Human Functions, Machine Tools, and the Role of the Analyst

Gordon R. Middleton
Patrick Henry College, grmiddleton@phc.edu

Human Functions, Machine Tools, and the Role of the Analyst

Abstract

In an era of rapidly increasing technical capability, the intelligence focus is often on the modes of collection and tools of analysis rather than the analyst themselves. Data are proliferating and so are tools to help analysts deal with the flood of data and the increasingly demanding timeline for intelligence production, but the role of the analyst in such a data-driven environment needs to be understood in order to support key management decisions (e.g., training and investment priorities).

This paper describes a model of the analytic process, and analyzes the roles played by humans and machine tools in each process element. It concludes that human analytic functions are as critical in the intelligence process as they have ever been, and perhaps even more so due to the advance of technology in the intelligence business.

Human functions performed by analysts are critical in nearly every step in the process, particularly at the front end of the analytic process, in defining and refining the problem statement, and at the end of the process, in generating knowledge, presenting the story in understandable terms, tailoring the presentation of the results of the analysis to various audiences, as well as in determining when to initiate iterative loops in the process.

The paper concludes with observations on the necessity of enabling expert analysts, tools to deal with big data, developing analysts with advanced analytic methods as well as with techniques for optimal use of advanced tools, and suggestions for further quantitative research.

Introduction

"Computers are useless. They can only give you answers."
> — *Pablo Picasso*

"Before you become too entranced with gorgeous gadgets and mesmerizing video displays, let me remind you that information is not knowledge, knowledge is not wisdom, and wisdom is not foresight. Each grows out of the other, and we need them all." (*emphasis added*)
> — *Arthur C. Clarke*

Data are proliferating and so are tools to help analysts deal with the flood of data and the increasingly demanding timeline for intelligence production. So where does that leave the analyst? What is the role of the analyst in such a data-driven environment? How important is the analyst's role and where? And what are the implications for organizations, analysts, and users of intelligence?

This article will suggest answers to these important questions: And these are important questions. Having served inside several of the intelligence agencies before and immediately after 9/11, the author observed firsthand the rush to 'throw tools at the analytic problem.' To be fair, this tendency did not begin with the post-attack rush to 'fix intelligence', but is part of the longer standing US approach to most problems—reorganizing,[1] and then 'throwing money at the problem'.[2] It also is part of the information revolution taking place in business and commercial applications that has spawned many innovative and often costly technical solutions ('machine tools' for purposes of this paper) to deal with Big Data.[3]

One of the results of this recent emphasis on technical solutions to the challenge of big data has been "tool burnout." Most organizations have had

---

[1] Posner, Richard A., *Preventing Surprise Attacks: Intelligence Reform in the Wake of 9/11* (New York: Rowman & Littlefield, 2005), p. 127.

[2] For example, the Army, Air Force and Navy each operate their own versions of the Defense Common Ground System (DCGS), designed to share intelligence data from many sources. It has been estimated to cost more than $10 billion. See, Mark Pomerleau, "Resistance to DCGS-A persists," March 31, 2015, available at: *http://gcn.com/articles/2015/03/31/ dcgs-a-palantir.aspx.*

[3] IBM, Microsoft, Oracle, SAP, and others each offer hardware and software tools for data-integration and database-management (DBMSs), as well as business intelligence and analytics software. For example, see: Doug Henschen, "Sixteen Top Big Data Analytics Platforms," *Information Week*, January 30, 2014; and Peter Wayner, "Seven Top Tools for Taming Big Data," *InfoWorld,* April 18, 2012.

their favorite contractor and each had their proprietary tool set and wanted to demonstrate how it could find the proverbial needle in the haystack, especially in the pressure for improved intelligence after the events of 9/11. Only, it had to be their haystack; and their needles; and under carefully scripted conditions, etcetera. After being sold what appeared to be a bill of goods that on multiple occasions did not deliver, cynicism toward advanced tools has begun to appear, which illustrates the need for improved understanding of the relationship between human functions and advanced machine tools.[4]

However understandable that sentiment may have been, the flood of data mandates that analytic tools will play a significant role in the intelligence process.[5] The timeliness of our intelligence requirements, coupled with the data volume, mandate some level of automation or computer assistance in the analytic process. We will never have nearly enough analysts to manually sort through the data, much less analyze and report on it in a timely fashion. The question to be addressed is: What is the role of analysts and what is the relationship of analysts to machine tools, especially in an era of Big Data? To get at the answer to this question, we must first look at the analytic process itself in order to gain perspective on the interplay of human functions and machine functions. Defining the process and understanding where and how humans and machine tools interact with that process may assist in framing questions for further quantitative research.

