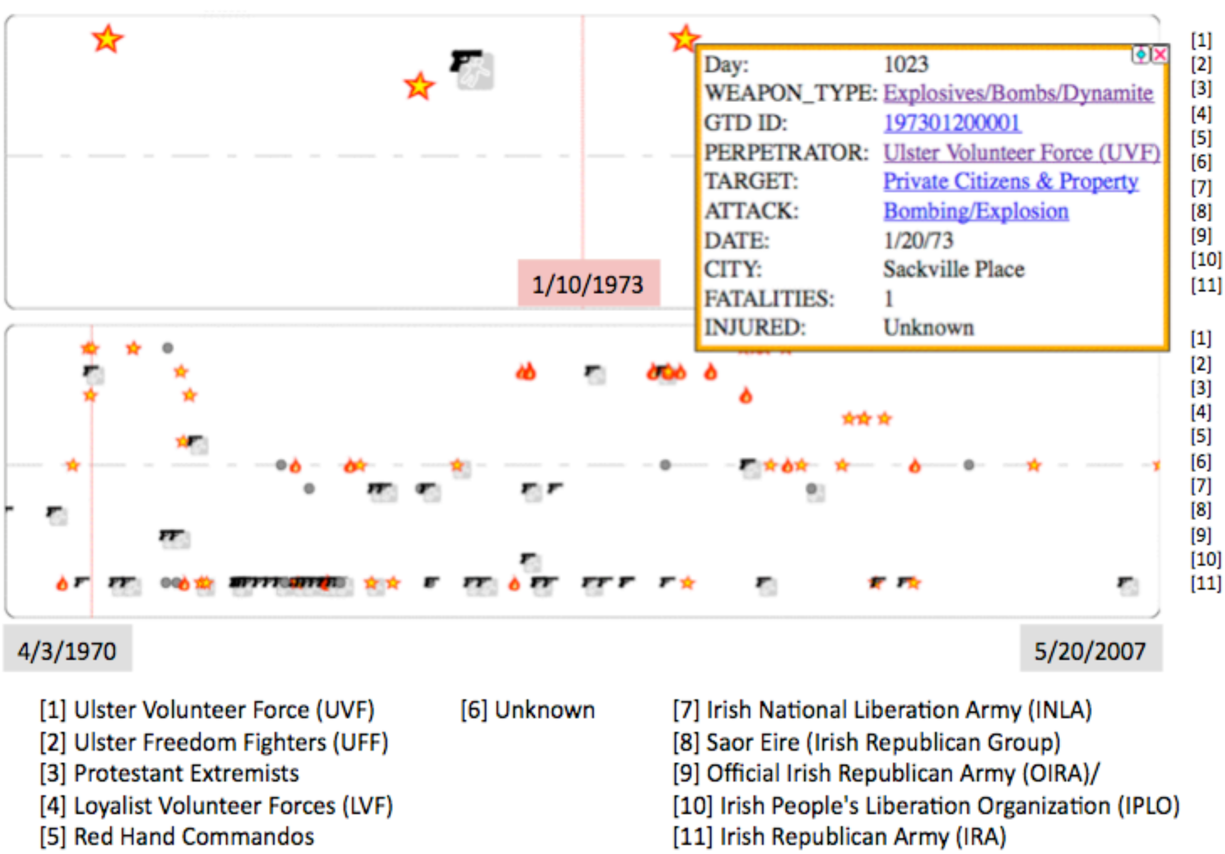

Figure 5: Terrorist activity in the Republic of Ireland (1970-2007) represented with ABSTRACT.

representation for other notable dissident groups. For example, one group that is vehemently opposed to the peace process is the Real Irish Republican Army (RIRA). In response to what they deemed to be Irish submission in the form of a peace deal, some members of the Provisional IRA broke off to form a more violent faction. This faction became known as the Real IRA. Had they been represented more comprehensively in the GTD, Figure 5 would illustrate the extent to which violence struck Ireland, Northern Ireland, and Britain in the wake of the GFA (post April, 1998). In the weeks and months following the signing of the GFA, the Real IRA conducted several operations, including bombings and mortar attacks. Despite its lack of representation in the GTD, data concerned with the activities of the Real IRA could be effectively illustrated with ABSTRACT. Doing so would (a) further illuminate the extent to which dissident and paramilitary activity has pervaded Ireland, Northern Ireland, and the rest of the UK in past decades, and (b) show the relationships between contextual events (e.g. signing of the GFA) and attacks by dissident groups or paramilitaries.

