Measuring Attributes of Rigor in Information Analysis

Daniel J. Zelik, Emily S. Patterson, and David D. Woods

Introduction

Information analysis describes the process of "making inferences from available data" and determining "the best explanation for uncertain, contradictory, and/or incomplete data" (Trent, Patterson, & Woods, 2007, p. 76). It is an accumulation and interpretation of evidence to support decision making (Heuer, 1999; Schum, 1987). And it represents an active, goal-directed, and often technologically mediated, sensemaking activity (Klein, Phillips, Rall, & Peluso, 2007).

This macrocognitive framing of information analysis as "sensemaking" (Klein, in press; Klein, Moon, & Hoffman, 2006a, 2006b) highlights that cognition employs narrative elements to recognize and communicate meaning, and that sensemaking—as with other macrocognitive functions such as (re)planning—is (1) joint activity distributed over time and space, (2) coordinated to meet complex, dynamic demands, (3) conducted in an uncertain, event-driven environment, (4) often performed with simultaneously complementary and conflicting goals with high consequences for failure, (5) made possible by effective expertise in roles, and (6) shaped by organizational constraints (see Patterson & Miller, in press; Klein et al., 2003; Schraagen, Klein, & Hoffman, 2008).

Thus, information analysis in both professional domains such as intelligence analysis and more everyday contexts like online shopping is a deliberate process of collecting data, reflecting upon it, and aggregating those findings into knowledge, understanding, and the potential for action. However, even in contexts where there are relatively low consequences for inaccurate conclusions, there lurk ever-present vulnerabilities exacerbated by ever increasing data availability and connectivity (Woods, Patterson, & Roth, 2002; Treverton & Gabbard, 2008; Betts, 2002, 1978). Of particular concern is the potential inherent in all analytical activity that an analysis process is prematurely concluded and is subsequently of inadequate depth relative to the demands of a situation—a vulnerability we characterize as the risk of shallow analysis.

This chapter examines rigor as a measure of sensemaking in information analysis activity that captures how well an analytical process reduces this risk. It describes eight attributes of analytical rigor that were synthesized from empirical studies with professional intelligence analysts from a variety of specialties and agencies (Patterson, Roth, & Woods, 2001; Miller, Patterson, & Woods, 2006; Trent, Patterson, & Woods, 2007; Zelik, Patterson, & Woods, 2007; Grossman, Woods, & Patterson, 2007). Examples are then provided that illustrate how these attributes were calibrated for two different studies of analytical activity. Finally, potential future directions are discussed for using the attributes to assess, negotiate, and communicate the rigor of information analysis processes.

Analytical Rigor as a Measure of Process

Across domains, conventional perspectives for measuring analytical rigor focus on identifying gaps between what was actually done versus a prescribed or "standard" method, with rigor variously defined as the "scrupulous adherence to established standards" (Crippen et al., 2005, p. 188), the "application of precise and exacting standards" (Military Operations Research Society, 2006, p. 4), "methodological standards for qualitative inquiry" (Morse, 2004, p. 501), and the "unspoken standard by which all research is measured" (Davies & Dodd, 2002, p. 280).

Unfortunately, such definitions suggest a conceptualization of analytical activity that is neither particularly likely to reflect rigorous analysis work as practiced (Dekker, 2005; Sandelowski, 1993, 1986), nor particularly useful in developing the concept of rigor as a measure of performance. The diverse nature of intelligence analysis in particular renders the identification of a single "standard" process an intractable, if not impossible task (Berkowitz & Goodman, 1991; Krizan, 1999; Marrin, 2007). Consequently, rather than on a standards-based notion of rigor, our measurement approach focuses on how the risk of shallow analysis is reduced via analyst-initiated strategies that are opportunistically employed throughout the analysis process. These strategies are alternatively conceptualized as "broadening" checks (Elm et al., 2005) insofar as they tend to slow the production of analytic product and make explicit the sacrifice of efficiency in pursuit of accuracy, a central tenet of the framework.

Stories of Shallow Analysis

In our research with professional information analysts, study participants were often asked to directly evaluate or critique the products and processes of other analysts (e.g., Miller, Patterson, & Woods, 2006; Zelik, Patterson, & Woods, 2007). Across these studies, consistent patterns emerged as experienced analysts identified critical vulnerabilities in information analysis by way of "stories" describing how analytical processes often can or did go wrong. We organize these stories into eight inter-related risks of shallow analysis.

Specifically, we found that shallow analysis:

1. *Is structured centrally around an inaccurate, incomplete, or otherwise weak primary hypothesis, which analysts sometimes described as favoring a "pet hypothesis" or as a "fixation" on an initial explanation for available data.*

2. *Is based on an unrepresentative sample of source material, e.g. due to a "shallow search", or completed with a poor understanding of how the sampled information relates to the larger scope of potentially relevant data, e.g. described as a "stab in the dark".*

3. *Relies on inaccurate source material, as a result of "poor vetting" for example, or treats information stemming from the same original source as if it stems from independent sources, labeled variously as "circular reporting", "creeping validity", or as the "echo chamber" effect.*

4. *Relies heavily on sources that have only a partial or, in the extreme, an intentionally deceptive stance toward an issue or recommended action, often characterized by analysts in terms of "biased", "slanted", "polarized", or "politicized" source material.*

5. *Depends critically on a small number of individual pieces of often highly uncertain supporting evidence proving accurate, identified by some individuals as an analysis heavily dependent upon "hinge evidence" or, more generically, as a "house of cards" analysis.*

6.   Contains portions that contradict or are otherwise incompatible with other portions, e.g. via the inclusion of lists or excerpts directly "cut and paste" from other documents or via an assessment that breaks an issue into parts without effectively re-integrating those parts.

7.   Does not incorporate relevant specialized expertise, e.g. an analyst who "goes it alone", or, in the other extreme, one who over relies on the perspectives of domain experts.

8.   Contains weaknesses or logical fallacies in reasoning from data to conclusion, alternatively described as having a "thin argument", a "poor logic chain", or as involving "cherry picking" of evidence.

In addressing each of these sources of risk, eight corresponding attributes of analytical rigor were identified (see Figure 1). Together these attributes represent a model of how experts critique analytical process, rather than a model of the analytical activities of those experts. Consequently, these attributes make salient the aspects of process that are attended to in making an expert judgment of analytical rigor.