Analysis Process

In the most general terms, the analysis process has been described as (1) defining the issue; (2) assembling the data; and (3) generating knowledge. While this basic model of the analytic process may provide some insights to the role of the analyst, a more detailed model with additional granularity is required to help address the questions posed in this paper about the role of analysts. Figure 1, below, presents a model of the analytic process that is adapted from the milestone publication on visual analytics, edited by James J.

---

[4] The author heard one senior executive in intelligence exclaim, "If I never see another tool demo, it will be too soon!"

[5] Matt McConnell, "8 Problems with Data Overload," Wired Blog, December 4, 2013, available at: *http://insights.wired.com/profiles/blogs/8-problems-of-data-overload#axzz3lIHTn6xd*; "Too much information: How to cope with data overload," *The Economist,* June 30, 2011, available at: *http://www.economist.com/node/18895468;* oods, David D.; Patterson, Emily S.; and Roth, Emilie M., "Can We Ever Escape From Data Overload? A Cognitive Systems Diagnosis," Cognition, Technology and Work, in press, available at: *http://csel.eng.ohio-state.edu/productions/woodscta/media/diagnosis.pdf.*

Thomas and Kristin A. Cook in 2004, *Illuminating the Path: R&D Agenda for Visual Analytics.*

This model (Figure 1) reflects that the analytic process includes major elements associated with (1) framing the issue; (2) assembling the data; and (3) multiple, iterative loops for data foraging, sense-making, story-telling, and reevaluation.[6] This model also provides the opportunity to compare human functions and those of machine tools throughout the process.[7] The following paragraphs define and provide examples of the functions performed by humans and information machine tools at each of the major elements in the model of the analytic process in Figure 1 (Defining the Issue; External Data

Figure 1. Model of Analysis & the Complementary Roles of Human Cognition and Machine Tools

Adapted from Illuminating the Path: R&D Agenda for Visual Analytics. Edited by Thomas, James J. and Cook, Kristin A. 2004

---
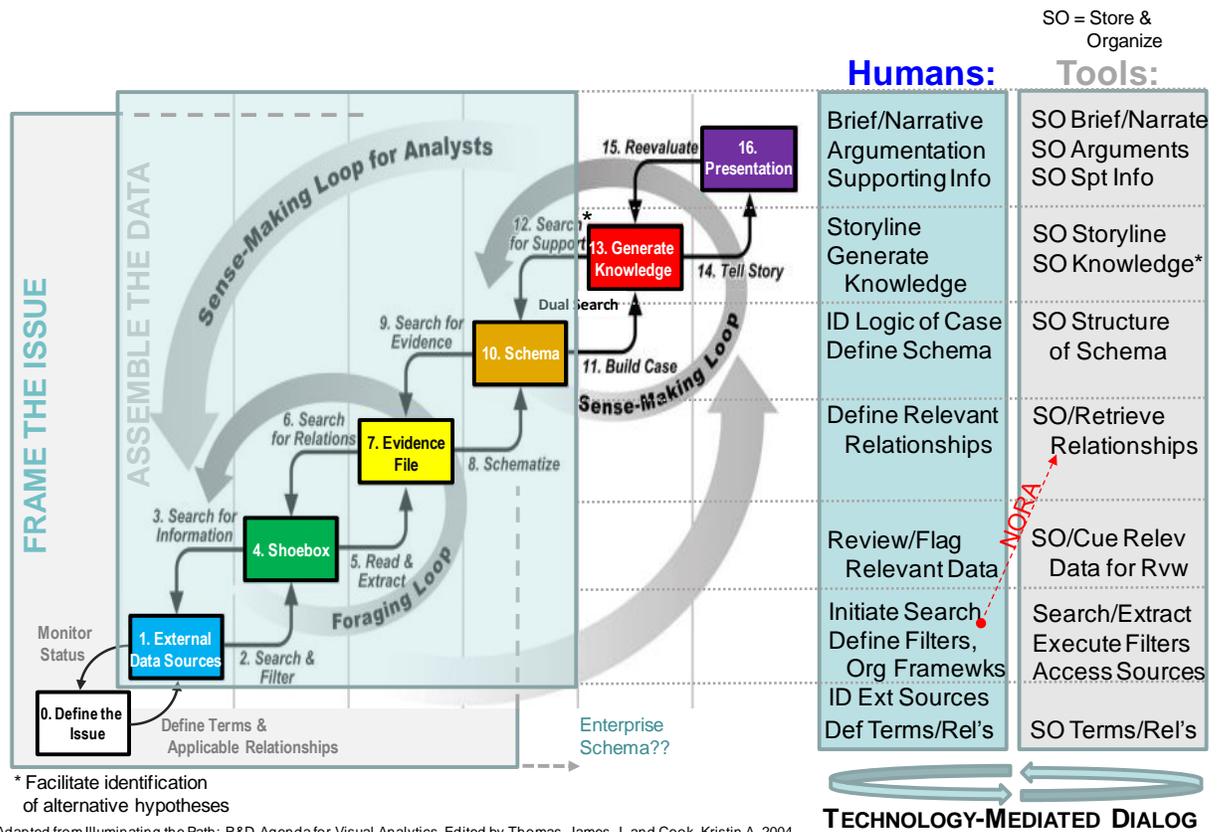
[6] Thomas, James J. and Cook, Kristin A. (eds)*., Illuminating the Path: R&D Agenda for Visual Analytics* (Pacific Northwest National Laboratory: National Visualization and Analytic Center, 2005), available at: *http://vis.pnnl.gov/pdf/RD_Agenda_VisualAnalytics.pdf.*

