

Information Fusion Process Design Issues for Hard and Soft Information: Developing an Initial Prototype

James Llinas

1 Introduction

The Data and Information Fusion (DIF) process can be argued to have three main functions: Common Referencing (CR) (also known as “Alignment”), Data Association (DA), and State Estimation, as shown in Fig. 1:

It can be argued that any DIF process can be architected as a network of such nodes (see [1]). In Fig. 1, we have either data or estimates (from a prior Fusion Node) entering this process. Data Alignment or CR is a function that transforms all input data to a common format and also, importantly, a common semantic framework. DA is a function that associates the evidence from multiple sources to asserted entities in the domain of interest; such entities can be not only physical objects but events, behaviors, situational substructures, etc. DA involves accounting for sensor and estimation errors, and also for semantic differences and similarities; the idea is to assemble and partition evidential sets of information so that subsequent inferencing and estimation processes are applied to the most robust collections of evidence about any such entity. DA comprises the three operations shown of Hypothesis Generation (defining feasible associations), Hypothesis Evaluation (a strategy for scoring inter-entity associability of the collective evidence), and Hypothesis Selection (typically some type of optimization scheme to define the best associations of all that are feasible and of higher scores). At this point, there is thus a set of entity-evidence groupings (evidence “assigned” to a given entity), and the assigned evidence is passed to whatever estimation process is at work on the given entity for next-time-increment processing. State Estimation (SE) follows, with various inferencing or estimation methods operating on these aligned and associated evidential sets. By a far greater proportion, the DIF research and development community has applied such techniques to evidential data coming from modern electromechanical sensors, i.e., what some call “physics-based” sensors. In the military/defense domains, these are the modern sensors used for Intelligence, Surveillance, and Reconnaissance (ISR) applications, ranging from satellite-based sensor systems to embedded, Unattended Ground Sensors (UGS’s). In

J. Llinas (✉)
Center for Multisource Information Fusion,
University at Buffalo, Buffalo, NY, USA
e-mail: llinas@buffalo.edu

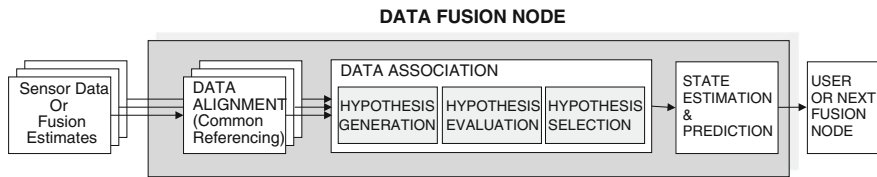


Fig. 1 Nominal fusion node processes

these cases, the designer of automated CR-DA-SE DIF operations enjoys the benefits of dealing with input sources that are well-calibrated and understood, and largely if not exclusively (in raw form at least) numerical. Data from such sources has come to be called “hard” data in the sense of these well-understood properties. For the type of operations involving traditional military engagements, where ISR must be conducted covertly and at a distance, the use of such ISR sensors and hard data has worked reasonably well.

However, subsequent to the end of the Cold War, the nature of defense and military operations has changed dramatically, from so-called “conventional” operations to what today are called “irregular” and “asymmetric” operations. These environments are characterized by a number of complicating features:

- They can be quite complex, involving terrorist, criminal, insurgent, and warfighting mixed operations
- They typically have no clearly defined or identifiable adversaries
- Hostiles/adversaries are mixed in with neutral, friendly persons or forces
- The goals involve not only destructive goals (“kinetic” actions) but establishment of influence and indirect effects

among other factors; such definitions are controversial and it is not our intent to be precise here but to give a flavor for these distinctions. Some of the other subtleties of these environments are that there are now improved data sources to better understand an “enemy” (a first principle of warfare), not only from a military point of view but from a socio-cultural point of view. Further, as noted above, friendly forces are often embedded in certain of these environments, which permits direct and close observation by these forces (humans, not sensors). Experiences in Iraq and Afghanistan and other places in the world in dealing with intelligence and security problems are typical of these new problems, and have required the (ongoing) formulation of new paradigms of intelligence analysis and dynamic decision-making. Broadly, these problems fall into the categories of counter-terrorism and counter-insurgency (COIN) as well as stability operations. Depending on the phases of counter-insurgency or other operations, the nature of decision-making ranges from conventional military-like to socio-political (sometimes also characterized as “hard” and “soft” decisions). Because of this wide spectrum of action, the nature of information support required for analysis has an equally wide range. Since automated DIF processes provide some of the support to such decision-making, requirements for DIF process design must address these varying requirements, resulting in considerable challenges in DIF

process design. One important driving factor for DIF process design is the new heterogeneity of the information supportive of DIF process design; these factors are discussed in the Sect. 2.

2 Heterogeneity of Supporting Information

2.1 *Observational Data*

As remarked above, the experiences in Iraq and Afghanistan, and in other similar involvements have also shown that some of the key observational and intelligence data in such operations comes not only from traditional sensor systems but from dismounted soldiers or other human observers reporting on their patrol activities. These data are naturally communicated in language in the form of various military and intelligence reports and messages. Such data, in textual, linguistic form, are entirely different than hard sensor data, as they are much more ambiguous, yet they can also be much more semantically rich; they are “soft” data in the sense that they are both largely uncalibrated and their content is much harder to fully understand (deep understanding begs the age-old challenge of forming automated methods for natural language understanding). Such “Soft” data finds its way into DIF processes as both structured and unstructured digitized text, and this input modality creates new challenges to DIF process designs, contrasted with more traditional DIF applications involving the use of highly-calibrated, numerically precise observational data from sensors. Combined with the data from the usual repertoire of “hard” or sensor data from various radio frequency (RF) sensors, video and other imaging systems, as well as SIGINT and satellite imagery, the observational data stream is a composite of data of highly different quality, sampling rates, content, and structure.