Our approach to operationalize the model involves translating empirically based descriptions of each attribute into differentiating categories of analytical behaviors. Each attribute is thus characterized in terms of low, moderate, and high indicators of rigorous process. Using this rubric, analytical processes can be rated or scored by comparing the methods employed during a given analysis to the categorical descriptions that appear in Figure 1. Scoring an analysis process across all attributes, in turn, reflects an overall assessment of analytical rigor.

Generally, if no broadening checks are observed in relation to a particular aspect of risk, we code it as "low" since analysts, both casual and professional, typically show some awareness of the possible risks of being inaccurate in an assessment with respect to a given attribute. For "moderate" rigor, encompassed are activities that display an active response to a particular type of risk, while strategies that are particularly effortful, systematic, or exhaustive are coded as "high." Finally, aspects of an analysis that are not applicable to a given context are generally not scored, for example attributes involving collaboration on a task where an analyst was expected to work alone.

| Attribute | Indicators of... | | |
Description	LOW Rigor	MODERATE Rigor	HIGH Rigor
Hypothesis Exploration The construction and evaluation of potential explanations for collected data.	– Little or no consideration of alternatives to primary or initial hypotheses. – Interpretation of ambiguous or conflicting data such that they are compatible with existing beliefs. – Fixation or knowledge shielding behaviors.	– Some consideration of how data could support alternative hypotheses. – An unbalanced focus on a probable hypothesis or a lack of commitment to any particular hypothesis.	– Significant generation and consideration of alternative explanations via the direct evaluation of specific hypotheses. – Incorporation of "outside" perspectives in generating hypotheses. – Evolution and broadening of hypothesis set beyond an initial framing. – Ongoing revision of hypotheses as new data are collected.
Information Search The focused collection of data bearing upon the analysis problem.	– Failure to go beyond routine and readily available data sources. – Reliance on a single source type or on data that are far removed from original sources. – Dependence upon "pushed" information, rather than on actively collected information. – Use of stale or dated source data.	– Collection from multiple data types or reliance on proximal sources to support key findings. – Some active information seeking.	– Collection of data from multiple source types in addition to the use of proximal sources for all critical inferences. – Exhaustive and detailed exploration of data in the relevant sample space. – Active approach to information collection.
Information Validation The critical evaluation of data with respect to the degree of agreement among sources.	– General acceptance of information at face value, with little or no clear establishment of underlying veracity. – Lack of convergent evidence. – Poor tracking and citation of original sources of collected data.	– Use of heuristics to support judgements of source integrity, e.g. relying on sources that have previously proven to be consistently accurate. – A few "key" high-quality documents are relied on heavily. – Recognizes and highlights inconsistencies between sources.	– Systematic and explicit processes employed to verify information and to distinguish facts from judgements. – Seeks out multiple, independent sources of converging evidence. – Concerned both with consistency between sources and with validity and credibility within a given source.
Stance Analysis The evaluation of collected data to identify the relative positions of sources with respect to the broader contextual setting.	– Little consideration of the views and motivations of source data authors. – Recognition of only clearly biased sources or sources that reflect a well-defined position on an issue.	– Perspectives and motivations of authors are considered and assessed to some extent. – Incorporates basic strategies to compare perspectives of different sources, e.g. by dividing issues into "for" or "against" positions.	– Involves significant research into, or leverages a preexisting knowledge of, the backgrounds and views of key players and thought leaders. – May involve more formal assessments of data sources, e.g. via factions analysis, social network analysis, or deception analysis.
Sensitivity Analysis The evaluation of the strength of an analytical assessment given possible variations in source reliability and uncertainty.	– Explanations are appropriate and valid at a surface level. – Little consideration of critical "what if" questions, e.g., "What if a given data source turns out to be unreliable?" or "What if a key prediction does not transpire as anticipated?"	– Considers whether being wrong about some inferences would influence the overall best explanation for the data. – Identifies the boundaries of applicability for an analysis.	– Goes beyond simple identification to specify the strength of explanations and assessments in the event that individual supporting evidence or hypotheses were to prove invalid or unreliable. – Specifies limitations of the analysis, noting the most vulnerable explanations or predictions on which the analysis is at risk of erring.
Information Synthesis The extent to which an analyst goes beyond simply collecting and listing data in "putting things together" into a cohesive assessment.	– Little insight with regard to how the analysis relates to the broader analytical context or to more long-term concerns. – Lack of selectivity, with the inclusion of data or figures that are disconnected from the key arguments or central issues. – Extensive use of lists or the restatement of material copied directly from other sources with little reinterpretation.	– Explicit, though perhaps not systematic, efforts to develop the analysis within a broader framework of understanding. – Depiction of events in context and framing of key issues in terms of tradeoff dimensions and interactions. – Provides insight beyond what is available in the collected data.	– Extracted and integrated information in terms of relationships rather than components and with a thorough consideration of diverse interpretations of relevant data. – Re-conceptualization of the original task, employing cross-checks on abstractions. – Performed by individuals who are "reflexive" in that they are attentive to the ways in which their cognitive processes may have hindered effective synthesis.
Specialist Collaboration The extent to which substantive expertise is integrated into an analysis.	– Minimal direct collaboration with experts. – Little if any on-topic, "outside" expertise is accessed or sought out directly.	– Involves some direct interaction with experts, though usually via readily available specialists. – Expertise is drawn from within preexisting personal or organizational networks.	– Independent experts in key content areas are identified and consulted. – Efforts to go beyond a "core network" of contacts to seek out domain-relevant experts, with additional resources and "political capital" potentially expended to gain access to such specialist expertise.
Explanation Critiquing The critical evaluation of the analytical reasoning process as a whole, rather than in the specific details.	– Few if any instances of alternative or "outside" criticisms being considered. – Reliance on preexisting channels of critiquing, primarily those supervisory.	– Brings alternative perspectives to bear in critiquing the overall analytical process. – Leverages personal or organizational contacts to examine analytical reasoning, e.g. by way of peer analysts, proxy decision makers, etc.	– Familiar as well as independent perspectives have examined the chain of analytical reasoning, explicitly identifying which inferences are stronger and weaker. – Use of formal methods such as "red teams" or "devil's advocacy" to challenge and vet hypotheses and explanations. – Expenditure of capital, political or otherwise, in critiquing the analytical process.

Figure 1. Attributes of rigorous analysis organized with respect to indicators of low, moderate, and high rigor.

