Assessing uncertainty handling representations of HLIF systems with URREF

MARK LOCHER
PAULO COSTA

Researchers have extensively explored uncertainty issues in Low Level Information Fusion (DFIG L0/L1 process levels) systems, and predominately use probabilistic uncertainty representations. However, this prominence does not happen in High-Level Information Fusion (HLIF) systems. One reason for this discrepancy is that HLIF systems ingest a wider range of evidence, with its associated uncertainties, and execute a broader scope of inferential reasoning than LLIF systems. Researchers developed multiple techniques to address these uncertainties and reasoning needs, but it is not clear when and where in a specific fusion system a particular technique should be applied. ISIF established the Evaluation of Technologies for Uncertainty Reasoning Working Group (ETURWG) to provide some clarity on this issue. As a first step, the ETURWG created the Uncertainty Representation and Reasoning Evaluation Framework (URREF). The framework formally represents concepts and criteria needed to evaluate the uncertainty management capabilities of HLIF systems. It provides 26 criteria for evaluating the effectiveness and resource efficiency of a fusion system's uncertainty management capabilities. However, given the recency of the framework and the complexity of the issues it addresses, practitioners face difficulties in understanding where and how each criterion is applicable across a general fusion process environment, including a generic fusion system model. This paper's primary contribution is to address this gap by providing a discussion of the significant application factors and considerations regarding the usage of the framework, while providing examples of such usage in the process.

Manuscript received February 20, 2018; revised October 31, 2018; released for publication December 10, 2018.

Refereeing of this contribution was handled by Johan de Villiers

Authors' addresses: Department of Systems Engineering and Operations Research and the Center of Excellence in C4I & Cyber, George Mason University, Fairfax, VA, 22020 USA (E-mail: {mlocher, pcosta}@gmu.edu).

Both authors state that they have no conflicts of interest or financial interests affected by this work.

Available as an OWL file at http://eturwg.c4i.gmu.edu/files/ontologies/URREF.owl.

1557-6418/18/$17.00 © 2018 JAIF

1. INTRODUCTION

This paper describes the use of the Uncertainty Representation and Reasoning Evaluation Framework (URREF) in evaluating an information fusion system's ability to appropriately handle the various uncertainties that arise in the fusion process. Information fusion transforms information from different sources and different points in time into a unified representation that supports human or automated decision-making [8]. This decision-making focus demands that information fusion results are sound. Unfortunately, data sources used are often "inconclusive, ambiguous, incomplete, unreliable and dissonant" [59]. It is important to evaluate the different forms of uncertainty a fusion system has to deal with, where and how they occur, and the impact they have on the fusion processes and system outputs. The URREF provides a set of uncertainty definitions and evaluation criteria to support such an evaluation.

High Level Information Fusion (HLIF) is defined as the situation (L2) and impact (L3) levels of the Data Fusion and Information Group (DFIG) model [72], [5]. It is distinguished from L0/1, which is called Low Level Information Fusion (LLIF). LLIF has been widely explored and issues of uncertainty determination and propagation are extensively documented. It typically uses crisp data from homogenous, credible sources. Classical probabilistic uncertainty representations with fixed probabilities, rather than belief functions or imprecise probabilities, predominate in LLIF [31]. HLIF involves more complex environments, reasoning about complex situations, with a diversity of entities and multiple relationships between those entities. HLIF uses more diverse information sources, with significant evidential vagueness or ambiguity, and incompleteness and inconsistencies between evidence items. The credibility of individual sources may vary significantly. The community has developed a range of techniques and models to address these issues, but there is no consensus on how to compare their effectiveness and system impacts.

The International Society for Information Fusion (ISIF) chartered the Evaluation of Technologies for Uncertainty Reasoning Working Group (ETURWG) to provide a forum to collectively address this common need in the ISIF community, coordinate with researchers in the area, and evaluate techniques for assessing, managing, and reducing uncertainty [25]. The group developed the Uncertainty Reasoning and Representation Evaluation Framework (URREF) as a first step towards sound evaluation of uncertainty representations in HLIF systems. First documented in [13], the current version and associated documentation can be found at the ETURWG website.[1] These criteria focus on evaluating the effectiveness and resource efficiencies of the uncertainty representation(s) within a fusion system. The ETURWG does not expect URREF to identify a "silver bullet" technique that will adequately address all the

[1]Use of an ontology editor such as Protégé suggested.

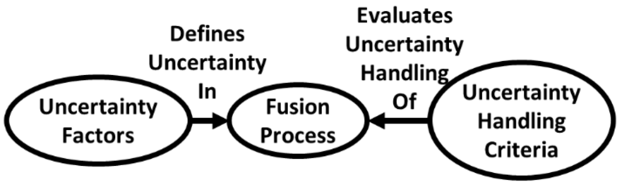


Fig. 1. URREF Top-Level Model

significant relevant uncertainties in a fusion system's environment but will assist designers in incorporating the appropriate range of techniques to meet their specific requirements.

This paper's primary contribution is to provide a discussion of the significant application factors and considerations regarding the usage of the framework, while providing examples of such usage in the process. Section 2 highlights the URREF and provides the criteria. Section 3 defines the key characteristics of the overall fusion process environment, including fusion system model, that affect uncertainty representation. Section 4 maps the URREF evaluation criteria to this environment and discusses how they are used to understand the uncertainty representation capabilities of a fusion system.

2. THE URREF

Figure 1 shows the URREF's top-level model. Uncertainty Factor provide a core description of the type, nature, derivation and models of the uncertainties that can be found in the fusion process. The Fusion Process includes the source, fusion system (in both a component and process view) and evidence/information.[2] These will be the subjects of an uncertainty handling evaluation. The Uncertainty Handling Criteria are measures useful for evaluating how well a specific fusion process handles its uncertainties. The ETURWG grounded the URREF on earlier work done by the W3C Incubator Group for Uncertainty Reasoning [47]. This work provides a basic framework of world/agent/sentence where an agent makes a statement about some aspect of the world using a logical sentence format. A logical sentence is a statement stated precisely enough that it can be assigned a truth value. This truth value may be binary, qualitative or numerical. The ETURWG identified three basic uncertainty characteristics: the nature, derivation and type of uncertainty, described in Table 1.