One main deficiency and critical path is on the soft data/human observation side, since it is generally agreed that DIF for observational data provided by hard, physical-science type sensors is much more mature; some have in fact argued that capabilities for Level 1 Fusion with hard data input is rather mature and that limited research investments should be made in this area. Although additional hard/Level 1 fusion research remains to be done, we generally concur with these judgments and believe that the first requirement is to define and prototype a viable processing paradigm for soft data fusion both for single and multiple input streams, so that the critical pre-estimation functions of Common Referencing and DA can be constructed. If we examine a notional processing diagram for multiple streams of human observational data expressed in linguistic terms (this is just one category of soft data), we envision something like the process in Fig. 2 (this is similar to a prototype of this process we have developed at our research center [2]):

In this depiction, each human observer processes the energy received from their sensing capability into a Perception-Cognition cycle, and a mental process judges how to express the observation in language (Linguistic Framing), resulting in a linguistic utterance, and the chosen instance of language. This utterance may be audio and need to be converted to digital text, and then formed into a message (that

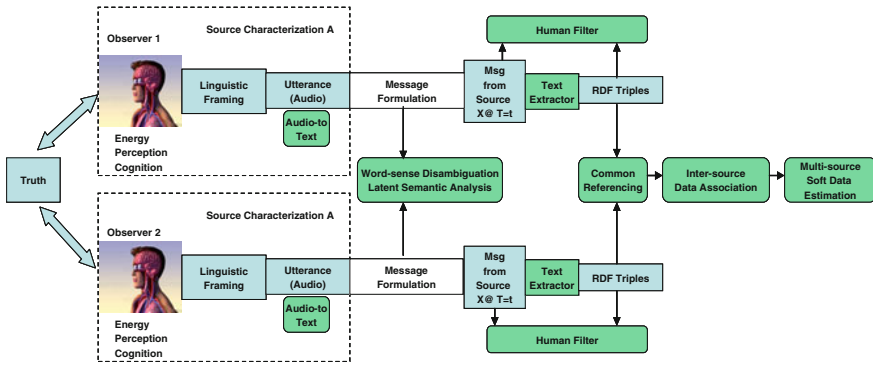


Fig. 2 Notional multi-message-stream soft data fusion process

may be sent over network communications channels, not shown). Today, the received message is typically parsed by a state-of-the-art Text Extractor, yielding for example RDF Triples of Subject-Verb-Object phrases, or some other representation (see [3] for the “propositional graph” approach we are using in our research). In virtually every military application, the message stream and/or the triples would be filtered through a human observer who functions as a first-level Quality Control process. Each filtered triples stream then comprises the raw data input into a downstream Data Fusion process. The meta-data (time-tags, uncertainty, etc) and the semantic content of these triples need to be framed in a normalized way by processing through the CR function, and then associated to determine if they relate to the same Entity in the true, unknown world, so that multisource, fusion-based estimation processes can exploit their informational content.

2.2 Open Source and Social Media Data

Soft or Hard data can also find its way into modern DIF processes in the form of monitored Open Source and Social Media feeds such as newswire feeds, Twitter and Blog sources judged to be possibly helpful. Getting such data into a DIF system will require automated web crawlers and related capabilities, and subsequent natural language processing capabilities, as much of these data are also represented in language.

2.3 Contextual Data

Modern problems also afford (and demand) the use of additional data and information beyond just observational data. A major category of such data and information is

Contextual Information. Contextual Information is that information that can be said to “surround” a situation of interest in the world (many definitions and characterizations exist but we will not address such issues here). It is information that aids in understanding the (estimated) situation and also aids in reacting to the situation, if a reaction is required. Contextual Information can be relatively or fully static or can be dynamic, possibly changing along the same timeline as the situation (e.g., weather). It is also likely that the full characterization and specification of Contextual Information may not be able to be known at system/algorithm design time, except in very closed worlds. Thus, we envision an “a priori” framework of exploitation of Contextual Information that attempts to account for the effects on situational estimation of that Contextual Information (CI henceforth) that is known at design time. Even if such effects are known at design time, there is a question of the ease or difficulty involved in integrating CI effects into a fusion system design or into any algorithm designs. This issue is influenced in part by the nature of the CI and the manner of its native representation, e.g., as numeric or symbolic, and the nature of the corresponding algorithm; for example, cases can arise that involve integrating symbolic CI into a numeric algorithm. Strategies for a priori exploitation of CI may thus require the invention of new hybrid methods that incorporate whatever information an algorithm normally employs in estimation (usually observational data) with an adjunct CI exploitation process. Note too that CI may, like observational data, have errors and inconsistencies itself, and accommodation of such errors is a consideration for hybrid algorithm design. Similarly, we envision the need for an “a posteriori” CI exploitation process, due to at least two factors: (1) that all relevant CI may not be able to be known at system/algorithm design time, and may have to be searched for and discovered at runtime, as a function of the current situation estimate, and (2) that such CI may not be of a type that was integrated into the system/algorithm designs at design time and so may not be able to be easily integrated into the situation estimation process. In this case then we envision that at least part of the job of posteriori CI exploitation would involve checking the consistency of a current situational hypothesis with the newly-discovered (and situationally-relevant) CI. There are yet other system engineering issues. The first is the question of accessibility; CI must be accessible in order to use it, but accessibility may not be a straightforward matter in all cases. One question is whether the most-current CI is available; another may be that some CI is controlled or secure and may have limited availability. The other question is one of representational form. CI data can be expected to be of a type that has been created by “native” users; for example, weather data, important in many fusion applications as CI, is generated by meteorologists, for meteorologists (not for fusion system designers). Thus, even if these data are available, there is likely to be a need for a “middleware” layer that incorporates some logic and algorithms to both sample these data and shape them into a form suitable for use in fusion processes. In even simpler cases, this middleware may be required to reformat the data from some native form to a useable form. In spite of some a priori mapping of how CI influences or constrains the way in which situational inferences or estimates can be developed, which may serve certain environments, the defense and security type applications, with their various dynamic and uncertain types of CI, demand a

more adaptive approach. Given a nominated situational hypothesis H_f from a fusion process or “engine”, the first question is: what CI type information is relevant to this hypothesis? Relevant CI is only that information that influences our interpretation or understanding of H_f . Presuming a “relevancy filter” can be crafted, a search function would explore the available CI and make this CI available to an “posteriori” reasoning engine. That reasoning engine would then use: (1) a CI-guided subset of Domain Knowledge, and (2) the retrieved CI to reason over H_f to first determine consistency of H_f with the relevant CI. If it is inconsistent, then some type of adjudication logic will need to be applied to reconcile this inconsistency between: (1) the fusion process that produced H_f and (2) the posteriori reasoning process that judges it as inconsistent. If however H_f is judged as consistent with the additional CI, an expanded interpretation of H_f could be developed, providing a deeper situational understanding. This overall process, which can be considered a “Process Refinement” operation, would be a so-called “Level 4” process in the context of the JDL Data Fusion Process Model (see [1]), that is, as an adaptive operation for fusion process enhancement. The overall ideas discussed here are elaborated in [4].

2.4 *Ontological Data*

DIF processes and algorithms have historically been developed in a framework that has assumed the a priori availability of a reliable body of procedural and dynamic knowledge about the problem domain; that is, knowledge that supports a more direct approach to temporal reasoning about the unfolding patterns of interest in the problem domain. In COIN and other complex problems, such a priori and reliable knowledge is most often not available—the Tactics, Techniques and Procedures (“TTP’s”) of modern-day adversaries are highly adaptive and extremely hard to model with confidence. The US DARPA COMPOEX Program [5] attempted to develop such models but only achieved partial success, experiencing gaps in the overall modeling space of such desired behavioral models. We label these types of problems as “weak knowledge” problems, implying that only fragmentary a priori behavioral model type knowledge is available to aid in DIF based reasoning, inferencing, and estimation.