Attributes of Analytical Rigor

1.   Hypothesis Exploration. Hypotheses are among the most basic building blocks of analytical work, representing candidate explanations for available data (Elm et al., 2005) and one of the necessary components of inferential analysis (Schum, 1987). As noted by Kent (1966), "what is desired in the way of hypotheses... is quality and quantity... a large number of possible interpretations of the data, [and] a large number of inferences, or concepts, which are broadly based and productive of still other concepts" (p. 174). Consistent with this perspective, rigorous analysis is identified by the depth and breadth of the generation and consideration of alternatives, by the incorporation of diverse perspectives in brainstorming hypotheses, by the evolution of thinking beyond an initial problem framing, and by the ongoing openness to the potential for revision.

Hypothesis exploration also represents a check against "premature narrowing" (Woods & Hollnagel, 2006; Cooper, 2005), given the concern "that if analysts focus mainly on trying to confirm one hypothesis they think is probably true... [they] fail to recognize that most of this evidence is also consistent with other explanations or conclusions" (Heuer, 1999, p. 96). Accordingly, indicators of weak hypothesis exploration include a focus on confirming a single explanation, a failure to consider alternative perspectives on available data, an interpretation of ambiguous or conflicting data such that it is compatible with existing beliefs, and analyst behavior indicative of fixation (Pherson, 2005; De Keyser & Woods, 1990).

2.   Information search. Similarly viewed as a fundamental component of analysis work, this attribute is alternatively described as the "collection of data bearing upon the [analysis] problem" (Kent, 1966, p. 157) or as the focused "extraction of an essential, representative, 'on analysis' sample from available data" (Elm et al., 2005, p. 297). Thus, information search encompasses all activities performed to gather task-relevant evidence— including those to broaden as well as deepen, those that are active as well as passive, and those hypothesis-driven as well as data-driven. Note that this framing of information search reflects the diverse nature of analytical activity and emphasizes the fact that, for the professional analyst, supporting evidence comes in many forms, and not simply as raw data (Krizan, 1999).

Information search is primarily concerned with where and how analysts look for supporting information. A strong information search process is characterized by the extensive exploration of relevant data, by the collection of data from multiple source types, and, most critically, by an active approach

6

to information collection. In this context, active collection describes efforts to seek information beyond that which is readily available (Serra, 2008). A weak information search in contrast is identified by failure to go beyond routine and readily available data sources, by reliance on a single source type or on "distant" data that is removed from original source material, and by passive dependence upon "pushed" rather than actively collected data.

3.   Information Validation. This attribute is concerned with the "critical evaluation of data" (Kent, 1966, p. 157) and with determining the level of conflict and corroboration—or agreement and disagreement—among sources (Elm et al., 2005). In rigorous analysis, analysts make an explicit effort to distinguish fact from judgment and are concerned with consistency and credibility among, as well as within, sources. Thus, a strong validation process involves assessing the reliability of sources, assessing the appropriateness of sources relative to the task question, and the use of proximate sources whenever possible (Krizan, 1999). It also involves an explicit effort to seek out multiple, independent sources of converging evidence for key findings.

In contrast, weak information validation is reflected in the uncritical acceptance of data at face value, little or no clear effort to establish underlying veracity, and a failure to collect independent supporting evidence. Poor tracking and citation of original sources also identify such analyses. Between strong and weak characterizations, a moderate validation process involves the recognition of inconsistencies among sources and, often times, involves the use of heuristics to support judgments of source integrity—such as deference to sources that have previously proven highly reliable and avoidance of those that have not. On the aggregate, then, information validation can be described as an intense concern with issues of agreement, consistency, and reliability with respect to the set of collected data.

4.   Stance analysis. "Stance" refers to the perspective of a source on a given issue and is alternatively discussed in the language of viewpoint, outlook, or affiliation—though more often it is characterized in terms of slant, bias, or predisposition. Stance analysis refers to the evaluation of information with the goal of identifying the positions of sources with respect to a broader contextual understanding and in relation to alternative perspectives on an issue. In the extreme, this analysis serves as a guard against deception (Johnson, Grazioli, Jamal, & Berryman, 2001; Moore & Reynolds, 2006; Heuer, 1987). More prosaically however, it serves as a check

against a reliance on unrepresentative data sets that reveal only partial perspectives on an issue.

A process in which little attention is paid to issues of stance reflects weak analysis. In such instances, the analysis may identify heavily slanted sources or sources that support a well-defined position on an issue but yet reflect little in the way of a nuanced understanding. A somewhat improved stance analysis would incorporate basic strategies for considering the perspectives of different sources. For example, dividing evidence into camps that are "for" or "against" an issue represents a simplifying heuristic for organizing and making sense of various stances on that issue.

A significantly stronger stance analysis involves research into, or leverages a preexisting knowledge of, the backgrounds and views of key individuals, groups, and thought leaders. Where appropriate, it may also include a more formal assessment that employs structured methods to identify critical relationships, to predict how the general worldview of a source is likely to influence his or her stance toward specific issues, or to detect the intentional manipulation of information. Social network analysis (Butts, 2008; Moon & Carley, 2007; Krebs, 2002), faction analysis (Feder, 1987; Smith, Branlat, Stephens, & Woods, 2008), and analysis of competing hypotheses (Heuer, 1999; Folker, 2000), for example, are techniques that may be selectively employed to those ends.

5.   Sensitivity analysis. The term "sensitivity", as it is used here, has a meaning most similar to its usage in the statistical analysis of quantitative variables, wherein it describes the extent to which changes in input parameters affect the output solution of a model (Saltelli, Chan, & Scott, 2000). However, rather than with the relationship between output and input variance, our concern is with the strength of an analytical assessment given the potential for low reliability and high uncertainty in supporting evidence and explanations. Phrased differently, sensitivity analysis describes the process of discovering the underlying assumptions, limitations, and scope of an analysis as a whole, rather than those of the supporting data in particular, as with the related attribute of information validation.

Many in the intelligence community emphasize the importance of examining analytical assumptions. Schum (1987), for example, identifies assumptions as foundational to analytic work. Similarly, Davis (2002) notes that "careful attention to selection and testing of key assumptions to deal with substantive uncertainty is now well established as the doctrinal standard for [analysis work]" (p. 5). As such, effective sensitivity analysis demands the

identification and explication of the key beliefs that tie an analysis together and that link evidence with hypotheses.

To that end, a strong sensitivity analysis goes beyond simple identification, meticulously considering the strength of explanations and assessments in the event that individual supporting evidence or hypotheses were to prove invalid. It also specifies the boundaries of applicability for the analysis. With weak sensitivity analysis, in contrast, explanations seem appropriate or valid at surface level, with little consideration of critical "what if" questions—e.g., "What if a key data source misidentified a person of interest?" Likewise, the overall scope of a weak analysis process may be unclear or undefined.