[7] Gordon R. Middleton, "Avatars or Robots? The Human Factor in Overcoming Information Overload," Presented at the Military Operations Research Symposium (MORS), June 2010.

Sources; Shoebox; Evidence File; Schema; Generate Knowledge; and Presentation).[8] In so doing, these descriptions exemplify the technology-mediated dialog that occurs between the human analysts in the process and the machine tools and provide a basis for more explicitly understanding these roles and interactions. They may also help identify where the relative advantages exist for human functions and information machine tools.

Define the issue. This process initiates the analysis by defining terms and primary applicable relationships. As Clark underscores, "the first and most important step an analyst can take is to understand the problem in detail."[9] It may also involve supporting efforts to define "enterprise ontologies" for use in tagging or indexing broad information sources for applicability to a range of identified topics of interest (e.g., the National Intelligence Priorities Framework, aka "NIPF").

Human Functions. The analyst plays the key role in understanding the problem and provides the foundation of understanding the problem topic well enough to know the question or questions that were not asked, but perhaps should have been included in the problem statement. The individual analyst in charge of the analytic process establishes the definitions, primary relationships, and even identifies other organizations or individuals that will need to participate in or review the specific analysis methodology and results. This may not be driven by a single analyst, but may be the result of other human-to-human interactions and organizational dynamics (e.g., team dynamics or organizational unit decisions).

Machine Tools. In some circumstances, computerization of multiple databases and information sources may make some tools useful at this stage. The analyst may also consult so-called, enterprise ontologies,[10] for applicability to the specific problem. However, ongoing issues of data and information system interoperability and security access make use of tools at this stage of the process effective only after what is sometimes extensive human intervention. The primary role of tools at this stage is to store and help organize the results of human interactions and human functions in

---

[8] Only the major process elements are addressed in this paper. Further insights might be obtained from exhaustively examining all of the intermediate steps and each of the iterative loops (e.g., 2 Search & Filter; 5 Read and Extract; etc) but that more detailed analysis is beyond the scope of this paper.

[9] Clark, Robert M., *Intelligence Analysis: A Target-Centric Approach* (Washington, D.C.: CQ Press, 2013), p. 20.

[10] Commonly used in the intelligence agencies as the collective set of tags or topical labels to identify the topics of interest to which a document or source may be applicable.

defining and understanding the problem. This may take the form of saved documents, briefings, emails, lists, initial reference documents, terms of reference, or other preliminary information.

Data sources. Once the problem, primary terms, and relationships are identified, the analysis process moves to searching and filtering available sources external to the data already in the possession of the analyst.

Human Functions. The analyst defines search filters and key terms, network, or organization associated with the problem; organizational frameworks (where to search for relevant information and where to store the results); and initiates search activities (in some instances this may include interacting with a legal process to obtain authorization for collection of new data or to access existing special collections or with a collection management process to stimulate additional collection).