Although uncertainty has been understood qualitatively since the Greek philosophers of the early 5th Century BCE, an understanding of the different types of uncertainty began with the development of quantitative probability, addressing randomness, started by Fermat, Pascal and Huygens in the 17th century [3]. In 1921, Knight distinguished between problems with known probabilities (which he called risk) from those with unknown probabilities (called uncertainty—also a form of ignorance) [41]. The concepts of vagueness and ambiguity were given formal form by Black in 1937 [4]. Since that time, numerous taxonomies of uncertainty have been developed, both for general use and for specific fields. Jousselme et al. reviewed six taxonomies for potential application in fusion systems [35]. The two most comprehensive characterizations they identified were by Smithson [68] and by Krause and Clark [43]. Both use the classic randomness (probability)/vagueness/ambiguity classification. Smithson also included knowledge incompleteness and distortion as types of uncertainty. Distortion occurs when biases/inaccuracies in one's knowledge or when the knowledge transformation process introduces confusion in the knowledge [68]. Krause and Clark's taxonomy made two important distinctions. The first was between uncertainty induced by the classic sources and uncertainty induced by conflict. Second was the need to distinguish between uncertainty in a single information item and uncertainty in a set of information items. Conflict (also called inconsistency) most often occurs in an information set, although equivocation is identified as an internal conflict in a single item. Incompleteness is also primarily a characteristic of a set, although a single item may have missing information as well [43].

TABLE 1
URREF Uncertainty Factors

Uncertainty Nature	**Uncertainty is either inherent in the phenomenon expressed by the sentence or is result of lack of knowledge about that phenomenon.**
Aleatory	Uncertainty is inherent property of the world.
Epistemic	Uncertainty from lack of complete knowledge
Uncertainty Derivation	**Uncertainty derivation refers to the way it can be assessed. That is, how the uncertainty metrics can be derived.**
Objective	Assessed in a formal way, e.g., via a repeatable derivation process.
Subjective	Assessed via a subjective judgment. Even if one uses formal methods for this assessment, if the assessment involves subjective judgment, the Uncertainty Derivation is subjective.
Uncertainty Type	**Underlying characteristics of the information that make it uncertain.**
Ambiguity	Sentence has multiple possible interpretations
Vagueness	No precise correspondence between terms in the sentence and referents in the world
Randomness	The information comes from a process whose outcomes are non-deterministic.
Inconsistency	No world exists that satisfies the sentence.
Incompleteness	Occurs when information is missing.

[2]This paper will use the terms evidence and information interchangeably.

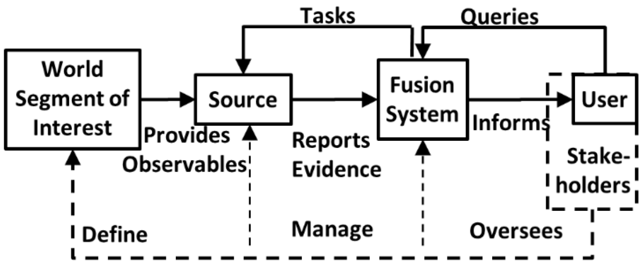


Fig. 2. Fusion Process Environment Model

Finally, the ETUWG identified the most common uncertainty representations (models):

- Belief functions[3]
- Fuzzy methods
- Probabilistic methods
- Random set
- Rough set

Additional choices can be found in Khaleghi et al. [40] or Castanedo [12]. The uncertainty handling criteria and their definitions are given in Table 2 below. The criteria are in four categories. Data criteria assess how a fusion process's design, including its uncertainty model(s), address aspects of uncertainty in data, both for individual items and for the collective set. Data Handling criteria focus on the effect of the uncertainty representation on explaining the reasoning used to create the output, and to maintain a record of what data was used in the process. Reasoning Criteria assess the overall approach to uncertainty handling in two areas:

- The correctness and consistency criteria assess the effects on the system outputs.
- The remaining criteria assess the effects on the overall system performance. These highlight the resource demands made by an uncertainty handling approach.

Representation Criteria assess internal characteristics of the uncertainty handling representation(s) and its integration with the fusion process.

It is an irony that the literature on uncertainty has a significant amount of ambiguity, redundant or overlapping terms, and conflicting definitions to describe aspects of uncertainty. In identifying these criteria, the ETURWG often had to select one term out of a range of choices for that aspect of uncertainty. In this paper, we generally do not attempt to identify synonymous terms or conflicting meanings.

3. FUSION PROCESS ENVIRONMENT

To apply the URREF criteria, one needs a model of the overall fusion process environment. We derived the model in Figure 2 from the DFIG model [5]. The main extension was to subsume the user in a larger group we call stakeholders, for reasons discussed below. This section describes each component, providing the context and key considerations for applying the URREF criteria.

3.1. Stakeholders/User

Any fusion system has a group of stakeholders, who collectively have an influence on the design and operation of the fusion system. The focus, scope and extent of a fusion system is driven by stakeholders' objectives, values and plans (collectively "stakeholders' interests"). A key subset of this group are the system users. These are the decision-makers, operators, and analysts who are the primary interactors with the system. Other stakeholders manage or influence aspects of the fusion process. For example, many fusion system users do not control the sources that provide evidence to their system. They submit information requests to one or more centralized management groups. Other stakeholders may require that the fusion process maintain records on how it created its outputs and the uncertainties associated with it. For example, the law of armed conflict requires a military commander to gather a reasonable amount of information to determine whether the target was a military objective and whether incidental damages to non-military targets are proportionate [48]. Uncertainties in the gathered information are a consideration in judging whether a commander acted properly. For such a system, the military legal community (as a stakeholder) may require that a fusion system be able to identify and trace the uncertainties in the evidence and how they were addressed in the fusion process to support a judgment of the legality of a commander's planned actions.

3.2. World Segment of Interest

The world segment is those aspects of a "real" world that stakeholders of the fusion system are interested in. Their points of view define the world segment. A world segment is defined as an area in the real world or cyber domain and possibly a time frame of interest (Figure 3).

The stakeholders' information needs define the world segments aspects of interest, including boundaries, key characteristics and entities of interest along with their attributes and relationships with other entities. In the same ocean area, a fusion system supporting a naval commander will focus on different entities than one supporting biologists studying marine mammals. The entities in the world segment generate observables, features detectable and reported by some source. Some entities may have a very limited set of observables, which may require a very specific approach to detect and collect the observable.

This is distilled into a world segment model using an ontological structure [9]. Entities should be categorized broadly, such as using Sowa's ontological categories. This allows for both concrete and abstract entities, with either time-stable (objects) or time varying (events) characteristics. It also allows for modeling

[3]Belief functions encompass approaches derived from Evidential Reasoning (Dempster-Shafer [62]). It includes Transferable Belief Model [66], Dezert-Smarandache Theory (DSmT) [18], and Subjective Logic [33].