6.   Information synthesis. Often emphasized by experts more than casual analysts is that rigorous analytical work is as much about putting concepts together as it is about breaking an issue apart (Cooper, 2005; Mangio & Wilkinson, 2008). That is to say, rigorous analysis demands not only "analytic" activity in the definitional sense, but "synthetic" activity as well. Thus, information synthesis is a reflection of the extent to which an analysis goes beyond simply collecting and listing data to provide insights not directly available in individual source data.

Weak information synthesis is reflected in analyses that succeed in compiling relevant and "on topic" information, but that do little in the way of identifying changes from historical trends or providing guidance for broader or more long-term concerns. Indicators of weak synthesis include extensive use of lists, copying material from other sources with little reinterpretation, and a lack of selectivity in what is emphasized by the analysis. A stronger synthesis is reflected by explicit efforts to develop an analysis within a broader framework of understanding. The depiction of events in relation to historical or theoretical context and the framing of key issues in terms of tradeoff dimensions and interactions also identify such analysis.

Stronger still is synthesis that has integrated information in terms of relationships rather than components, with a thorough consideration of diverse interpretations of relevant data. In addition, such synthesis is performed by reflexive analysts who are attentive to ways in which their particular analytical processes may hinder effective synthesis and who are attuned to the many potential "cognitive biases" that manifest in analytical work (Moore, 2007; Krizan, 1999; Heuer, 1999).

7.   Specialist collaboration. Inevitably, analysts encounter topics on which they are not expert or that require multiple areas of expertise to fully

make sense of (Kent, 1966). Even in instances where an analyst has expertise in pertinent topics, success for the modern analyst still demands the incorporation of multiple perspectives on an issue (Clark, 2006; Medina, 2002). Accordingly, analytical rigor is enhanced when substantive expertise is brought to bear on an issue (Tenet, 1998).

The level of effort expended to incorporate relevant expertise defines effective specialist collaboration. In a process with little collaboration, minimal outside expertise is sought out directly. A moderately collaborative analysis process involves some interaction with experts, though at this level such expertise is often drawn from existing personal or professional networks, rather than from organizationally external sources. In a high-rigor process, independent experts in key content areas are identified and consulted. Thus, a strong specialist collaboration process is defined by efforts to go beyond a "core network" of contacts in seeking out domain-relevant expertise. In many cases, additional resources and "political capital" are expended to gain access to such specialized knowledge.

8.   Explanation critiquing. Specialist collaboration and explanation critiquing are related in that both are forms of collaborative analytical activity that reflect the influence of diverse perspectives (Hong & Page, 2004; Guerlain et al., 1999). However, whereas specialist collaboration primarily relates to the integration of perspectives relative to information search and validation, explanation critiquing relates to the integration of perspectives relative to hypothesis exploration and information synthesis. More succinctly, explanation critiquing is concerned with the evaluation of the overall analytical reasoning process, rather than with the evaluation of content specifically.

Similar to specialist collaboration however, this attribute is largely defined by the extent to which analysts reach beyond immediate contacts in collecting and integrating alternative critiques. A low quality explanation critiquing process has limited instances of such integration, while a more moderate process leverages personal and professional contacts to examine analytical reasoning. In the latter case, it is often peer analysts, supervisors, or managers who serve as the primary source of these alternative critiques.

In a still stronger analysis process, independent as well as familiar reviewers have examined the chain of analytical reasoning and explicitly identified which inferences are stronger and which are weaker. In addition, the use of formal methods such as "red teams" or "devil's advocacy" (Defense Science

Board Task Force, 2003; Pherson, 2005) as well as the expenditure of capital, political or otherwise, serve as hallmarks of strong explanation critiquing.

Examples of Calibrating the Attributes of Rigor to a Particular Task

The attribute-based framing of analytical rigor represents a model in the sense that it "constrains what can be measured by describing what is essential performance"; it represents an operational framework in that "the model parameters... become the basis for specifying the measurements" (Dekker & Hollnagel, 2004, p. 82). We now provide examples that illustrate how the attributes of rigor can be scored as low, moderate, and high for two very different analysis tasks.

The first example involves scoring the analytical rigor of twelve ad hoc, three-member teams that completed a logistics analysis and planning task set in a military context used in a prior study (Trent, Patterson, & Woods, 2007). The study task was for each team to develop a plan to move a predetermined number of troops and supplies to a desired location by the fastest, cheapest, and most secure routes possible, given a number of scenario-specific constraints. Half of the teams were collocated and worked together directly while the other half were distributed and limited to audio communication only.

For all cases, verbal transcripts of the study sessions were used to score the analysis and planning processes of the teams with respect to the attributes of rigor. Two independent coders rated each team as low, moderate, or high on each attribute as shown in Figure 2, with specialist collaboration not scored due to limitations imposed by the design of the study. Note that both coders were also provided with feedback as to how well the plan developed by each team met the stated scenario requirements.

Of the 84 items scored (12 teams x 7 scored attributes), 76 were judged by both raters to fall into the same low, moderate, or high category, implying strong agreement between raters ($\kappa_w = 0.86$). Disagreements in coding were most common with respect to the attribute of information search, but also occurred in scoring hypothesis exploration, information validation, and stance analysis. Overall however, there was general consistency in how the coders applied the framework to assess the analytical rigor of the processes employed by the teams.

Attribute	Team											
	1	2	3	4	5	6	7	8	9	10	11	12
Hypothesis Exploration	Low	Moderate	**High**	Moderate Low	Moderate	Moderate	Moderate Low	Moderate	**High**	Moderate	Moderate	Low
Information Search	Moderate Low	Moderate	**High**	**High**	Moderate	Moderate	Moderate	**High**	Moderate **High**	Moderate	Moderate **High**	Moderate
Information Validation	Low	Low	Moderate **High**	Moderate **High**	Moderate	Low	Low	Low	Low	Moderate	Low	Moderate
Stance Analysis	Low	Moderate	Moderate **High**	Low	Moderate	Moderate	Moderate	Moderate	Low	Moderate	Moderate	Low
Sensitivity Analysis	Low	Low	**Moderate**	Low	Low	Low	Low	Low	Low	Low	Low	Low
Information Synthesis	Low	Low	Moderate	Moderate	Low	Low	Low	**High**	Low	Low	Low	Low
Specialist Collaboration	—	—	—	—	—	—	—	—	—	—	—	—
Explanation Critiquing	Low	Moderate	**High**	Low	Low	Low	Moderate	Moderate	Moderate	Low	Moderate	Low
Solution Score (out of 60)	**44/60**	**53/60**	**60/60**	**31/60**	**33/60**	**19/60**	**24/60**	**53/60**	**60/60**	**60/60**	**53/60**	**39/60**

Figure 2. Scoring the rigor of the analysis processes of teams in the first example, with differences between raters highlighted.