Ontological information however, that does not attempt to overtly form such comprehensive behavioral and temporal models but does include temporal primitives along with structural/syntactic relations among entities, can be specified a priori with reasonably good confidence, and thus provides a declarative knowledge base to support DIF reasoning and estimation. Note that such knowledge is also represented in language and is available as digital text, in the same way as data from messages, documents, Twitter, etc. The use of ontological information in DIF systems can be varied; ontological information can augment observed data, can aid in asserting possible relationships, help in directing search and also in sensor management (to acquire expected information based on ontological relations), and yet other ways. Importantly, specified ontologies can also serve as providing consistent and grounded semantic terminology for any given system. In our current research,

we employ ontologies primarily for augmenting observational data with asserted ontological data whose relevance is algorithmically determined using “spreading activation” and then integrated to enrich the evidential basis for reasoning [6]. The broader implications of ontologies for intelligence analysis are described in [7], that comes from our university’s National Center for Ontological Research (see http://ncorwiki.buffalo.edu/index.php/Main_Page).

2.5 Learned Information

Finally, there is the class of information that could be learned (online) from all of the above sources if the DIF process is designed with a Data Mining/Inductive or Abductive Learning functional component. Very little research and prototyping of such dual-inferencing-process type DIF systems has been done although the conceptualization of such DIF schemes and architectures has been put forward some time ago by Waltz (e.g., [8]), as shown in Fig. 2. Any DIF system that incorporates such dual-inferencing schemes will encounter the challenge of knowledge management; whether and how any runtime learned knowledge gets integrated into runtime operations, or gets saved for later operations, or any other scheme for employment of learned knowledge is a challenge for storing, managing, and integrating that knowledge. The runtime integration of learned information raises a number of both algorithmic issues as well as architectural issues. For example, if meaningful patterns of behavior can be learned and can be measured/judged as persistent or enduring, such patterns could be incorporated in a dynamically-modifiable knowledge base to be reused. In Fig. 3, Waltz shows that the management of such knowledge evolving from what he calls Data Mining operations is handled by the “Level 4”, process refinement function of the traditional JDL DIF process.

Such learning processes will also not be perfect and have some uncertainty that also needs to be factored into the traditional CR and DA functions of the target fusion process.

The heterogeneity of data and information as just described also creates new challenges and complexities for the traditional functions of DIF as depicted in Fig. 1. In the next section, we address the impacts of these modern defense/security problems and of data heterogeneity on the DIF functions of Data Alignment or Common Referencing, and on DA.

3 Common Referencing and Data Association

As pointed out in Fig. 1, CR is that traditional DIF system function that is sometimes called “Alignment” and is the function that normalizes these input sources for any given fusion application or design. CR addresses such things as coordinate system normalization, temporal alignment issues, and uncertainty alignment issues

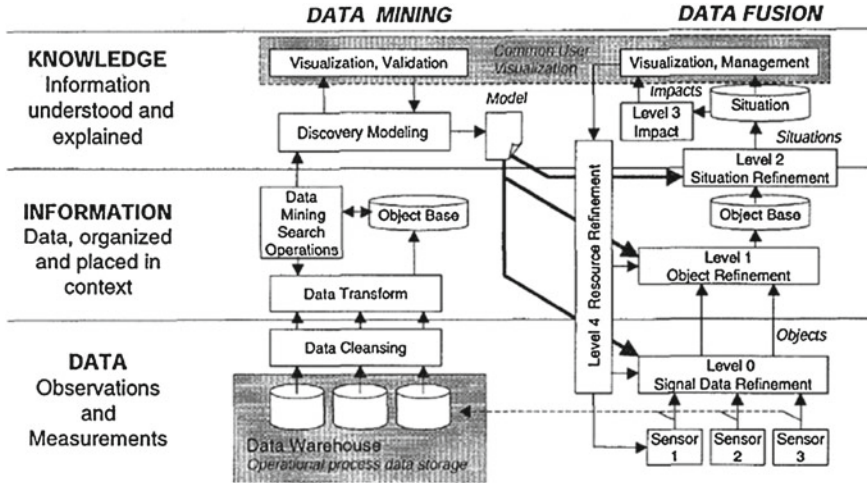


Fig. 3 Notional fusion process architecture combining data mining and data fusion (from [8])

across the input streams, among other issues. With the highly-disparate input streams described above, the design of required CR techniques is a non-trivial challenge. There are at least two major CR issues that this heterogeneous data represent: temporal alignment and uncertainty alignment. Consider a textual input message whose free text, in just a few lines, could have past-present-future tense expressions, e.g., “3 days ago I saw...”, “past precedents lead me to believe that tomorrow I should see...”, etc. Other sources can also have varied temporal structures regarding their input. Such data lead to the issue of what the DIF community has called “OOSM: out-of- sequence-measurements” for hard/sensor data but the issue carries over to all sources as well. Dealing with these issues requires complex temporal alignment techniques for CR and also raises the issue of retrospective fusion processing operations to correct for delayed inputs (if warranted; this is a design choice). For example, such process designs impute the need to set a threshold for allowable delays (how far back in time will we adjust for), and this also sets a requirement for memory capacity to save all data in that window to allow undoing and redoing the inferences when such time-late or past-referenced data arrives. Temporal alignment methods we have used for Soft data are described in [9].

The uncertainty alignment requirement evolves due to the high likelihood that any uncertainty in the widely disparate sources described above will be represented in inconsistent forms. Consider the basic differences between the uncertainty in sensor (hard) data and textual (soft) data; sensor data uncertainty is sensibly always expressed in probabilistic form whereas, due to the problem of imprecise adjectives and adverbs in language, linguistic uncertainty is often expressed in possibilistic (fuzzy) terms. It can be expected that uncontrolled Open Source or Social Media data may use yet other uncertainty formalisms to express or tag inputs (e.g., beliefs and subjective confidence measures). Transformation and normalization of disparate

forms of uncertainty is a specialized topic in the uncertainty/statistical literature (e.g., [10]), and is among the high-priority issues in the DIF community [11]. It should be noted that such transformations can only be developed by invoking some statistical type qualities that are preserved across the transform, such as some form of total uncertainty; that is, the transform of some probability value does not create an “equivalent” value of a probability in, say, a possibilistic space; instead the transformed value is one that satisfies some statistical constraint about which the transform is structured. For the interested reader, seminal papers on the probability-possibility transformation issue are in [12–14]. In our research, we have addressed the probabilistic- possibilistic transformation issue in an approach that satisfies the consistency and preference preservation principles [15], resulting in the most specific distribution for a specified portion of a probabilistic representation; this yields a truncated triangular transformation in our case [16].