TABLE 2
Uncertainty representation handling criteria

Data Criteria			Criteria on aspects of data and their relationship to source, reported object, and objectives of the fusion process.
	Credibility		Degree to which an evaluation subject can be believed or accepted as true, real, or honest
		Objectivity	Whether an evaluation subject reports in an unbiased manner
		Observational Sensitivity	Effect of a change in a result of evaluation subject and the corresponding change in a value of a quality being observed
		Self Confidence	Evaluation subject's assessment of its own credibility
	Quality		Assessment of the informational quality of the data
		Accuracy	Closeness of agreement between reported value and true value of reported entity
		Precision	Closeness of agreement between reported values obtained by replicate measurements on the same or similar evaluation subjects under specified conditions
		Veracity	Extent to which a source reports what it assesses to be the case
	Relevance To Problem		Degree to which information has direct bearing on the objectives of the fusion process
	Weight Of Evidence		Ability to assess the degree of impact of an evaluation subject on the result of fusion
Data Handling Criteria			Measure of the way data is managed by the fusion system
	Interpretation		Fusion system's ability to provide a coherent explanation that can be used to guide assessment, to understand the system's conclusions, and to provide a basis for reasoning and action
	Traceability		Ability to provide an accurate and unbroken historical record of its inputs and the chain of operations that led to its conclusions
Reasoning Criteria			How an information fusion system transforms its input data into knowledge. Focus on uncertainty model effects
	Computational Cost		Amount of system's computational resources required by a given representational technique to produce its results
	Consistency		Ability to produce the same results when provided with the same data under the same conditions
	Correctness		Ability to produce correct results, as measured against ground truth or an accepted gold standard
	Performance		Assess how well the fusion system and its representational model handle the functional requirements of an information fusion system
		Throughput	Measure the average and peak rate of conversion of inputs to outputs
		Timeliness	Ability to produce results within a required timeframe
	Scalability		Ability to handle an expanded work load
Representation Criteria			Encompasses criteria related to how uncertainty is characterized, captured and stored in a manner that can be processed by the fusion system
	Adaptability		Ability of the representational model to allow for different configurations of the model.
	Compatibility		Degree to which a given knowledge representation complies with data standards, and is related to the degree of flexibility it has in being coded with various standards
	Expressiveness		Ability to convey all relevant aspects of a given fusion problem
		Assessment	Ability to capture the types of uncertainty present in the evaluation subject
		Outcomes	Ability to represent appropriate scale for the outcomes
		Dependency	Ability to capture dependency among propositions (e.g., cause and effect, relevance, statistical association)
		Relational	Ability to represent uncertainty about domains with relational structure, i.e., domains in which there are types of objects with type-specific attributes and structure, having relationships to other types of objects
		Higher order uncertainty	Ability to capture uncertainty about the uncertainty model, including parameters, structure, and/or type of model
		Configurality	Ability to combine different types of uncertainty in multiple entities / relationships
	Knowledge Handling		Ability of a given uncertainty representation technique to convey knowledge
	Simplicity		User 's ability to execute common operations without requiring deep knowledge about its inner workings

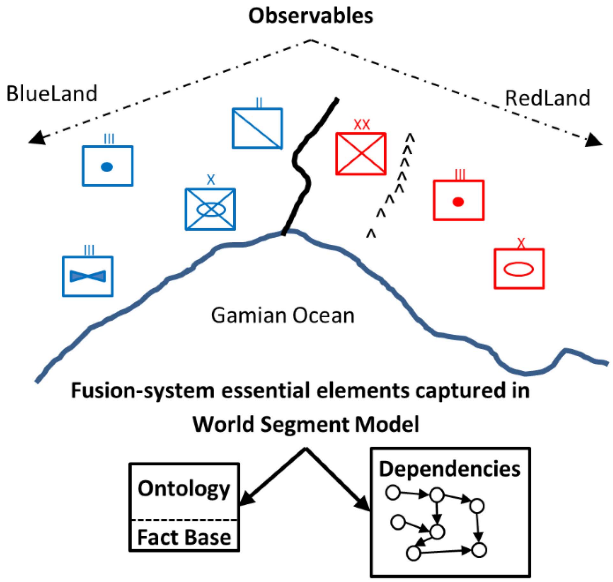


Fig. 3. World Segment of Interest defined by stakeholders/users needs and interest. Defines relevant observables, entities/attributes/relationships (captured in an ontology) and key dependencies that can infer new information

complex structures or situations, along with assigning attributes like purpose to them [70]. The world segment model generally becomes part of the fusion system, and mismatches between the world and model can result in significant errors and uncertainties. A key part of the world segment model is the dependencies. These are linkages between the attributes and relationships of entities, both within an entity and between entities. They have an "If A, then B" structure. The dependency between A and B is established from prior knowledge (include expert elicitation) or learned from collected evidence. The core of HLIF reasoning hinges on dependencies; when we have good reasons to believe A exists, then our understanding of B's existence, attributes or relationships change. Dependencies are expressed as rules, clauses (for logic programs) or graphical models (e.g. Bayesian networks, Markov networks).

A system may have multiple world segments within it (e.g. a global health epidemic system may be divided into regions or countries) or it may have multiple system copies, each with a different world segment. A system may also be deployable and load different world segments models as needed.

3.3. Source and Evidence

A source gathers observables and transforms them into evidence on some aspect of a world segment, through new observation or analysis of previously collected data (Figure 4). Source here means a specific mode of accessing data (e.g. panchromatic imagery, communications intercept, seismic detection, human reporting, database searches, etc.). When humans are part of the source process, at least some of the functions in Figure 4 are done mentally. Some source systems

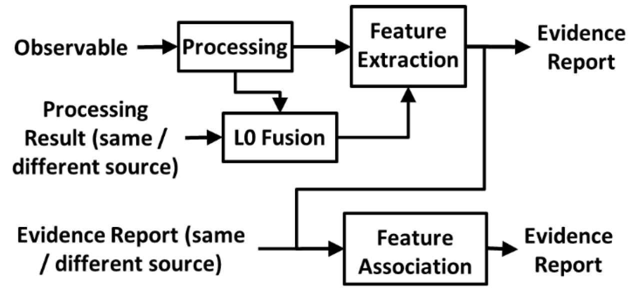


Fig. 4. Source process model

are multi-mode (e.g. radar with both Synthetic Aperture Radar and Surface Moving Target Indication modes) or multi-sensor (e.g. imaging and signals intercept on the same platform). Uncertainty should be assessed for each mode. A source may be dedicated to a specific fusion system or provide data to multiple fusion systems. A source may perform L0 fusion of observable samples (e.g. SAR change detection) using either internally generated data or integrating externally provided data. A source system may also conduct Level 1 fusion, using either self-generated or externally provided evidence. When data from those different sources are fused, the overall fusion process must be aware of this to avoid multiple counting of the same evidence.