The second example involves another study in a laboratory environment in which participants were asked to respond to a quick reaction task in the form of a recorded verbal briefing that answered the question: "What were the causes and impacts of the failure of the maiden flight of the Ariane 501 rocket launch?" The task was conducted under conditions of data overload, time pressure, and low to no expertise in the area, and the scenario was based on the case of the Ariane 5 rocket, whose maiden launch on June 4, 1996, ended in a complete loss of the rocket booster and scientific payload when it exploded shortly after liftoff, an accident which was significant because it departed from historical launch failure patterns in its underlying causes.

Data originally collected by this study for other purposes were later re-analyzed with respect to the attributes of rigorous analysis for two junior intelligence analysts who had not yet received their security clearances (Miller & Patterson, 2008) and for eight senior intelligence analysts, all from the same intelligence agency (Patterson, Roth, & Woods, 2001). In this case, a single individual coded the processes of each analyst (see Figure 3), however note that neither specialist collaboration nor explanation critiquing were scored, as the study design did not allow participants to access individuals with specialized expertise or to have their problem framing and briefing critiqued by other analysts. The results of this evaluation are represented graphically in Figure 4.

12

| Attribute | Novice | | Expert | | | | | | | |
	N1	N2	E1	E2	E3	E4	E5	E6	E7	E8
Hypothesis Exploration	**Moderate** No obvious initial bias.	**Moderate** No obvious initial bias.	**Moderate** Relied on high quality sources early on to formulate hypotheses.	Low Generally sloppy, rushed to get an acceptable answer.	**Moderate** Spent a lot of time working out the impacts, but less on the diagnosis of the failure.	**Moderate** Explicit strategy for sampling in time, bracketing the event, and playing forward to help break fixations and detect overturning updates.	**Moderate** Considered holistically if there is disagreement in what had been read. When there is disagreement, picked one to rely upon.	**Moderate** Framing evolves over time, even though participant has more subject-matter knowledge than other study participants.	**Moderate** Went quickly, but no obvious initial bias.	Low Seemed biased towards "mechanical failure" hypothesis and ended up there.
Information Search	**Moderate** Read 10 documents.	**Moderate** Read 9 documents (2 repeats).	**High** Time spent.	Low Time spent.	**High** Time spent.	**High** Time spent.	**Moderate** Time spent.	**High** Time spent.	**Moderate** Time spent, though it appeared more efficient than sloppy.	**Moderate** Time spent.
Information Validation	Low Wrong date for accident.	**Moderate** Corroborated some information.	**Moderate** Worried about cut and paste, past dates more reliable than predictions, did not trust exact cost estimates.	Low Would rely on customer to clarify inconsistencies in data.	**Moderate** Noted discrepancies informally and kept track of when they came up again.	**Moderate** Copied and pasted notes to bring information together on the screen.	Low Since no explicit statement of disagreement at a high level in documents, believed details were right (some were not).	**Moderate** Clearly noticed and worked on judging how to resolve discrepancies (but without looking for new documents).	Low Seemed to trust everything that was copied and pasted in, but holistically judged no obvious disagreements.	**Moderate** Highlighted phrases that appeared more than once.
Stance Analysis	Low Malaysian MBA students treated the same as other sources.	Low No judgment of source stances.	**Moderate** Sensitive when judging documents contributions.	**Moderate** Showed sensitivity to potential for bias in what was read, but did not obviously do anything when it was detected.	**Moderate** Impressively nuanced judgments of documents.	**Moderate** Sensitive to stances when reading.	**Moderate** Mostly looked for deception or disagreement and for reasons why, used time as a major cue.	**Moderate** Sensitive to stances when reading.	**Moderate** Sensitive to stances when reading.	**Moderate** Somewhat sensitive to stances when reading.
Sensitivity Analysis	Low No sensitivity to risk.	Low No sensitivity to risk.	**Moderate** Wanted to do timeline or ask customer to verify the right accident.	Low Little sensitivity to risk.	**Moderate** Particularly on impacts, circled the answers to see if similar from different sources.	**Moderate** Spent time and effort to see if new info was different.	Low No sensitivity to risk.	Low No sensitivity to risk.	Low No sensitivity to risk.	Low No sensitivity to risk.
Information Synthesis	Low Found document and forwarded as answer.	Low Personal interests guided what he read next, with no obvious strategy for pulling together information.	**Moderate** Provided insights that went beyond what was read, based on patterns and past experience.	Low Briefing at level of what was written and not put in context.	**High** Pulled it all together in the briefing in synthesized fashion, and also gave supporting details.	Low Gave detailed briefing at the level of what was read in documents; briefed a compiled list.	Low No time spent "getting a feel" or contextualizing what was found; briefed a compiled list.	**High** Clearly utilized knowledge to pull together ideas in coherent framework.	**High** Impressive coherent narrative showed a command in general of what he had read at the "gist" level.	**Moderate** Noticed payload not asked about, put things into his own words and framed by question, briefed more than a compiled list.
Specialist Collaboration	— —	— —	— —	— —	— —	— —	— —	— —	— —	— —
Explanation Critiquing	— —	— —	— —	— —	— —	— —	— —	— —	— —	— —