Regarding the DA function, that some consider the heart of a fusion process, these highly-varied data raise the level of DA complexity in significant ways. The soft data category, that inherently is reporting about Entities and (judged) Relationships, and is inherently in semantic format (language/words), raises the important issue of how to measure semantic similarity of such elements as reported in these various input streams. Such scores are needed in the “Hypothesis Evaluation” step of the DA process (see [17] on these DA subfunctions). But there are further DA complications that arise due to the soft data: linguistic phrases have verbs that reflect inter-Entity (noun) relationships; also of note is that the Natural Language Processing (NLP) community has employed graphical methods for the representation of linguistic structures. As a result, the DA process now involves inter-association of both Entities (nouns) and Relations (verbs), and of graphical structures. This requirement extends to the hard data as well since that data needs to be cast in a semantic framework in order to enable the overall DA process for the combined Hard and Soft data. Developing DA methods for graphical structures represents an entirely new challenge for the DA function. In such approaches for these applications, a scoring approach also needs to be developed to assess Relational similarity as well as Entity similarity, and a composite association scheme for these graphical substructures needs to be evolved. Historical approaches to DA have often employed solution methods drawn from assignment problems in Operations Research. When association is required between many non-graphical data sources (i.e., among entities and attributes, as in the multisensor-multitarget tracking DA problem), this can be handled by such methods as the multidimensional assignment problem [18, 19]. The main difference between the multidimensional assignment problem and graph-based association is how topological information from the graphs is used. Our research center has attacked this problem and has developed research prototype algorithms, as described in [20] where the graph association problem is formulated as a binary linear program and a heuristic for solving the multiple graph association is developed using a Lagrangian relaxation approach to address issues involving a between-graph transitivity requirement.

In virtually all computer applications involving the estimation or inferencing about some state of affairs such as a “situation”, there is the issue of constructing computer-

based processes (software) that is able to work with notions of “meaning”. Dealing with notions of meaning becomes more difficult in DIF processes as one attempts to build methods for so-called “high-level” fusion, involving more abstract hypotheses such as situations and threats, etc. In modern problems and with hard and soft data sources, these problems are aggravated; some aspects of these issues are discussed in the next section.

4 Semantics

The introduction of linguistic information, as well as the transformation of sensor+algorithm estimation process outputs (hard fusion outputs) into a semantic frame, also adds to the complexity of DIF process design and development. Semantic complexity is also added by the very nature of modern intelligence and security problems wherein the situations of interest relate to both military operations and also socio-political behaviors and entities. Clear meanings of such notions of interest in modern intelligence or ISR problems such as “patterns of life”, “rhythm of the city”, “radicalization” as patterns or situations of interest—to be estimated by DIF systems—have proven difficult to specify in clear semantic terms, that is, to specify their meaning with adequate specificity for computer-based processes. While the use of ontologies helps in this regard, standardization issues remain when considering networked and distributed systems, which are typical in the modern era. For example, in distributed intelligence or military systems there is typically no single point of architectural authority that can mandate a single ontological framework for the network. For large-scale real systems there is also the problem of large legacy systems that were never designed with ontological formalisms in mind; this creates a “retrofit” problem of adjusting the semantic framework of that system to some new ontological standard, which can be a costly and complex operation.

It must also be noted that the way in which all textual/linguistic information gets into a DIF system is through processing in some type of NLP or text extraction system. Such systems serve as a front-end filter for the admission of fundamental entity and relationship data, the raw Soft data of the system, and so any imperfections in such extractions bound the capture of semantically-grounded evidential information for the subsequent reasoning and estimation processes; that is, the meaning of the text can be lost. While errors in hard sensor data are typically known with reasonable accuracy due to sensor calibrations, the errors in text extraction and NLP systems are either weakly known or unknown, sometimes as a result of proprietary constraints. Other strategies to deal with the complexities of semantics involve the use of controlled languages, to bound the grammatical structures and also the extent of the vocabulary that has to be dealt with. A good example for military/intelligence applications is the “Battle Management Language” or BML [21] that has been under development since about 2003 for both Command and Control simulation studies but also for DIF applications (e.g., [22, 23]).

There is a corresponding need to better understand the nature of semantic (and syntactic) complexity in language, and also to develop measures and metrics that aid in developing better NLP processes and controlled languages. There is a reasonably rich literature on these topics (e.g., [24]) that should be exploited in regard to the integrated design of DIF systems that today have to deal with a wide range of semantic difficulties.

As hinted at in our discussion regarding DA, many of these current problems involve graphical data representations and therefore impose the use of graphically-based algorithmic techniques. Some of these issues are addressed in the next section.

5 Graphical Representations and Methods

There are a number of reasons that, for COIN and asymmetric warfare-type problems, graphs are becoming a dominant representational form for the information in and the processes involved in DIF systems. In the information domain, many of the components discussed in Sect. 2 are textual/linguistic and to capture this information in digital form, graphs are the representational form of choice. The problem domain is also described in the ontologies that are also typically couched in graphical forms. Note that ontologies describe inter-entity relations of various types. Note too that the inferences and estimates of interest in these problems are of the “higher-level” type in the sense of the JDL Model of Information Fusion, that is, estimates of situations and threat states. These higher-level states—the conditions of interest for intelligence and security applications—are also best described as graphs, since situations can in the most abstract sense be considered as a graph of entities and relations.

As a result, it is not unexpected to see that the core functions of DIF such as DA as previously described, are employing graphical methods in these fusion function operations. The U.S. Army’s primary intelligence support system, the Distributed Common Ground Station-Army (DCGS-A) employs a “global graph” approach to capture all of the evidentiary information that supports DIF and other intelligence analysis operations; see [25] and Fig. 4 that shows the top-level structure of this graphical concept.

Developing a comprehensive understanding of these problems thus involves a logical synthesis of the many situational substructures or subgraphs in these problem domains; the fusion-process-generated subgraphs can be thought of situational components or hypotheses. The subgraphs are somewhat thematic and can be thought of as revolving about the “PMESII” notion of the heterogeneity of the classes of information of interest in such problems (PMESII stands for Political, Military, Economic, Social, Infrastructure, and Information categories). Thus, it is also not surprising to see Social Network Analysis tools—that are by the way graph-theoretic and graph-centric—employed in support of intelligence analysis, here with the focus on the Social and Infrastructure patterns and subgraphs of the problem space.

In our own work for such problems, we considered that it would be broadly helpful in analysis to enable a subgraph-querying capability as a generalized analysis tool.