A common source differentiator is the hard/soft distinction, which aligns with the URREF Uncertainty Derivation criterion of objectively or subjectively derived evidence. Technical sensors are considered to provide hard or objective evidence, based on a repeatable derivation process. They generally provide consistent data with little possibility of source-generated untruthfulness, bias or deception. Evidence developed from human reporting is considered soft or subjective, with issues of source credibility, including deception; significant use of vague or ambiguous terms, or inconsistent application of terms between individual human sources [37]. The distinction is useful but benefits from being refined. Many sources have a machine/human partnership, where the extraction of useful information is done by humans. Imagery and communications intercepts sources are two examples. Such sources are generally classified as hard sources. In classifying a source as hard or soft, there are at least four considerations:

- Degree of calibration. Almost all technical sources undergo some type of calibration prior to employment, to ensure a level of accuracy and consistency. For some sources, human data exploiters undergo training to provide a level of consistency across different individuals. This consistency may not be tight as for a technical source.
- Use of source quality standards and reporting reviews prior to evidence release.
- Source recording. If the source maintains a record of the data that generated the evidence, it can be

TABLE 3
Classes of Evidence

Unequivocal testimony	Statement from a source (written, verbal)
Equivocal testimony	Hedged source statement ("I think I saw….")
Tangible	Evidence that may be physically examined: e.g. objects, documents, images, recordings
Missing evidence	Evidence one expects to find but does not.
Accepted facts	Statements whose truthfulness as evidence is not questioned (e.g. gold has a higher density than iron).

reviewed in cases where there are questions about the evidence.
- Source quality improvement efforts to identify and correct deficiencies, adjusting their accuracy and credibility over time.

Each source has its own characteristics that define how it gathers and processes its data. The source model describes, to some level of detail, how the source gathers and processes its data. An accurate model for each type of source is necessary for doing an uncertainty assessment on that source.

Sources generate evidence that is used in the fusion process. Evidence can be expressed using logical sentences with an uncertain truth value (which include "100% true" and "0% true"). Evidence can take a variety of forms. Table 3 provides a classification scheme [61]. Testimony is a statement made by a source. The statement may be based on direct observation, or on secondhand sourcing/hearsay. The statement may be either unequivocal ("It is the case that…") or equivocal ("I think that….", "I'm not positive, but…"). An equivocal assertion may include a reason for the equivocation ("It was dark, but I'm pretty sure I saw…"). Opinion is a form of equivocal testimony. It is defined as "A view or judgement formed about something, not necessarily based on fact or knowledge" or "A statement of advice by an expert on a professional matter." [53]. The key here is whether an opinion statement comes from a competent and knowledgeable source, able to support that statement. Expert judgment is a form of opinion that is a valid form of evidence. Missing evidence is not negative evidence, which is evidence that something does not exist at a point in time one is interested in. In some cases, missing evidence can be significant. For example, evidence intentionally destroyed can have a negative connotation for the destroyer.

Evidence may be at any level of the DFIG model, and it does not have to come from a process that moves sequentially through the levels. While sensor-derived data goes through L0 processing, human derived data often does not (although some may go through a form of preprocessing, such as summation or statistical processing). Evidence, especially from human or communications intercept sources, can also be about relationships between entities, situation or structure identification, or intentions (specific plans and objectives).

3.4. Fusion System

Understanding how to apply the URREF criteria to a HLIF process benefits from a generic system fusion model allows aligning the criteria with fusion system processes/components. After initially exploring the literature, we established these model requirements:

- Identifies key functions within a fusion process.
- Maps the flow between the functions, including feedback and reevaluation requests.
- Allows varying human/machine divisions of effort.
- Is not bound to a specific uncertainty representation or fusion methodology.
- Uses general domain-independent terminology.

According to Salerno, over 30 fusion process models had been proposed by 2002 [58]. Several teams have reviewed selected subsets, including Esteban et al. [24], Bedworth and O'Brien [2], Whitney, Posse and Lei [78] and Roy et al. [57]. Foo and Ng published an updated review in 2013 [26]. We found most of the models before 2005 very limited in their functional description. These included Pau's Model [55], Intelligence Cycle model [2], Thomopoulos' model [74], JDL model [30], Dasarathy model [16], Waterfall model [2], Extended OODA loop [63], Omnibus model [2] and the General Data Fusion Architecture [11]. Although they also had limited functional details, models that incorporated humans as part of the fusion process included the Visual Data-Fusion model [39], JDL level 5 [7], Endsley's situation awareness model [22], [23] and Lambert's Unified Data Fusion Model [44]. Four models published between 2002 and 2016 included significant details about their functions, shown in Figure 5. They were by Salerno [58], Steinberg [71], Lambert [45] and García, Snidaro, and Llinas [27]. There is a high degree of commonality in the functions described. All have some form of data ingest function that performs reference base alignment and semantic (ontological) registration. Some models explicitly depicted entity extraction from unstructured information sources (e.g. free text reporting). García et al.'s model was the only one to explicitly depict an uncertainty characterization process, while Steinberg's model discussed it in the text describing the model. Salerno's model explicitly depicted a number of information development activities to support the overall fusion process, including

- Data mining activities, including link analysis, pattern learning and pattern matching.
- Model development support, including pattern identification and model generation. Models may be built ahead of time, or created from the data stream.