Machine Tools. The actions of the analyst initiate machine access to identified and authorized data sources (including big data, social media, or social network sources) and may include search, extraction, execution of data filters of various types using the key terms and other search methodologies. The primary role of tools at this stage is to store and help organize the results of human interactions and human functions in defining the key terms, network, or organization associated with the problem and to implement them in the search process. This may take the form of saved documents, lists, terms, network relationships, or organizational descriptions.

Shoebox. The results of these searches are returned to the analyst's work area, sometimes referred to as a "shoebox." This terminology dates from the practice of some analysts using actual shoe boxes to help organize and store index cards with information on topics of interest related to official research or of more general, personal interest related to the analyst's area of responsibility.

Human Functions. The analyst reviews these initial search results and flags clearly relevant data, and may also categorize other information of lesser or questionable relevance for possible follow-up at a later time. The intuitive or abductive capabilities of the analyst may be a critical element in the process at this juncture by identifying seemingly extraneous information, but which may become critical elements in the final analysis.

73

Machine Tools. The actions of the analyst initiate machine tools to sort and store information in relevance categories. Tools at this stage store and help organize the results of human interactions and human functions in validating the relevance of initial results of the search process. Increasingly, tools are becoming capable of translating the human inputs in the prior stage into cues as to the relevant data for priority human review. This includes capabilities for "non-obvious relationship analysis" (NORA),[11] often associated with multi-tiered relationship analytics. The actions of the analyst layers additional information and the primary role of tools at this stage is to store and help organize the results of the initial search results for future use. Like the previous stage, this may take the form of saved annotations, notes, comments on priorities for later action, or other ideas from the mind of the analyst associated with documents, lists, terms, network relationships, or organizational descriptions, or other search data.

Evidence file. Once the analyst has read and extracted applicable information, these results are further placed in an evidence file. This information is directly related to the question under review.

Human Functions. The analyst identifies and defines relevant relationships and data for detailed analysis. The human functions performed by the analyst at this stage are particularly important – and subject to biases and distracting or corrupting influences that are particularly 'human' in their nature.[12]

Machine Tools. The actions of the analyst initiates tools at this stage to store and help organize and retrieve the results of search results for analysis and review, as well as to surface non-obvious relationships, as above. As with previous stages, this may take the form of saved annotations, notes, comments on priorities for later action, or other ideas from the mind of the analyst associated with documents, lists, terms, network relationships, or organizational descriptions, or other search data.

---

[11] P.M. Kogge, "Comparative performance analysis of a Big Data NORA problem on a variety of architectures," *IEEE International Conference, Collaboration Technologies and Systems (CTS)*, May 2013; NORA is an acronym for Non-Obvious Relationship Awareness, a technology that mines data resources to determine the relationships between people. Non-Obvious Relationship Awareness was created by Systems Research and Development (SRD). SRD developed this technology for the Las Vegas gaming industry to help the casinos detect relationships between customers and parties named by the Nevada Gaming Control Board as excluded persons. SRD has been acquired by IBM.
[12] Heuer Jr., Richards J., *Psychology of Intelligence Analysis* (Central Intelligence Agency: Center for the Study of Intelligence, 1999).

Schema. To help organize and assist making sense of all relevant data, analysts will often develop a structured approach to organizing the relevant data, called a "schema." The schema may take many different forms, including a map, theory, concept, paradigm, or guiding image.[13]

Human Functions. Analysts identify the logic of the case, to include such factors as confidence levels required of the evidence in order to be used in court or to be assessed as conclusive; and define the schema for organizing available data and for structuring analysis of the data.

Machine Tools. Tools at this stage store the data in the schema format defined by the analysts or indexed to the schema and help retrieve the results of search results for analysis and review. As with previous stages, this may take the form of saved annotations, notes, comments on priorities for later action, or other ideas from the mind of the analyst associated with documents, lists, terms, network relationships, or organizational descriptions, as well as indexed data and index metadata, or other search data.

Generate knowledge. In turn, the schema (or data model) helps the analyst generate knowledge, through systematic organization of the relevant information that is known. It may also help reveal what information is not known and imply the need for additional search or collection activities.