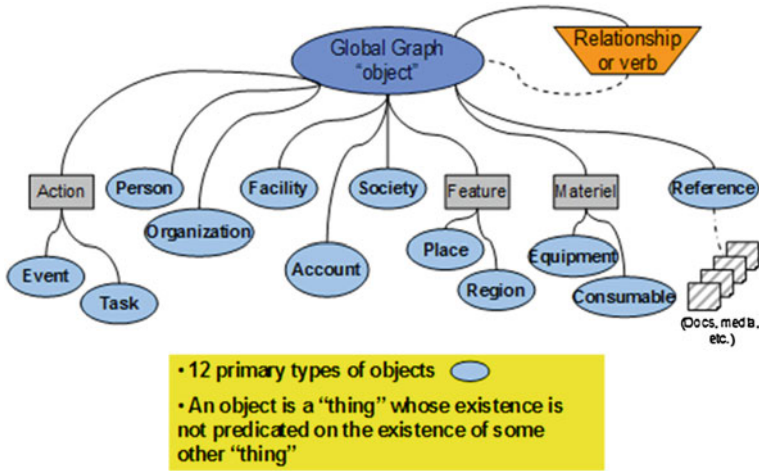


Fig. 4 U.S. Army’s “Global Graph” concept for DCGS-A (from [25])

In such an approach, the analyst forms a query in text that can be transformed to a graph (we call these “template” graphs in that they are subgraph structures of interest—a textual/graphical question in effect) that is then searched for in the associated-evidence graph that is formed by the DA process. This search operation is a stochastic, inexact graph-matching problem, since the nodes and arcs of the evidential data set have uncertainty values associated with them (or perhaps the template graph as well, if the query has stochastic/uncertain aspects), and also because what is sought is the best match to the query, not an exact match, since there may be no exact match in such unpredictable problem situations. Other complexities arise in trying to realize such capability, such as executing such operations incrementally for streaming data, and also doing them in a computationally-efficient way since the graphs can get quite large. As a consequence of several PhD efforts, we have realized today a rather mature graph-matching capability for intelligence analysis that is implemented in a cloud-based process; see [26–28], among other of our works.

6 Analytics, Sensemaking, and Decision-Making

We have noted previously that for the problems of interest here, those so-called “irregular” and “asymmetric” problems, the amount and reliability of a priori knowledge about the problem spaces is typically very limited. By and large, this means that analyzing the associated, multisource evidential data involves a mixture of strategies as has been suggested in Fig. 2. System designers and analysts must understand that there will be no singular tool or analytical technique that provides the “answer” at the level of abstraction desired. Such analysis environments are not entirely new to intelligence and military ISR analyses but these modern problems impose new and

additional difficulties in analysis methods and strategies. Commanders and analysts do not approach these environments totally absent of knowledge, and they usually have some type of focal topics and issues of interest. For commanders and analysts both, it is usual to have a set of Priority Intelligence Requirements (PIR's) that are ideally interrelated to anticipated Course of Action (COA) decision options. However, the action space for these problems involves the range of political, economic, military, paramilitary, psychological, and civic actions, i.e., not only "kinetic" actions involving the use of weaponry. As remarked previously, the decision space can thus also be labeled as "soft" in that the decision-space includes such decisions as those resulting in the realization of desired levels of influence (e.g., onto tribal leaders etc). It can be seen immediately from this definition that both the understanding of a current situation and its various elements, and the space of possible decisions and actions both have a much larger dimensionality than traditional military decision-making in force-on-force operations. Collectively, the broad elements of this action space can be broken into "direct" and "indirect" classes of actions, where direct actions are those focused on adversarial structure in the traditional military sense, and indirect actions those focused on undermining support to the adversaries while simultaneously attacking them militarily. It can also be argued that the End States of any decision sequence are "Effects" created by the sequence of actions (the COA). The concepts of Effects Based Operations (EBO), not a new term but actively revisited for these modern problems (e.g. [29]), shows that many references suggest that EBO is a viable concept for irregular/asymmetric problems, in part because effects are soft-type results, and subsume behavioral end-states, reflecting a human focus. One simple taxonomy of Effects is shown in Fig. 5 (from [30]), a main distinction being "Physical" versus "Behavioral", which could be equated to "Kinetic" versus "Non-kinetic".

The development of an interlinked COA to create these behavioral, non-kinetic Effects as end-states is very difficult and involves a web of interdependencies that make EBO a process involving notions of Complex Adaptive Systems (CAS). Smith [31] elaborates on this in various ways, and this CAS notion is also discussed in [32] that emphasizes the non-deterministic aspect of any Course of Action producing an intended Effect. Smith [31] has an extended development of the Effects-Based approach for asymmetric operations, and in consideration of what Smith calls an action-reaction cycle model (sensibly equivalent to Situation Management) puts forward a linked process that specifically shows the influences of understanding the Social Domain as part of the "Sensemaking" process that ultimately drives the COA development.

A very important notion (see [33]) is that the COA development process starts with a projected "Plausible Future" state so that actions are taken not necessarily on the basis of the current situation but one that *is* expected to exist at the time actions are taken on it, i.e., so that the situational state and actions onto it are as synchronous as possible. Note that, ideally, the DIF system should be supportive of some type of situational projection of such plausible future states, as part of an analysis suite. Supported under Air Force Research Laboratory funding, we have explored the ideas involved with, and the prototyping of automated DIF techniques for such estimation

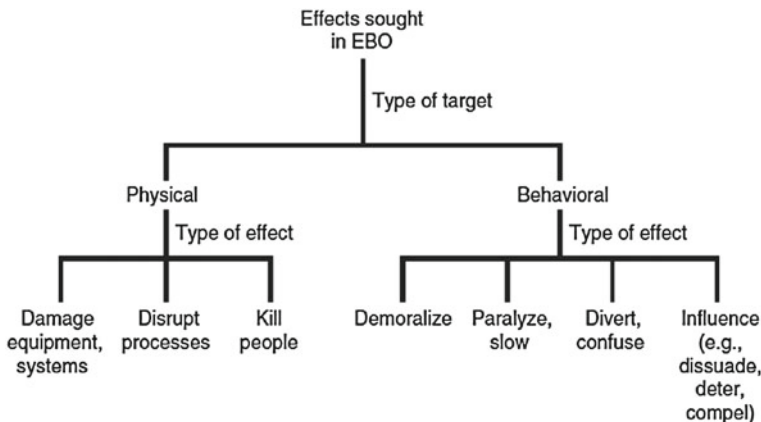


Fig. 5 Sample taxonomy of effects (from [30])

of plausible futures; see [34]. Additional remarks on the issues surrounding DIF, decision-making, COA development in the counterinsurgency environment can be seen in Llinas [35].