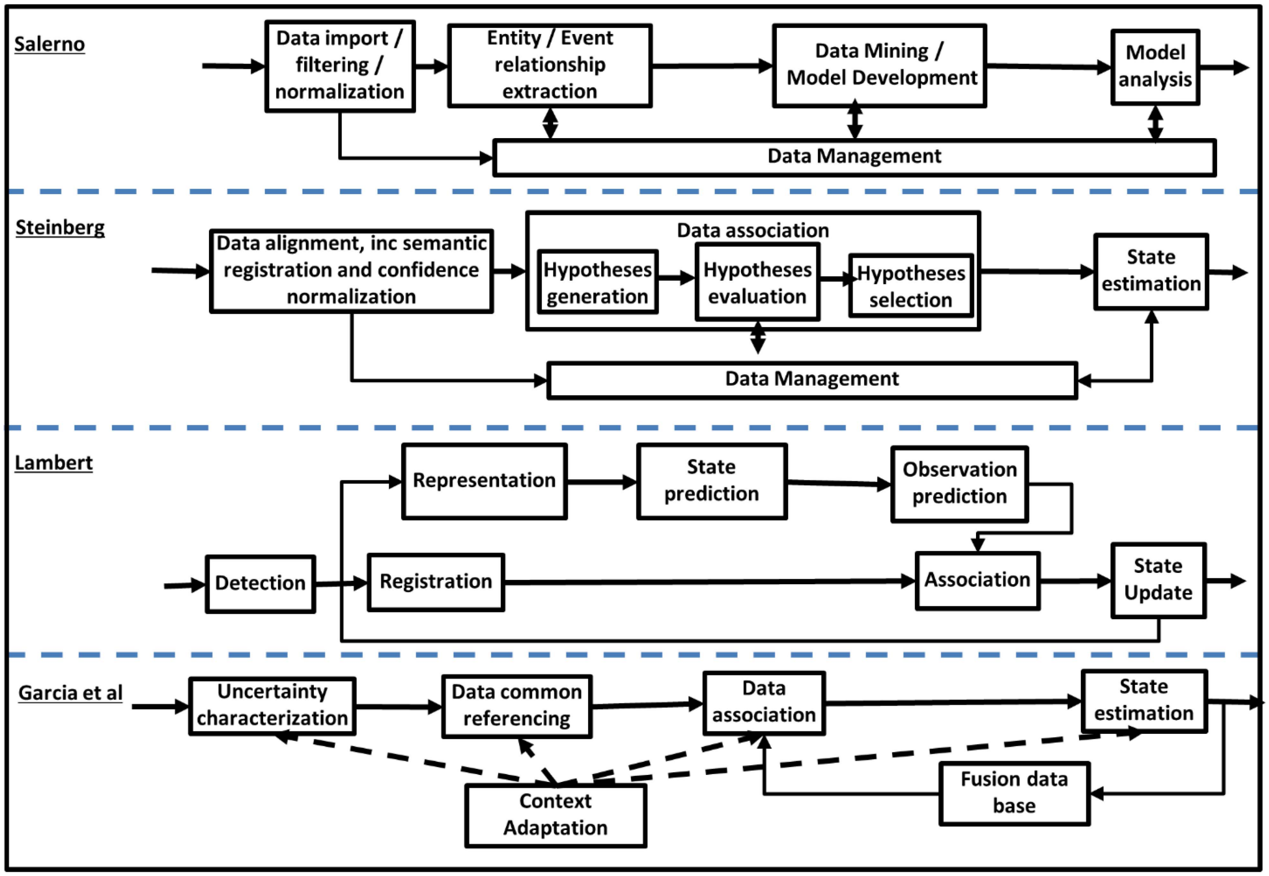


Fig. 5. Functional process/component elements of four major fusion system models

The other three models call out these functions as data association. For example, at level 2 HLIF, Steinberg model focused on finding and estimating relationships in the data, expressed as possible hypotheses. This is done by three subfunctions: hypotheses generation, hypotheses evaluation and hypotheses selection.

All four models had state estimation or state modeling. For HLIF, this process can use a variety of techniques, including link analysis, graph matching, templating methods, belief networks, compositional methods for model detection and development, and various algorithmic techniques [71].

Lambert's model differed from the others in using state transitions as a focusing element. This concept extends the idea of a Kalman filter to observing, predicting and updating state data, including tracking which scenario is being executed (L3 fusion) [45].

Because of differences between soft and hard sensors García et al.'s model have data from each type flow through a distinct path designed for the characteristics of that data [27]. They also explicitly include the use of context information. In the last five years, there has been significant work done on incorporating contextual information such as map data, weather, and procedural data (e.g. traffic rules, doctrinal concepts, patterns of life, hierarchies) for HLIF. Such non-sensor information can be used to both constrain and explain behaviors seen in sensor data [27] [67] [69].

To identify where to apply the various criteria, we merged these four models together to create the generic fusion system model shown in Figure 6. Based on our criteria, we realized that we needed to explicitly include several processes that one or more models discussed in their text but did not include in their visual model. The model assumes that input data may be L1, L2 or L3 data, including contextual data. The model has eight basic processes. Many source systems transmit free text reports, not structured text. Some form of entity and relationship extraction is required to transform those reports into machine-understandable data. The *Data Extraction/Alignment/Registration* process does this, including named entity recognition, coreference resolution, relationship extraction, and event extraction [56]. It also aligns the incoming data to a common reference base and ontological structure, appropriate for follow-on use. If the data is already structured according to an understood ontology, then this process is unnecessary.

For incoming evidence, *Source Uncertainty* manages all aspects of source uncertainty, as described in section 4.3. The *Data Store* captures all incoming evidence for access by the various processes. This includes both current and historical source evidence and reference information such as maps and equipment capability records.

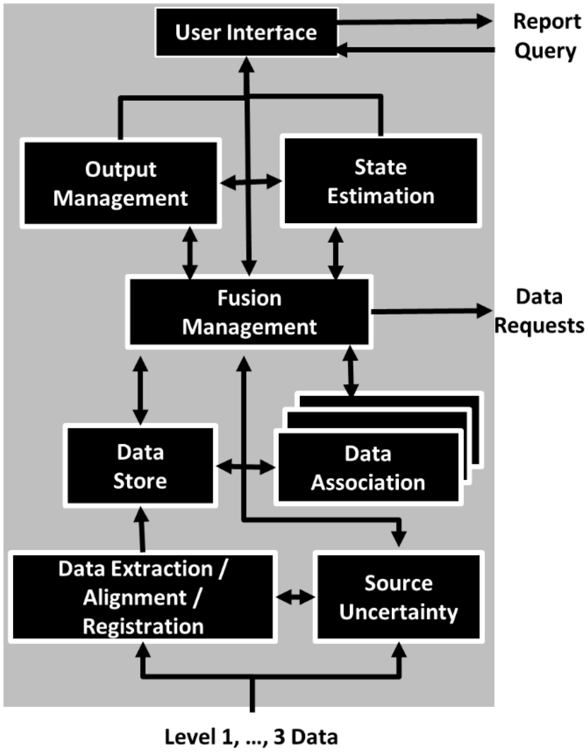


Fig. 6. HLIF fusion system model

An important aspect of this model is that not all the information is assumed to be in an immediately usable form for high level fusion processing in the State Estimation module. *Data Association* provides one or more services in which some or all of evidence, include context information, undergo to have the appropriate information extracted from them. For example, a fusion reasoning process may require relationship information. But the raw level 2 data may be a series of people association data, which must be combined into a social network analysis to reveal the full extent of the relationships. A key distinction between LLIF and HLIF is the significantly broader range of information in a HLIF, requiring a diverse set of data association processes to create that information [58], [24], [14], [71]. These processes can be implemented via middleware services [69].

Fusion Management involves all activities necessary to marshal information for the various fusion processing components and to sequence the fusion processes. This function can use multiple schemes to arrange the information to best provide insights into potential reasoning arguments and output hypotheses. It also identifies what additional information is needed to in the fusion process, and requests it [60].