7. Discussion and Conclusion

We have yet to explore the full potential of this approach with data that would contain more information about the full process of terrorism activity. We may consider extensive detainee history such as published by Bruning and Alexander (2008) or terrorist narratives like those assembled by Sageman (2004). Our work on the GTD data provides a high-level, relatively abstract, description of the events contained within the database. As we obtain more data, we expect we will be able to more readily identify persistent signature patterns of activity, and connect the bottom-up modeling and the top-down modeling together. Using GTD data has allowed us to make a start in that direction and to identify important features for future process-oriented trace-based approaches. We have found having an online tool invaluable for not only capturing semantic content but also facilitating cooperation between team members from different origins, namely terrorism study and information sciences. In addition, our experiences modeling GTD events underscore the importance of analyst-driven tools that readily support the creation and placement of new symbolic representations that in turn support the visualization of salient differences. Finally, this approach allows the data to speak for itself by enabling the user to visualize timeline of events represented by symbols and providing links to complementary information.

We have examined an approach for modeling process, an approach that acknowledges and attempts to address the data/knowledge gap emerging across the social sciences. We specifically address the modeling of terrorist activity, however, we believe trace-based methods may be applicable to other domain areas where modeling emergence and reflexivity are important. For specialists in terrorism studies, we believe these methods will contribute to our understanding of data-rich processes and sub-process such as Improvised Explosive Devices (IED) development, online recruitment, and the movement of money and resources. We also believe that the insights we obtain from formalizing our understanding of the influence of low-level psychological and social factors may have implications for less tractable terrorist processes.

As we strive to deepen our understanding and formalize our knowledge, some analyses of processes describing events may integrate perspectives from a variety of contexts, others may focus on particular discipline or problem perspectives. It is possible that understanding some processes will necessarily draw on perspectives from particular disciplines or professions. The nature of the activity, the perspective taken, and the degree of conceptual complexity and understanding are all presumably variables that will affect the overall understanding of the phenomenon and its relationship to its environment and context.

The modeled traces that we obtain are sets of symbols and relations assembled as chronological representations. We must take these representations pragmatically (Wittgenstein, 1953), and assume that they are neither right nor wrong, neither true nor false: they are merely useful for the particular applications in which we apply them. These representations are also intended to evolve with our knowledge and with the data available. Our current level of analysis and the inherent assumptions we make about starting points for analysis and end products will influence further analysis.

We recognize the evolutionist and pragmatic aspect of this analysis, and attempt to support analysts operating in a variety of contexts and levels of analysis by synthesizing bottom-up and top-down approaches into a common framework. We, in fact, believe that a commitment to a pragmatic approach requires this from us while simultaneously obligating us to try to evaluate theory through the modeling of actual events. We believe this is not only possible but increasingly feasible as interdisciplinary communities cognoscente of data-mining and data-sharing tools emerge.

8. Acknowledgment

Support for this study was provided by ONR (contract N00014-06-1-0164 and N00014-08-1-0481) and DTRA (contract HDTRA 1-09-1-0054). We gratefully thank Dr

Frank E. Ritter for his useful comments on this study and report.

9. References

Bruning, J., & Alexander, M. (2008). *How to Break a Terrorist: The U.S. Interrogators Who Used Brains, Not Brutality, to Take Down the Deadliest Man in Iraq*: Free Press.

Georgeon, O. (2008). Analyzing traces of activity for modeling cognitive schemes of operators. *AISB Quarterly, 127*, 1-3.

Georgeon, O., Henning, M. J., Bellet, T., & Mille, A. (2007). *Creating Cognitive Models from Activity Analysis: A Knowledge Engineering Approach to Car Driver Modeling*. Paper presented at the International Conference on Cognitive Modeling, Ann Arbor, MI, 43–48.

Hedström, P. (2005). *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge, MA: Cambridge University Press.

Henning, M. J., Georgeon, O., & Krems, J. (2007). *The quality of behavioral and environmental indicators used to infer the intention to change lanes*. Paper presented at the 4th International Driving Symposium on Human Factors in Driver Assessment, Stevenson, Washington USA.

Hilbert, D. M., & Redmiles, D. F. (2000). Extracting Usability Information from User Interface Events. *ACM Computing Surveys (CSUR), 32*(4), 384–421.

Horgan, J. (2009). *Walking Away from Terrorism, Accounts of Disengagement from Radical and Extremist Movements*: Routledge.

Lee, J. (2008). Exploring Global Terrorism Data: A Web-based Visualization of Temporal Data. *ACM Crossroads, 15*(2), 7-16.

Popper, K. (1972). *Objective knowledge*. Oxford: Oxford University Press.

Sageman, M. (2004). *Understanding terror networks*. Philadelphia, PA: University of Pennsylvania Press.

Taylor, M., & Horgan, J. (2006). A Conceptual Framework for Addressing Psychological Process In the Development of the Terrorist. *Terrorism and Political Violence, 18*, 1-17.

Wittgenstein, L. (1953). *Philosophical Investigations*.

[i] http://www.start.umd.edu/gtd/

[ii] http://www.simile-widgets.org/timeline/

[iii] http://semantic-mediawiki.org/

[iv] http://protege.stanford.edu/

[v] http://acs.ist.psu.edu/~georgeon/abstract/View_v10_v11/tI01/p01/3447