We see, as shown later, what today are called Sensemaking processes, as lying between DIF and DM processes, in a stage wherein “final” situation assessments and understandings (in the human mind) are developed. Thus, our view of this meta-process is as a three-stage operation: DIF as an automated process that nominates algorithmically-formed situational hypotheses (including nominations of “plausible future situations”), Sensemaking that dynamically interacts with DIF and human judgment in a kind of mixed-initiative operation to produce a final situational hypotheses upon which then DM operations are triggered. While there is also a substantive literature on Sensemaking, we address here three models: those of Pirolli and Card [36], of Klein et al. [37], and of Kurtz and Snowden [38]. The first two have many similarities and so we will show a figure of just one. These models depict Sensemaking as an iterative operation involving a hopefully-converging dynamic between a supporting information-space and an evolving situation hypothesis space. Here, the former is considered to be an automated DIF process and the latter is seen as occurring in the human mind, possibly aided by automated utilities. In [36], the overall Sensemaking process is “organized into two major loops of activities: (1) a “foraging” loop that involves processes “aimed at seeking information, searching and filtering it, and reading and extracting information, possibly into some schema, and (2) a sensemaking loop that involves iterative development of a mental model (a conceptualization) from the schema that best fits the evidence.” The Klein et al. Sensemaking model [37], called a Data-Frame Model, has many similarities to the process characteristics just described. The Kurtz and Snowden model, organized around their framework called Cynefin [38] is really based on the idea that categorizing the nature of the problem at hand, thereby partitioning it (in a “divide and conquer” strategy) and applying appropriate solution methods, is a part of the Sense-

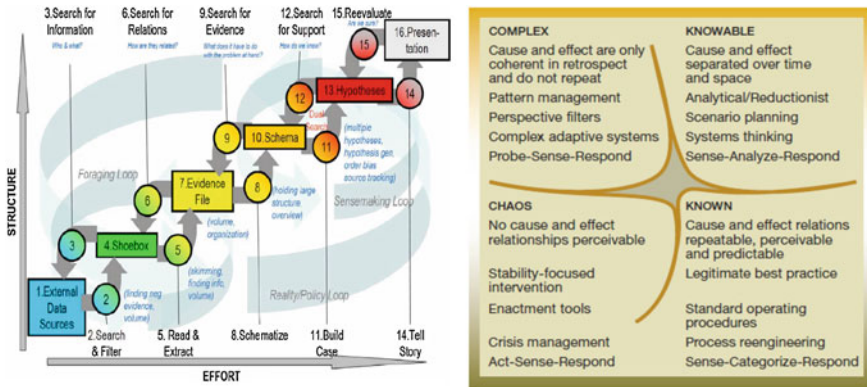


Fig. 6 Pirolli and card (left) and Kurtz and Snowden (right) models of Sensemaking (from [36, 38])

making process. Cynefin partitions problems into four categories called: Known (soluble by known methods), Knowable (soluble by analytical/reductionist techniques), Complex (soluble by what [38] calls Probe-Sense-Respond iterative discovery type processes), and Chaos (soluble by actions to reduce disorder, sensing the results, and responding or acting again). Cynefin takes a broader view of the states of complexity as ranging from order-to-complexity-to-chaos than do the models of Pirolli, Card or Klein; we include them because our concerns are for modern irregular, asymmetric warfare applications that often can have such properties. Diagrams showing the Pirolli/Card and Kurtz/Snowden models are provided in Fig. 6.

7 Connected Processes

So how do these processes interact, as we are asserting here? Fig. 7 shows a functional characterization of how:

- DIF, a largely-automated inferencing/estimation providing process that offers:
 - Algorithmically-developed situational estimates
 - Organized raw observational data—note these are hard (sensor) and soft (linguistic)
 - Controllable collection management of observational data
 - An Analytical Suite of useful but typically disparate tools
- Sensemaking, a semi-automated, human-on-the-loop process that:
 - Considers the DIF-provided estimates
 - Forages over these hypotheses as well as the data (e.g. drill-down etc)
 - Assesses the “Cynefin-category” nature of the problem at hand

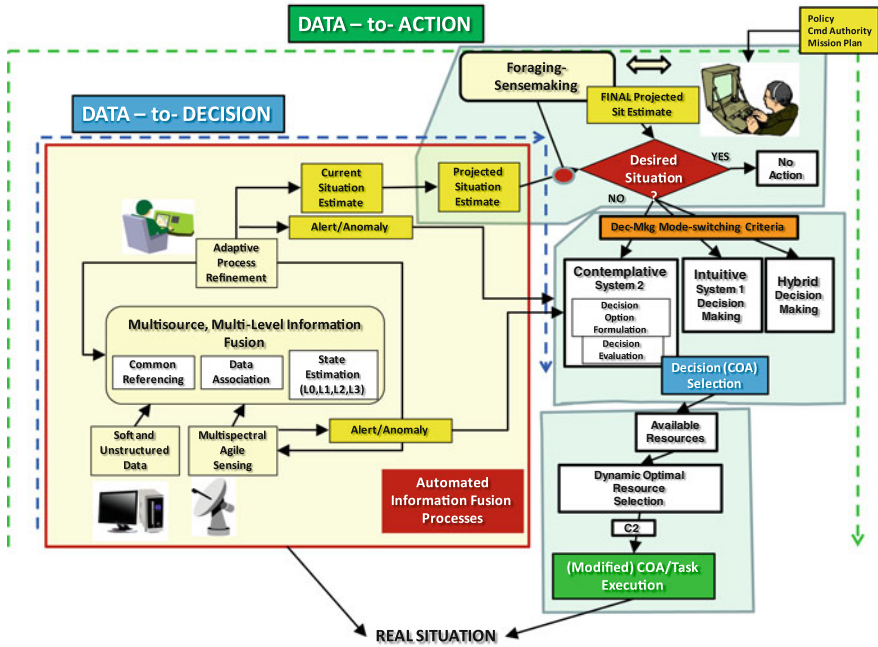


Fig. 7 Interconnected/dependent DIF-Sensemaking-DM-Resource Mgmt processes

- Considers possible Policy, Authority, and Mission factors
- Culminates in a “Final Adjudicated Situation Hypothesis” that is also judged as to acceptability; if not, this hypothesis is the starting point for decision-making and action-taking to “manage the situation”
- Decision-making, also a semi-automated, human-on-the-loop process that:
 - Operates in a System 1 (intuitive), 2 (contemplative, analytic) or “hybrid/mixed” DM mode
 - Yields a selected Course of Action
 - That triggers a Resource Optimization process to define specific resources that physically enable the selected COA onto the real-world situation

In [33], some of the issues regarding inter-process interdependencies were discussed (such as temporal dependencies), although that paper’s focus was on the various metrics involved across these processes. Another point we will make in this chapter is that yet another consideration related to decision-making is that most models of DM depict it as an analytical, contemplative process (analytical DM or ADM). It is important we think to realize that the DM community also discusses intuitive DM (IDM) that has considerably different properties than ADM. If we examine the disparate features of ADM and IDM, shown below in Table 1, we see that DIF process designs will need to be quite different to service the distinct functionalities

Table 1 Comparative features of IDM/System 1 and ADM/System 2 DM modalities (from [43])