The *State Estimation* process is the core of the fusion process. This process can take one or both of two forms. In less complex HLIF systems, it takes some form of direct symbolic reasoning, often a model-based process. To account for the uncertainty in the data and process, current models often take the form of Bayesian networks [71], [15], [46], although alternative approaches have been proposed using graphical belief models [1] and general-purpose graphical modeling using a variety of uncertainty techniques [64]. For more complex situation assessments, such as forensic reconstruction, the reasoning management process is a meta process, responsible for constructing the model used to provide the response. As such, there is a close interaction between reasoning management and output management.

The seventh process is *Output Management*. This process maintains the active hypotheses under consideration. It provides the output interpretation process (how did system arrive at this conclusion) and the traceability function (what evidence and functions did it use). It also is involved in generating hypotheses and in the pruning of hypotheses [32], [49].

The final process is the *User Interface*, which provides the information output and accepts user queries.

4. UNCERTAINTY ASSESSMENT

This section describes where and how the URREF criteria in Section 2 are applied to the process described in Section 3. The focus is on HLIF systems, but the criteria can also be applied to Level 0/1 systems as well. They do not cover the fusion management process levels (L4/5/6). These criteria guide fusion system developers and assessors through a comprehensive assessment of how well their uncertainty representations addresses the uncertainties both embedded in the evidence and generated by the fusion system's processes. Of the 26 criteria, thirteen can be specified as quantitative uncertainty measures, while the other thirteen are qualitative measures.

4.1. Stakeholder/User Uncertainty Tolerance Assessment

Identifying the stakeholders' concerns should drive the overall system uncertainty assessment. The first need is to understand their sensitivities to different kinds of uncertainties in the system. This focuses the main areas of evaluation, including the relative importance of different types of uncertainty. A second consideration is the uncertainty—system effects trade-off of addressing the various uncertainties via different uncertainty handling representations. Collectively, this information will focus and scope the uncertainty handling assessment.

4.2. World Segment Uncertainty Assessment

A fusion system uncertainty evaluation assesses the world segment to understand two important items:

- The uncertainties inherent in the observables.
- The uncertainties in the world segment model.

Uncertainties exist as variability in the world segment's observables and can propagate to the accuracy and precision of the collecting source. One needs to know the types and nature of these uncertainties. For any fusion system assessment, one must assume the world segment has a factual state. It is possible that the ground truth of that state may never be completely known, but it must be estimable well-enough to conduct meaningful assessments on the overall performance of a fusion system. The key component here is the world segment model. This model is a central part of fusion system, used both in data association and state estimation. Any model is an abstraction of a reality, and the fit with reality is imperfect. The key question is whether the fit is good enough. This is part of an overall assessment of the suitability and acceptability of a fusion system. For the uncertainty assessment, the primary question is whether the world segment model incorporates the key uncertainties inherent in the world segment. These will propagate through the source and into the fusion model, affecting both the correctness of the output and the demands placed on the fusion system's resources to address those uncertainties [17].

Second, epistemic uncertainties exist in world segment model and affect both the fusion system's output's correctness and consistency criteria and the data input's relevance criterion. In addition, limits on the expressiveness of the world segment model can induce uncertainty. The three characteristics are dependency uncertainty, higher order uncertainty and relational uncertainty. Dependency uncertainty occurs when there is significant doubt about the existence of or strength of the dependency between two or more world segment elements. This is a problem encountered during the model building effort. While the exact degree of dependency is often uncertain, the issue here is when is the uncertainty significant enough to affect the outcome (often detected by a sensitivity analysis). This leads to epistemic uncertainty because one does not know whether the model should include the dependency, or what strength value should be assigned to dependencies that are possible but not required (e.g. a probabilistic dependency). Higher order uncertainty is when one has significant doubt about the quantification values assigned in the model. All uncertainty representations require some form of quantification (e.g. basic probability assignments, membership functions). It is very possible to have uncertainties about the specific quantification scheme. This also leads to an epistemic uncertainty about the outcomes. Relational uncertainties occur in world segment models that allow for a varying number of entities and relationships. If so, then sources may make mistakes in assigning observables to entities. The evidence, including extracted information, will then have relational uncertainties. This can also occur in the fusion system when associating multiple evidence from different sources, or from the same source at different time periods. These are also a significant form of epistemic uncertainty in HLIF systems. There are five types of relational uncertainty:

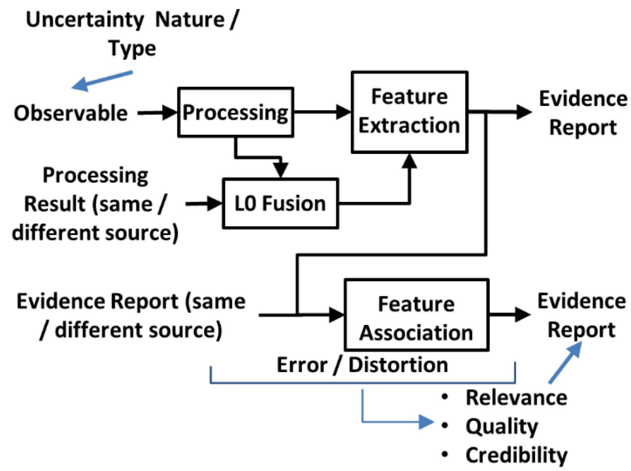


Fig. 7. Source errors and distortions combine with the uncertainty in the observables to create relevance, quality and credibility uncertainty

- Existence uncertainty for a key relationship or entity [28].
- Reference uncertainty is a dependency between two entities, but which specific entity has the dependency is uncertain (from a choice of several possible entities) [28].
- Type uncertainty is when one has determined the existence of an entity, but its reference class is uncertain [42].
- Identity uncertainty occurs when one is not certain if an entity is a new instance or one that has been previously identified [54].
- Number uncertainty occurs when the number of possible entities varies in a specific situation [52].

The primary effects of these uncertainties are seen when comparing the outputs of the fusion system to ground truth estimates in the world segment. This will be taken up in Section 4.4.2.

4.3. Source Uncertainty Assessment

Source uncertainty assessment focuses on the uncertainty in the evidence. The source ingests the variability, vagueness and ambiguity inherent in the observable. In the process, it often reduces the effects of variability, but can add uncertainty via process errors/distortions/limitations, especially for human-involved sources (Figure 7). For example, vagueness occurs when the source cannot apply a quantitative value to the observable. The discussion below follows Schum's classic work on evidence analysis and effects in probabilistic reasoning [59]. There are two basic questions when assessing uncertainties regarding evidence from a source:

- Is it relevant to the issues of interest to the fusion system's users?
- Is the evidence right?