Figure 3. Scoring the rigor of the analysis processes of participants in the second example, with brief rationales for each rating.
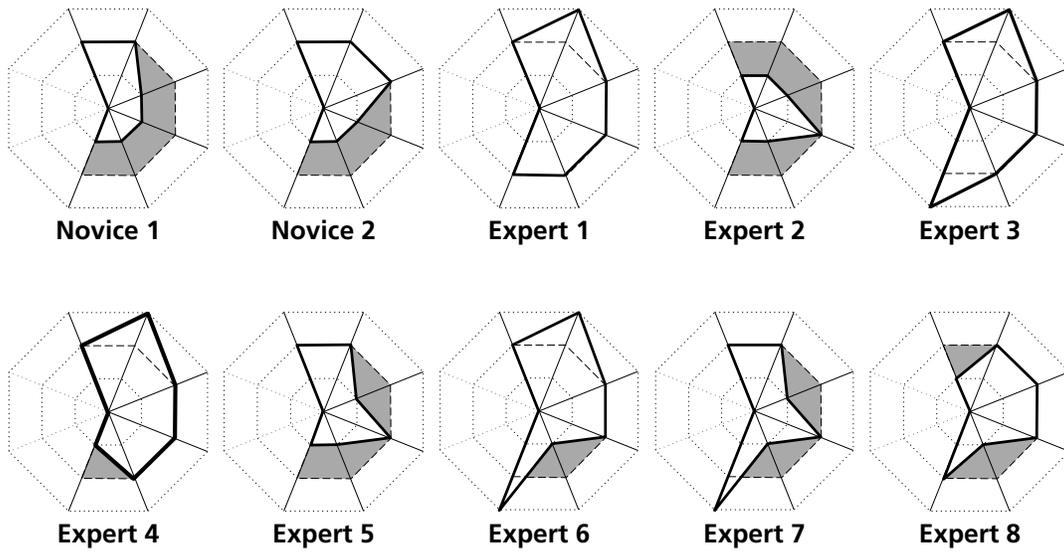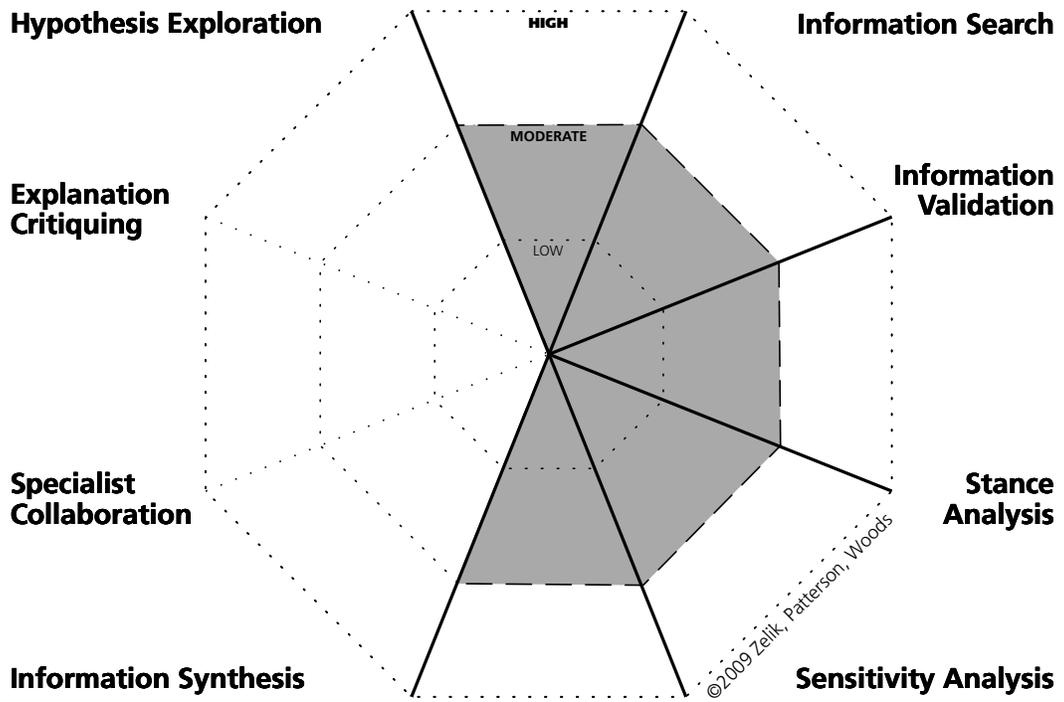
Hypothesis Exploration

Information Search

Explanation Critiquing

Information Validation

Specialist Collaboration

Stance Analysis

Information Synthesis

Sensitivity Analysis

HIGH

MODERATE

LOW

©2009 Zelik, Patterson, Woods

Novice 1 Novice 2 Expert 1 Expert 2 Expert 3

Expert 4 Expert 5 Expert 6 Expert 7 Expert 8

Figure 4. Graphical representation of the rigor attribute scores shown in Figure 3.

Note that this depiction of the attributes affords an integrated, pattern-based assessment of rigor that reveals an overall picture of the relationships among the attributes, as well as insight into the level of rigor achieved on individual attributes with respect to the demands of a given context (Bennett & Flach, 1992). In this example, interestingly, stance analysis—and to a lesser extent information synthesis and sensitivity analysis—were most diagnostic of expertise on the study task. In considering both examples presented here, it is our observation that there is flexibility in the rigor attribute framework that allows for its application in a variety of analytical contexts and to a variety of assessment tasks.

Discussion

We have presented a framework for and examples of applying an approach to codify the rigor of an information analysis process from a macrocognitive perspective that reflects the diversity of strategies employed by professional analysts to cope with data overload, to make sense of an uncertain and dynamic world, and to reduce the risk of shallow analysis. In practice, however, scoring a given analytic process with respect to these eight attributes is not without challenges, and there are many opportunities for advancing this line of work.

Of principal concern is the fact that the rigor attribute model is grounded largely in the domain of intelligence analysis. Accordingly, the extent to which it proves relevant in other analytical domains remains an open issue, though preliminary research on activities such as information search by students working on educational projects, medical diagnosis using automated algorithms for x-ray interpretation, and accident investigation analyses (Patterson, Zelik, McNee, & Woods, 2008) suggest that the concepts may be fairly robust. Of course, the framework presented here does not purport to be the final word on rigor in information analysis either; rather, it represents a current iteration of an ongoing direction of exploration.

In contrast with the conventional, standards-based notions discussed earlier, our framework supports a view of analytical rigor that is better aligned with calls for guidance—as opposed to guidelines—in conducting analyses of complex and highly interconnected data (Patton, 2002). This is not to suggest that standard methods play an insignificant role in achieving rigor, but that such standards do not necessarily and completely define rigor for the practicing analyst. Rather it is the extent to which these methods and practices speak to the individual attributes that define the overall rigor of an

analysis. Of course, in the intelligence community in particular, this is hardly a surprising conclusion (Mangino & Wilkinson, 2008; Marrin, 2007; Johnston, 2005).

A related implication of the rigor attribute model is that, although it elucidates the aspects of analytical process that are perceived as most diagnostic of rigor, it does not directly reveal what is required to be "rigorous enough". Again, this is not a surprising conclusion, though it is no less an important one.