Intuitive	Analytical
Experiential-inductive	Hypothetico-deductive
Bounded rationality	Unbounded rationality
Heuristic	Normative reasoning
Gestalt effect/pattern recognition	Robust decision making
Modular (hard-wired) responsivity	Acquired, critical, logical thought
Recognition-primed/thin slicing	Multiple branching, arborization
Unconscious thinking theory	Deliberate, purposeful thinking

for each DM mode. Thus, it can be argued that the DIF process should, ideally, be informed of the DM modality that users are in at any moment, so that, assuming the DIF design can be made DM-mode-sensitive, switch its operating mode to best service any DM mode at the moment. The DIF research community has conducted very minimal research on designing DM-mode-sensitive DIF processes; we see only two papers in the recent literature addressing this topic (see [39, 40]). In regard to IDM in particular, it could be argued that Case-Based Reasoning (CBR) techniques (similar to Klein’s RPD process, that enable intuitive, experientially-based inferencing and DM) might be a preferred inferencing mode in DIF for the IDM modality. While there are similarities between IDM and CBR/RPD, there is an important distinction for (probably most) modern operational domains about the notion of novelty in situations, and the true underlying capability of a human to deal with situations that are “seriously different” from their experience base. Naturalistic decision making using the RPD model fails in theory if there is a lack of experience or when encountering a completely novel scenario [41]. A review of most IDM models suggests that the inherent limits of IDM are the decision-makers personal range of situational experience combined with what has been “implicitly learned”. Any presented situation that is not adequately similar to this body of experience requires adaptation and learning. Boin et al. [42], state that “if the situation is radically different from those stored in memory, a somewhat different kind of sense-making process will be necessary.”

Another dependency area is between the DIF and Sensemaking processes. Clearly the Foraging function within Sensemaking implies that the DIF process will have to be open to, and enable, a range of queries that will be in regard to: raw or processed observational data, DIF functional operations (e.g., DA¹), and nominated situational

¹ The Army’s DCGS-A future system requirements for example include user-modifiable DA capabilities.

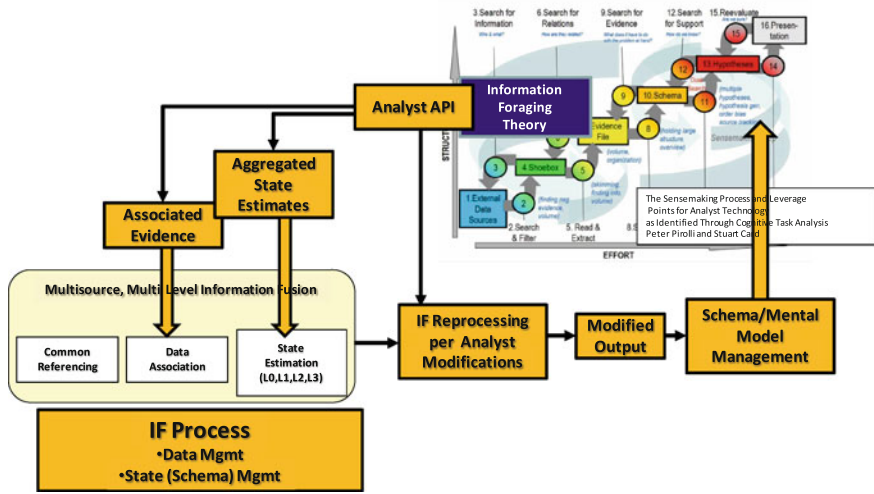


Fig. 8 Notional functional operations of the DIF-Sensemaking interactions

hypotheses, among possibly other runtime interactive operations. The notion of this interaction is depicted in Fig. 8, showing an analyst API that allows runtime modification of either or both of the Association and Estimation functions, then followed by IF reprocessing to generate new results that then get absorbed (possibly with automated support, not shown) into the analyst’s schema and mental models.

8 The Human Role in DIF, Sensemaking, and DM

This general process model can be seen to have at least two human points of involvement (assuming that the analyst is not the decision-maker). In our prototype hard-soft DIF system, we also have a possible role for a human in regard to editing the automated text extraction process for the soft/message data stream due to the considerable difficulty in achieving high quality extraction with automated methods. It can be appreciated that the complexity of natural language understanding, the complexities of the problems domain and the hard-soft fusion process all impute a serious consideration for placement of human intelligence in system design. The human role in DIF processes has been discussed for some time in the DIF community, and there are some works addressing the issues [44, 45]. In our judgment however, this general issue has been inadequately addressed at the community level, probably as a result of the DIF community having a quantitative bias, as can be seen in any review of community publications. The assertions and discussion here expands the challenge of addressing not only the human role in DIF but in Sensemaking and Decision-Making as well. The larger issue is a meta-system design question across the DIF-Sensemaking-DM

meta-process, as regards the placement of human intelligence and judgment for interpretation, control, and decision-making. The usual issues of quality of interpretation, quality of decision-making, quality of control versus timeliness need to be dealt with in developing approaches to designing this meta-system.

9 Summary

The world is dynamic in many ways. Looking at world politics and technology, no one should be surprised that there have been dramatic changes in the nature of security aspects driven by world politics; over the span of a decade or so, there should similarly not be any surprise that technology has advanced considerably. It is in this setting that this chapter was written, to offer perspectives on what those meta-changes have implied for the design and development of DIF systems as they sit in the interdependent environment with sensemaking and either analysis or decision-support systems. DIF system designers need to both take a larger view of their system's design but also reach out to and collaborate with those designing the related major functional capabilities for sensemaking and analysis and decision-making. DIF has always been a multidisciplinary area of study; this larger view further complicates that aspect but it is the opinion taken here that those interdependencies are inescapable, and that effective and efficient DIF designs can only be realized in the context discussed herein.

References

1. Bowman, C.L.: The dual node network (DNN) data fusion & resource management (DF&RM) architecture AIAA intelligent systems conference, Chicago, 20–22 Sept 2004
2. Prentice, M., Kandefer, M., Shapiro, S.C.: Tractor: a framework for soft information fusion. In: Proceedings of the 13th international conference on information fusion (fusion 2010), Edinburgh, Scotland, UK, July 2010
3. Prentice, M., Shapiro, S.C.: Using propositional graphs for soft information fusion. In: Proceedings of the 14th international conference on information fusion (fusion 2011), Chicago IL, USA, July 2011
4. Gómez-Romero, J., García, J., Kandefer, M., Llinas, J., Molina, J.M., Patricio, M.A., Prentice, M., Shapiro, S.C.: Strategies and techniques for use and exploitation of contextual information in high-level fusion architectures. In: 13th international conference on information fusion (fusion 2010), Edinburgh, Scotland, UK, July 2010
5. A. Kott, P.S. Corpac, COMPOEX Technology To Assist Leaders in Planning And Executing Campaigns In Complex Operational Environments, 12th International Command and Control Research and Technology Symposium: Newport. Rhode Island, USA (2007)
6. Kandefer, M., Shapiro, S.C.: Evaluating spreading activation for soft information fusion. In: Proceedings of the 14th international conference on information fusion (fusion 2011), Chicago, IL, USA, 5–8 July 2011
7. Smith, B., et al. Ontology for the intelligence analyst, in crosstalk. *J. Def. Softw. Eng.* 18–25, (2012)