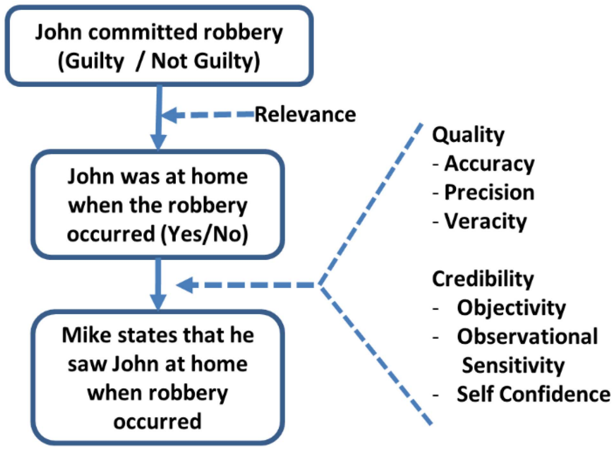


Fig. 8. Robbery Scenario

We use the example in Figure 8 to illustrate applying the criteria. John is accused of committing a robbery. If he did so, he would not have been at home when it occurred. If he did not do it, then he may or may not have been at home. This makes knowledge of John's whereabouts relevant to whether he committed the crime. A useful definition of relevance comes from the US Rules of Evidence [76]:

"Evidence is relevant if:

(a) it has any tendency to make a fact more or less probable than it would be without the evidence; and

(b) the fact is of consequence in determining the action."

Relevance measures the force of an item of evidence on some intermediate or final output of reasoning process. Probabilistically, relevance means that for a specific hypothesis H and any information E that could affect the belief in that hypothesis:

$$\text{Relevance} \stackrel{\text{def}}{=} P(H) <> P(H \mid E) \tag{1}$$

Relevance, as force of evidence, is always conditional on a particular hypothesis. It is not an inherent source characteristic. But we introduce it here because source uncertainty can modify the force of the evidence, sometimes in surprising ways. Relevance assumes a piece of evidence is true. There are several relevance measures in the literature [21]. The Bayes factor is one measure of the force of evidence:

$$\text{Relevance} = \frac{P(E \mid H)}{P(E \mid \bar{H})}, \quad (\bar{H})\text{is the complement of } H \tag{2}$$

In Figure 8, we have a testimonial statement from Mike that he saw John at home at the robbery. Is his statement right? This is assessed by the Credibility and Quality criteria.

Credibility assesses the source's ability to understand the information in the observables. Although Credibility is most applicable to human sources, there are elements that may occur with technical sources. It

TABLE 4
Credibility measures

Credibility	
Objectivity	u_{o}(Source Understood State | Competence, Bias)
Observational Sensitivity	u_{os}(Source Understood State | Environment, sensor factors)
Self Confidence	u_{sc}(Source Understood State | Source Equivocation)

has three subcriteria: Objectivity, Observational Sensitivity, and Self Confidence. Table 4 provides mathematical measures for each, where u is a general uncertainty measure which assigns a value between 0 to 1. This measure represents common measures of uncertainty (probability, belief, fuzzy or possibility measures). These measures represent a dependency, where "|" is "Given" or "If", modeling "If B, then A." If u_x is a probability measure where A and B have discrete states, "|" becomes the conditioning operator, and u_x is measured via a conditional probability table on A and B's states (e.g. a confusion matrix). Observe that these measures focus on what the source understands from the observable, not what it reports. Objectivity has two elements: competence and bias. Competence addresses two areas. One, did the source have the access and ability to observe what the source reported? Ability in this case refers to the source's general capabilities. Two, in the case where the source is providing an opinion, does the source has the competence and data necessary to make the judgment expressed in the opinion. Incompetent sources cannot make an objective statement. Bias is any source characteristic that affects the source's ability to objectively understand the received data and influences them to ignore or misinterpret the data. Both human and technical bias are well-documented in the literature. Both can be hard to detect, especially if one is not looking for them. Bias can also be dependent on what is being reported on.

Observational sensitivity complements objectivity by noting when adjustments need to be made for situation-specific differences. For example, descending darkness near the time of the robbery could impair Mike's ability to correctly identify John. Technical sensors can also suffer from transient environmental effects that impair but not eliminate the ability to detect an observable. Self-confidence is the criterion that assesses equivocal evidence. This is evidence where the source specifically casts doubt on the accuracy of what it is reporting. Human sources may use vague or nonspecific phrases such as "It was getting dark, so I'm not sure…" or "I think it was him." Technical equivocation occurs when a source reports using abnormal sensor settings, system limitations or releasing below normal quality standards. Both human and technical equivocation affect the fusion system's understanding of the

TABLE 5
Quality measures

Quality	
Accuracy	u_{acc}(Reported State | Actual State)
Precision	$u_{\mathrm{pre}} = \sum_{\mathrm{trials}}$ Loss Function (Reported, Actual
Veracity	u_{ver}(Reported State | Source Understood State)

source's accuracy and requires an adjustment for that specific evidence.

Quality recognizes that even trustworthy sources makes mistakes. Quality has three subcriteria: Accuracy, Precision and Veracity. Quality measures are in Table 5. In these measures, the focus is on what the source reports.[4]

Accuracy assesses how close the reported information is to what is true in the world segment. It recognizes that no source is infallible. Whether technical or human, there is always the possibility that a source makes a mistake, with no intention to do so. Confusion matrices, Receiver Operating Characteristics, or Precision/Recall are all measures of accuracy. The Precision criterion complements Accuracy by assessing the degree of measurement variability between repeated observations of the same or similar entities under similar conditions. It is a measure of the consistency of the observation.[5] Precision is related to variability in the sensing environment, which can change a sensing measurement over time. A source with low precision will vary significantly more than a high precision source, decreasing the confidence one may have in the evidence. Veracity measures whether the source believes it is telling the truth (even if the evidence statement itself is not true). As such, Veracity is applicable to sources that have humans in a significant judgment role.

The evidential force of a source report as a standalone item depends on a function of its relevance, credibility and quality. The predominant understanding of credibility and quality is that they reduce (discount) the evidential force. But not always. Schum's explorations of the effects of veracity and credibility show that under some circumstances, knowledge about credibility and veracity factors can give more evidential force than the evidence contents themselves [59].

Figure 9 extends the model in Figure 8 to demonstrate this, giving two approaches to modeling veracity effects. In both, the prior probability of John's guilt is 10%. If John is guilty, he could not have been home at the time of the robbery. If not guilty, there is still a 70%

[4]Which is why Veracity was classified as a quality criterion, not a credibility criterion

[5]The term Precision has at least three different uses in uncertainty discussions. The one given is the most common. Other uses include the proportion of true positives out of the total items classified as true in a confusion matrix (precision/recall), and the value of the least significant digit in a measurement.