In short, judging the sufficiency of rigor depends upon many contextual factors—analyst characteristics such as experience, specialization, and organizational affiliation; format and presentation characteristics such as writing style, presentation medium, and customer requirements; and production characteristics such as time pressure, resource availability, and priority, to name only a few (Krizan, 1999; Zelik, Woods, & Patterson, 2009). Moreover, just as there are situations where an analyst must rely on one or very few sources of information, so too are there situations in which exploring all available data is infeasible. Thus, contextual and situational demands largely dictate the best way in which to apply the rigor attribute model as a measure of process quality. Ultimately it is practitioners who negotiate and decide what is appropriate for a given set of circumstances (Gardiner, 1989; McLaughlin, 2008).

And so, in closing, we return to the risk of shallow analysis as a hallmark challenge of present-day information analysis. In the face of significant production pressures and rapidly proliferating data availability (Woods, Patterson, & Roth, 2002), it is increasingly easy for analysts and decision makers alike to fall into the trap of shallow analysis—believing an analysis to be sufficient in a context where it is not.

In responding to this challenge, the measurement of analytical rigor is advanced as a check against this risk. The model of analytical rigor presented here provides both direction for supporting this critical judgment task and a framework for a macrocognitive measure of performance that is based on how expert analysts assess rigor. To that end, the concept of analytical rigor in information analysis warrants continued exploration and diverse application as a macrocognitive measure of analytical sensemaking activity.

Acknowledgments

References

Bennett, K. B., & Flach, J. M. (1992). Graphical Displays: Implications for divided attention, focused attention, and problem solving. *Human Factors, 34,* 513–533.

Berkowitz, B., & Goodman, A. (1991). *Strategic intelligence for American national security*. Princeton, NJ: Princeton University Press.

Betts, R. K. (1978). Analysis, war, and decision: Why intelligence failures are inevitable. *World Politics, 31*, 61–89.

Betts, R. K. (2002). Fixing intelligence. *Foreign Affairs, 81*, 43–59.

Butts, C. T. (2008). Social network analysis: A methodological introduction. *Asian Journal of Social Psychology, 11*, 13–41.

Clark, R. M. (2007). *Intelligence analysis: A target-centric approach.* Washington, DC: CQ Press.

Cooper, J. R. (2005, December). *Curing analytic pathologies: Pathways to improved intelligence analysis* (Center for the Study of Intelligence). Washington, DC: U.S. Government Printing Office.

Crippen, D. L., Daniel, C. C., Donahue, A. K., Helms, S. J., Livingstone, S. M., O'Leary, R., & Wegner, W. (2005). Annex A.2: Observations by Dr. Dan L. Crippen, Dr. Charles C. Daniel, Dr. Amy K. Donahue, Col. Susan J. Helms, Ms. Susan Morrisey Livingstone, Dr. Rosemary O'Leary, and Mr. William Wegner. In Return to Flight Task Group (Eds.), *Return to Flight Task Group final report* (pp. 188–207). Washington, DC: U.S. Government Printing Office.

Davies, D., & Dodd, J. (2002). Qualitative research and the question of rigor. *Qualitative Health Research, 12*, 279–289.

Davis, J. (2002). Improving CIA analytic performance: Strategic warning. *The Sherman Kent Center for Intelligence Analysis Occasional Papers, 1*(1). Retrieved October 27, 2008, from https://www.cia.gov/library/kent-center-occasional-papers/vol1no1.htm

De Keyser, V., & Woods, D. D. (1990). Fixation errors: Failures to revise situation assessment in dynamic and risky systems. In A. G. Colombo & A. Saiz de Bustamante (Eds.), *Systems reliability assessment* (pp. 231–251). Dordrecht, The Netherlands: Kluwer Academic.

Defense Science Board Task Force. (2003). *Final report of the Defense Science Board Task Force on the role and status of DoD red teaming activities*. Washington, DC: U.S. Government Printing Office.

Dekker, S. (2005). *Ten questions about human error: A new view of human factors and system safety*. Mahwah, NJ: Lawrence Erlbaum.

Dekker, S., & Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology & Work, 6*, 79–86.

Elm, W., Potter, S., Tittle, J., Woods, D., Grossman, J., & Patterson, E. (2005). Finding decision support requirements for effective intelligence analysis tools. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (pp. 297–301). Santa Monica, CA: Human Factors and Ergonomics Society.

Feder, S. (1987). FACTIONS and Politicon: New ways to analyze politics. *Studies in Intelligence, 31*(1), 41–57.

Folker, R. D. (2000). *Intelligence analysis in theater joint intelligence centers: An experiment in applying structured methods* (Joint Military Intelligence College Occasional Paper Number Seven). Washington, DC: U.S. Government Printing Office.

Gardiner, L. K. (1989). Dealing with intelligence-policy disconnects. *Studies in Intelligence, 33*(2), 1–9.

Guerlain, S. A., Smith, P.J., Obradovich, J. H., Rudmann, S., Strohm, P., Smith, J. W., Svirbely, J., & Sachs, L. (1999). Interactive critiquing as a form of decision support: An empirical evaluation. *Human Factors, 41*, 72–89.

Grossman, J., Woods, D. D., & Patterson, E. S. (2007). Supporting the cognitive work of information analysis and synthesis: A study of the military intelligence domain. In *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting* (pp. 348–352). Santa Monica, CA: Human Factors and Ergonomics Society.

Heuer, R. J. (1987). Nosenko: Five paths to judgment. *Studies in Intelligence, 31*(3), 71–101.

Heuer, R. J. (1999). *Psychology of intelligence analysis*. Washington, DC: U.S. Government Printing Office.

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences, 101*, 16385–16389.

Johnson, P. E., Grazioli, S., Jamal, K., & Berryman, R. G. (2001). Detecting deception: Adversarial problem solving in a low base-rate world. *Cognitive Science, 25*, 355–392.

Johnston, R. (2005). *Analytic culture in the US Intelligence Community: An ethnographic study*. Washington, DC: U.S. Government Printing Office.

Kent, S. (1966). *Strategic intelligence for American world policy*. Princeton, NJ: Princeton University Press.

Klein G. (in press). Macrocognitive measures for evaluating cognitive work. In E. S. Patterson & J. Miller (Eds.), *Macrocognition metrics and scenarios: Design and evaluation for real-world teams*. Aldershot, UK: Ashgate Publishing.

Klein, G., Moon, B., & Hoffman, R. R. (2006a). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems, 21*(4), 70–73.

Klein, G., Moon, B., & Hoffman, R. R. (2006b). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems, 21*(5), 88–92.