Human Functions. Analysts create briefings, narrative reports, media storyboards; develop the line of argumentation for logic of presentations, and identify supporting information. Analysts use hypotheses and available data to create a storyline that describes the actors, their relationships, motivations or objectives to address the question or topic under analysis. Analysts generate knowledge through reasoning – using inductive reasoning (combining separate fragments of information, to form general rules or conclusions), deductive reasoning (applying general rules to specific problems to arrive at conclusions), or abductive processes (non-linear insight or intuition to generate novel hypotheses).[14] The process of generating knowledge also includes searching for information that may falsify or support various hypotheses about the issue under consideration.[15] Analysts are also uniquely responsible for turning knowledge into wisdom and insight—and

---

[13] Clark uses the term "model," as an alternative term for a data schema. Ibid, p. 32.
[14] Moore, David T. and Krizan, Lisa, "Core Competencies for Intelligence Analysis at the National Security Agency," in Russell G. Swenson (ed.) *Bringing Intelligence About* (Joint Military Intelligence College, 2003), pp. 110-111.
[15] Heuer Jr., Richards J., and Randolph H. Pherson, *Structured Analytic Techniques for Intelligence Analysis* (Washington, D.C.: CQ Press, 2010).

75

occasionally into foresight (as referenced in the second quote at the opening of this paper, by Arthur C. Clarke).

Machine Tools. Tools at this stage store or animate the storyline and other knowledge products. This may take the form of additional annotations, notes, or other ideas from the mind of the analyst associated with documents, lists, terms, network relationships, or organizational descriptions, as well as draft narrative reports, storyboards, briefings, or media presentations.

Presentation. Getting to the 'right answer' is only half the challenge facing analysts (as alluded to in the opening quotation from Picasso). In addition to asking the right questions, conveying the critical results of the analyses to the right place or organization, or the right person, in a timely manner, and in the most appropriate or desired form or format are key to decision makers actually understanding and possibly even acting on the intelligence they've received. Story telling is one of the analyst's greatest abilities in effectively conveying the results of all the preceding steps, whether the form it takes is written narrative, audio-visual media centered, or some blend of these with personal, verbal presentation.

Human Functions. Analysts tailor presentations of their analysis (briefings, narrative reports, media storyboards; line of argumentation in their presentation) based on the individual audiences to whom they present their results.

Machine Tools. Tools at this stage store multiple versions of complex presentations of analytic products developed by the analysts, which reflect the nature and interests of the individual audiences. These may take the form of additional annotations, notes, or other ideas from the mind of the analyst associated with documents, lists, terms, network relationships, or organizational descriptions, as well as draft narrative reports storyboards, briefings, or media presentations.

Iterative Loops. As alluded earlier, these process elements are not done in an exclusively linear manner, but involve multiple, iterative loops for data foraging, sense-making, story-telling, and reevaluation.

Human Functions. The analysts and their supervisory chain are the major arbiter of when and under what circumstances these iterations are accomplished. After presenting their results, feedback to the analysts may suggest they reevaluate certain aspects of their analysis or modify the

methodology for presenting the information (story telling), which may also depend on the specific audience. Analysts may also assess the need for additional or different research to support or falsify new hypotheses that result from the knowledge generation step. They may also come to believe additional evidence may be required due to changes in the schema or model, resulting from new or changed hypotheses. This in turn may convince analysts they should search for other relationships or of the need to select individuals who may need to be added to the social network associated with the topic. These iterations may accumulate to result in some reconsideration of the basics of the problem formulation and even the form and content of the question or issue being analyzed.

Machine Tools. Tools store the additional or updated data associated with the iterations to accomplish the functions at each step in the process as described, above.

Discussion

The above description of the elements in the analytic process, along with an indication of the role of humans and machine tools, provide a basis for answering the questions about the role of the analyst in such a data-driven environment; the relative importance of the analyst's role and how that is different in the various stages of the process; and the implications for organizations, analysts, and users of intelligence of these interactions and differences. Figure 2 provides a graphical overview of the analytic process, the key capabilities bearing on those process elements, and the key capabilities for each process element based on the above analysis. This analysis conceptually demonstrates that humans and information machine tools each have important roles to play in the elements of the analysis process.

Human functions are critical elements in nearly every step in the process, especially at the front end of the process in defining and refining the problem statement, and at the end of the process, in generating knowledge, in tailoring the presentation of the results of the analysis to various audiences, and in determining when to initiate iterative loops in the process. In most of the elements, it is not an 'either-or' issue between analysts and machine tools, but rather there is a critical dialog between the analysts and the machine tools to produce the most effective results

Machine tools play a particularly key role in middle elements in the process, particularly where scaling is important. This role is one that is especially

77

important in big data environments. Advances in automated indexing may be a particularly useful capability development in emerging tool capabilities.