8. Waltz, E.: Information understanding: integrating data fusion and data mining processes. In: Proceedings of the IEEE international symposium on circuits and systems, ISCAS '98, Ann Arbor, Michigan, USA, 1998
9. McMaster, D.: Temporal alignment in soft information processing. In: Proceedings of the 14th international conference on information fusion (fusion 2011), Chicago, Illinois, USA, July 2011
10. Klir, G.J.: A principle of uncertainty and information invariance. *Int. J. Gen. Syst.* **17**, 249–275 (1990)
11. Blasch, E., Llinas, J., Lambert, D., Valin, P., Das, S., Hong, C., Kokar, M., Shahbazian, E.: High level information fusion developments, issues, and grand challenges—fusion10 panel discussion. In: International conference on info fusion (fusion10), Edinburgh, Scotland, UK, July 2010
12. Oussalah, M.: On the probability/possibility transformations: a comparative analysis. *Int. J. Gen. Syst.* **29**(5), 671–718 (2000)
13. Geer, J.F., Klir, G.J.: A mathematical analysis of information- preserving transformations between probabilistic and possibilistic formulations of uncertainty. *Int. J. Gen. Syst.* **20**, 143–176 (1992)
14. Klir, G., Parviz, B.: Probability-possibility transformations: a comparison. *Int. J. Gen. Syst.* **21**(1), 291–310 (1992)
15. Dubois, D., Prade, H.: Unfair coins and necessity measures: a possibilistic interpretation of histograms. *Fuzzy Sets Syst.* **10**, 15–20 (1983)
16. Gross, G., Nagi, R., Sambhoos, K.: A fuzzy graph matching approach in intelligence analysis and maintenance of continuous situational awareness. *J. Inf. Fusion* **18**, 43–61 (2013)
17. Hall, D.L., Llinas, J.: Handbook of multisensor data fusion. CRC Press, Boca Raton (2001)
18. Poore, A., Lu, S., Suchomel, B.: Data association using multiple-frame assignments. In: Liggins M., Hall D., Llinas J. (eds.) Handbook of multisensor datafusion, chapter 13, 2nd ed., pp. 299–318. CRC Press, Boca Raton, 2009
19. Poore, A., Rijavec, N.: A lagrangian relaxation algorithm for multidimensional assignment problems arising from multitarget tracking. *SIAM J Optim* **3**, 544–563 (1993)
20. Tauer, G., Nagi, R., Sudit, M.: The graph association problem: mathematical models and a lagrangian heuristic. *Nav. Res. Logist.* **60**(3), 251–268 (2013)
21. Mark Pullen, J., et al.: Joint battle management language (JBML)—US contribution to the C-BML PDG and NATO MSG-048 TA. IEEE European simulation interoperability workshop, Genoa, Italy, June 2007
22. Schade, U., Biermann, J., Frey, M., Kruger, K.: From battle management language (BML) to automatic information fusion. In: Proceedings of the IF and GIS, pp. 84–95, (2007)
23. Lee, H., Zeigler, B.P.: SES-based ontological process for high level information fusion. *J. Def. Model. Simul. Appl. Methodol. Technol.* **7**(4), (2010)
24. Pollard, S., Biermann, A.W.: A measure of semantic complexity for natural language systems. In: Proceedings of NLP complexity workshop: syntactic and semantic complexity in natural language processing systems, Stroudsburg, Pennsylvania, USA, 2000
25. Walsh, D.: Relooking the JDL model for fusion on a global graph. In: National symposium on sensor and data fusion, Las Vegas, Nevada, USA, July 2010
26. Stotz, A., et al.: Incremental graph matching for situation awareness. In: 12th international conference on information fusion (fusion 2012), Seattle, WA, USA, 6–9 July 2009
27. Sambhoos, K. et al.: Enhancements to high level data fusion using graph matching and state space search. *Inf. Fusion* **11**(4), 351–364 (2010)
28. Gross, G.: Continuous preservation of situational awareness through incremental/stochastic graphical methods. In: Proceedings of the 14th international conference on information fusion (fusion 2011), Chicago, IL, USA, 5–8 July 2011
29. Davis, P.K.: Effects based operations: a grand challenge for the analytical community. RAND Report, Santa Monica, CA (2001)
30. Smith, E.A.: Complexity, networking, and effects-based approaches to operations, US DoD Command and control research program publication. www.dodccrp.org (2006)

31. Smith, E.A.: Effects-based operations: applying network centric warfare in peace, crisis and war. CCRP, Washington, DC (2002)
32. Senglaub, M.: Course of action analysis within an effects-based operational context, SANDIA REPORT SAND2001-3497 Nov 2001
33. Llinas, J.: Quantitative aspects of situation management: measuring and testing situation management concepts, intelligent sensing, situation management, impact assessment, and cyber-sensing. Proc. SPIE **7352** (2009)
34. Holsopple, J., Yang, S.: Fusia: future situation and impact awareness. In: Proceedings of fusion 2008: the international conference on information fusion, Cologne, Germany, July 2008
35. Llinas, J.: Situation management in counterinsurgency operations: an overview of operational art and relevant technologies. In: Proceedings of the 14th international conference on information fusion (fusion 2011), Chicago, IL, USA, July 2011
36. Pirolli, P., Card, S.K.: The sensemaking process and leverage points for analyst technology. In: Proceedings of the 2005 international conference on intelligence analysis, McLean, Virginia, USA, 2005
37. Klein, G., Moon, B., Hoffman, R.F.: Making sense of sensemaking I: alternative perspectives. IEEE Intell. Syst. **21**(4), 70–73 (2006)
38. Kurtz, C.F., Snowden, D.J.: The new dynamics of strategy: sense-making in a complex and complicated world. IBM Syst. J. **42**(3), 462–483. (2003)
39. Louvieris, P., et al. Smart decision support system using parsimonious information fusion. In: Proceedings of the 7th international conference on information fusion(fusion 2005), Philadelphia, PA, USA, July 2005
40. Solano, M.A., et al.: High-Level fusion for intelligence applications using recombinant cognition synthesis. Inf. Fusion JI **13**(1), (2012)
41. Diaz, S.K.: Where do I start? Decision making in complex novel environments, naval post-graduate school thesis, Sept 2010
42. Boin, A., et al.: The politics of crisis management: public leadership under pressure. Cambridge University Press, Cambridge (2005)
43. Croskerry, P.A.: Universal model of diagnostic reasoning. Acad. Med. **84**(8), 1022–1028, (2009)
44. Blasch, E.P., Plano, S.: JDL level 5 fusion model: user refinement issues and applications in group tracking. In: Proceedings SPIE 4729, signal processing, sensor fusion, and target recognition XI, p. 270, 31 July 2002
45. Hall, D.L., Jordan, J.M.: Human-centered information fusion. Artech House, Norwood, MA (2010)