Klein, G., Phillips, J. K., Rall, E., & Peluso, D. A. (2007). A data/frame theory of sensemaking. In R. R. Hoffman (Ed.), *Expertise out of context: Proceedings of the 6th International Conference on Naturalistic Decision Making*. Mahwah, NJ: Erlbaum.

Klein, G., Ross, K. G., Moon, B., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent Systems, 18*(3), 81–85.

Krebs, V. (2002, April 1). Uncloaking terrorist networks. *First Monday, 7*(4). Retrieved September 30, 2009, from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/941/863

Krizan, L. (1999). *Intelligence essentials for everyone* (Joint Military Intelligence College Occasional Paper Number Six). Washington, DC: U.S. Government Printing Office.

Mangio, C. A., & Wilkinson, B. J. (2008, March). *Intelligence analysis: Once again*. Paper presented at the International Studies Association 2008 Annual Convention, San Francisco, California.

Marrin, S. (2007). Intelligence analysis: Structured methods or intuition? *American Intelligence Journal, 25*(1), 7–16.

McLaughlin, J. (2008). Serving the national policymaker. In R. Z. George & J. B. Bruce (Eds.), *Analyzing intelligence: Origins, obstacles and innovations* (pp. 71–81). Washington, DC: Georgetown University Press.

Medina, C. A. (2002). The coming revolution in intelligence analysis: What to do when traditional models fail. *Studies in Intelligence, 46*(3). Retrieved February 10, 2007, from https://www.cia.gov/csi/studies/vol46no3/article03.html

Military Operations Research Society. (2006, October 3). *MORS special meeting terms of reference*. Paper presented at Bringing Analytical Rigor to Joint Warfighting Experimentation, Norfolk, VA. Retrieved December 11, 2008, from http://www.mors.org/meetings/bar/tor.htm

Miller, J. E., & Patterson, E. S. (2008). Playback technique using a temporally sequenced cognitive artifact for knowledge elicitation. In *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (pp. 523–527). Santa Monica, CA: Human Factors and Ergonomics Society.

Miller, J. E., Patterson, E. S., & Woods, D. D. (2006). Elicitation by critiquing as a cognitive task analysis methodology. *Cognition, Technology & Work*, 8, 90–102.

Moon, I., & Carley, K. (2007). Modeling and simulating terrorist networks in social and geospatial dimensions. *IEEE Intelligent Systems, 22*(5), 40–49.

Moore, D. T. (2007). *Critical thinking and intelligence analysis* (Joint Military Intelligence College Occasional Paper Number Fourteen). Washington, DC: U.S. Government Printing Office.

Moore, D. T., & Reynolds, W. N. (2006). So many ways to lie: The complexity of denial and deception. *Defense Intelligence Journal, 15*(2), 95–116.

Morse, J. M. (2004). Preparing and evaluating qualitative research proposals. In C. Seale, G. Gobo, J. F. Giubrium, & D. Silverman (Eds.), *Qualitative research practice* (pp. 493–503). London: SAGE.

Patterson, E. S., & Miller, J. (Eds.). (in press). *Macrocognition metrics and scenarios: Design and evaluation for real-world teams*. Aldershot, UK: Ashgate Publishing.

Patterson, E. S., Roth, E. M., & Woods, D. D. (2001). Predicting vulnerabilities in computer-supported inferential analysis under data overload. *Cognition, Technology & Work, 3*, 224–237.

Patterson, E. S., Zelik, D., McNee, S., & Woods, D. D. (2008). Insights from applying rigor metric to healthcare incident investigations. In *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (pp. 1766–1770). Santa Monica, CA: Human Factors and Ergonomics Society.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Pherson, R. (2005, March). *Overcoming analytic mindsets: Five simple techniques*. Paper presented at the annual meeting on Emerging Issues in National and International Security, Washington, DC.

Saltelli, A., Chan, K., & Scott, E. M. (Eds.) (2000). *Sensitivity analysis*. New York: John Wiley & Sons.

Sandelowski, M. (1986). The problem of rigor in qualitative research. *Advances in Nursing Science, 8*(3), 27–37.

Sandelowski, M. (1993). Rigor or rigor mortis: The problem of rigor in qualitative research revisited. *Advances in Nursing Science, 16*(2), 1–8.

Schraagen, J. M., Klein, G., & Hoffman, R. R. (2008). The macrocognition framework of naturalistic decision making. In J. M. Schraagen, L. G. Militello,

T. Ormerod, & R. Lipshitz (Eds.), *Naturalistic decision making and macrocognition* (pp. 3–26). Aldershot, England: Ashgate.

Schum, D. A. (1987). *Evidence and inference for the intelligence analyst* (Vol. 1). Lanham, MD: University Press of America.

Serra, J. (2008). Proactive intelligence. *Futures, 40*, 664–673.

Smith, M. W., Branlat, M., Stephens, R. J., & Woods, D. D. (2008, April). *Collaboration support via analysis of factions*. Papers presented at the NATO Research and Technology Organisation Human Factors and Medicine Panel Symposium, Copenhagen, Denmark.

Tenet, G. (1998). *Unclassified recommendations of the Jeremiah report*. Retrieved December 26, 2008, from http://ftp.fas.org/irp/cia/product/ jeremiah-decl.pdf

Trent, S. A., Patterson, E. S., & Woods, D. D. (2007). Challenges for cognition in intelligence analysis. *Journal of Cognitive Engineering and Decision Making, 1*, 75–97.

Treverton, G. F., & Gabbard, C. B., (2008). *Assessing the tradecraft of intelligence analysis* (RAND National Security Research Division). Retrieved October 21, 2008 from http://www.rand.org/pubs/technical_reports/2008/ RAND_TR293.pdf

Woods, D. D., & Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering*. New York: Taylor & Francis.

Woods, D. D., Patterson, E. S., & Roth, E. M. (2002). Can we ever escape from data overload? A cognitive systems diagnosis. *Cognition, Technology & Work, 4*, 22–36.

Zelik, D. J., Patterson, E. S., & Woods, D. D. (2007). Judging sufficiency: How professional intelligence analysts assess analytical rigor. In *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting* (pp. 318–322). Santa Monica, CA: Human Factors and Ergonomics Society.

Zelik, D., Woods, D. D., & Patterson, E. S. (2009). *The supervisor's dilemma: Judging when analysis is sufficiently rigorous*. Paper presented at the CHI 2009 Sensemaking Workshop, Boston, MA.