Figure 2. Summary of Key Capabilities Contributing to Each Analytic Process Element
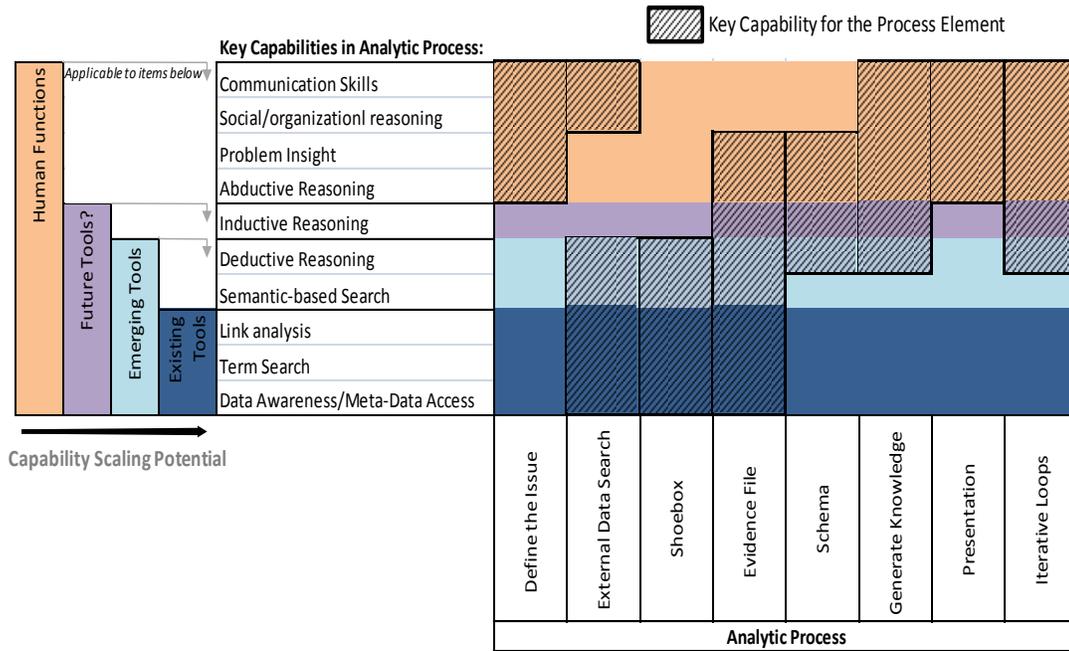
Conclusions and Implications

This paper provides rationale for the view that analysts are as critical in the intelligence process as they have ever been, and perhaps even more so due to the advance of technology applications in intelligence. The flood of data available from open sources and from highly classified sources makes the effective integration of the human role performed by analysts in the collection phase particularly critical, if the increasingly sophisticated tools are to be used effectively. Implications from these circumstances include:

Enabled, Expert Analysts. Timely, insightful analysis to protect U.S. national security requires exceptionally enabled and expert analysts to accomplish their significant challenges in the process due to the high dependence in the process on human functions.

Tools to Deal with Big Data. Timely, insightful analysis to protect U.S. national security requires the best machine tools for analysts in order to deal with the flood of data. Advanced machine tools, like NORA and other commercial tools, can provide significant analytic power to analysts,

78

especially in a big data environment and need to be fully integrated into the national security analytic environment.

Balanced Training/Development. Organizations need to pursue balanced investments to train and develop their analysts with advanced analytic and reasoning methods associated with the front and back ends of the analytic process, as well as with techniques for optimal use of advanced tools in the middle elements of the process.

Future Research. Quantification of the benefits and costs of big data tools is a critical step that is required to support key investment and training decisions in intelligence. Such quantification may provide particular benefit if studies would adopt a common functional taxonomy for the analytic process, such as proposed in this paper. This would help focus future testing on key elements and permit more direct comparison of results. Specific assessments across the spectrum of commercially available tools applicable within each of the various elements of the analytic process could shed light on the relative cost-benefit of increased training of analysts versus implementation of such advanced tools. Structured testing against an analysis model would also help assess the relative effectiveness of tools within each analytic element. This research approach would significantly aid in the further understanding of the analytic process and may provide insights to relative advantage for investment in further developing the skills of human analysts or investment in more advanced tools. Research along these lines would have high potential to provide practical insight to government and business managers in planning their investments in advanced information tools and training of their analysts and could be highly cost-effective for government and business enterprises.