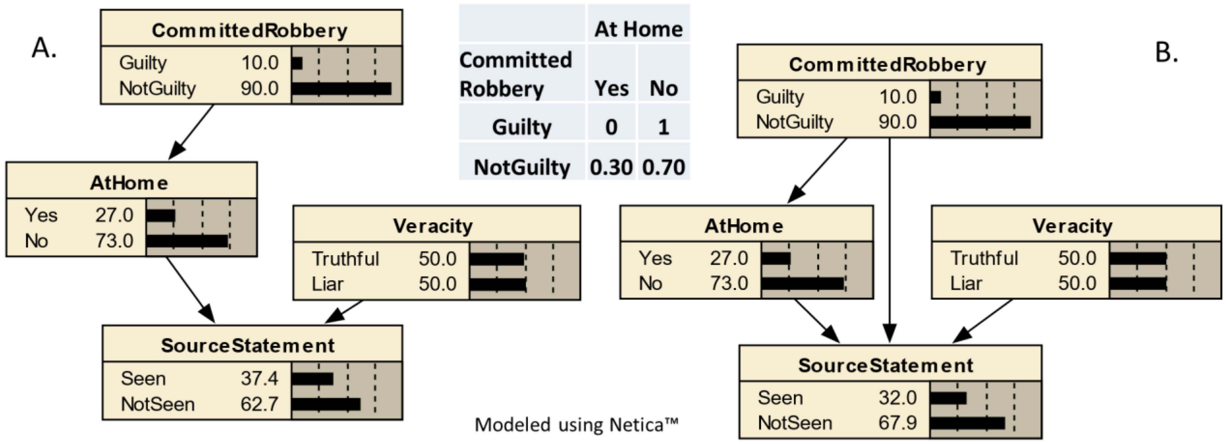
TABLE 6
Results of two different credibility models

Common data	**Guilty**	**Not Guilty**
Initial Belief (priors)	0.10	0.90
At Home—Yes	0	1
At Home—No	0.14	0.86
Single Thread		
Truthful—Source "Seen"	0.02	0.98
Truthful—Source "Not Seen"	0.13	0.87
Liar—Source "Seen"	0.09	0.91
Liar—Source "Not Seen"	0.11	0.89
Multi-Thread (Mike may know John's role in robbery)		
Truthful—Source "Seen"	0.02	0.98
Truthful—Source "Not Seen"	0.13	0.87
Liar—Source "Seen"	0.17	0.83
Liar—Source "Not Seen"	0.06	0.94

chance he was not at home at the time of the robbery. Finally, if Mike is a truthful witness, his accuracy is 95%.

Figure 9A gives a classical discounting approach, using a single thread model. Here, the source Mike is a suspected liar, and the probability of his evidence being true in either case is assessed at 60%. In Figure 9B, one suspects that Mike has some knowledge about whether John committed the robbery, and that he is willing to lie to protect John if John is guilty. If he has some knowledge that John is not guilty, he will tell the truth about what he observed (he will not risk perjury in this case). If John did commit it, Mike has only a 60% chance of telling the truth (we are not certain he will lie). Because Mike's statement has a dependency on whether John is guilty or not, as well on whether John was at home, this is a multi-thread model. Table 6 gives the results of the two models. First, see the effect of knowing for certain whether John was at home or not. If he was, then he is not guilty. If he wasn't, then the probability that he is guilty increases from 10% to 14%.

Second, in both models, a truthful source has the same result: 2% guilty if Mike says he saw John at home, 13% if he says he did not. This is a dilution of the 0%/14% result of John's actual state and results from the 95%/5% accuracy distribution. Now look at the liar results. In the single thread model (Mike has 60% of telling the truth in any case), one sees a further dilution of the relevance. It stays closer to the 10% prior probability than the case where the source is credible. But there is a surprise in the multi-thread case. If Mike lies when he knows John committed the robbery and says that he saw John at home, the probability of being guilty climbs to 17%. This is opposite of what happens in the truthful case. This is because



Committed Robbery	At Home Yes	At Home No
Guilty	0	1
NotGuilty	0.30	0.70

Veracity	At Home	Source Statement Seen	Source Statement Not Seen
Truthful	Yes	0.95	0.05
Truthful	No	0.05	0.95
Liar	Yes	0.60	0.40
Liar	No	0.40	0.60

Veracity	Committed Robbery	At Home	Source Statement Seen	Source Statement Not Seen
Truthful	Guilty	Yes	0.95	0.05
Truthful	Guilty	No	0.05	0.95
Truthful	Not Guilty	Yes	0.95	0.05
Truthful	Not Guilty	No	0.05	0.95
Liar	Guilty	Yes	0.60	0.40
Liar	Guilty	No	0.60	0.40
Liar	Not Guilty	Yes	0.95	0.05
Liar	Not Guilty	No	0.05	0.95

Fig. 9. Simplified robbery scenario with a suspected lying witness

if we think that someone who has knowledge about the ultimate hypothesis we are seeking will lie under certain circumstances, then telling the lie increases our probability of the ultimate hypothesis being true if the lie is told. There are many subtleties like this in doing source evidence assessments. See Schum [59] for an in-depth discussion on this issue.

The bottom line is that all source uncertainty assessment must determine to what these quality and credibility issues exist in their sources and select an uncertainty model and associated representation that address all the significant issues.

4.4. Fusion Model Uncertainty Assessment

Fusion model assessments focus on uncertainty representation in three areas: input evidence, output information, and the components of the fusion system. Figure 10 maps the URREF criteria in Section 2 to the fusion model in section 3.

4.4.1. Input Uncertainty Assessment Criteria

The input uncertainty assessment criteria can be divided into two categories: criteria applicable to individual evidence items and those for the collective set of evidence. For individual evidence items, the criteria are Credibility, Quality, and Assessment (an Expressiveness criterion). Credibility and Quality were discussed in the previous section. Assessment evaluates whether the fusion system can appropriately address the range of uncertainty types in the evidence. Uncertainty types identifies the basic uncertainty introduced by the world segment uncertainties and the specific characteristics of the source's process. In the fusion model, the characteristics of source evidence establish the uncertainty models needed for the source uncertainty, data association and state estimation modules. For individual evidence items, the source uncertainty module has the primary responsibility, since it establishes the credibility and veracity of each item.

In addition to uncertainty in the individual items of evidence, there is also uncertainty associated with the collective set of evidence. There are three criteria that apply: Assessment, Relevance, and Weight of Evidence. Assessment evaluates the fusion's system's ability to address the uncertainty types of incompleteness and inconsistency. Incompleteness is missing data, either partial (missing fields in a piece of evidence) or entirely. The most likely causes often are lack of source resources to obtain the evidence, observational problems in collection (e.g. cloud obscured image) and failure to request the evidence. When missing data is not available in time, the fusion process needs to be robust enough to provide its best estimate without the data, and able to identify what data was missing and its effects on the

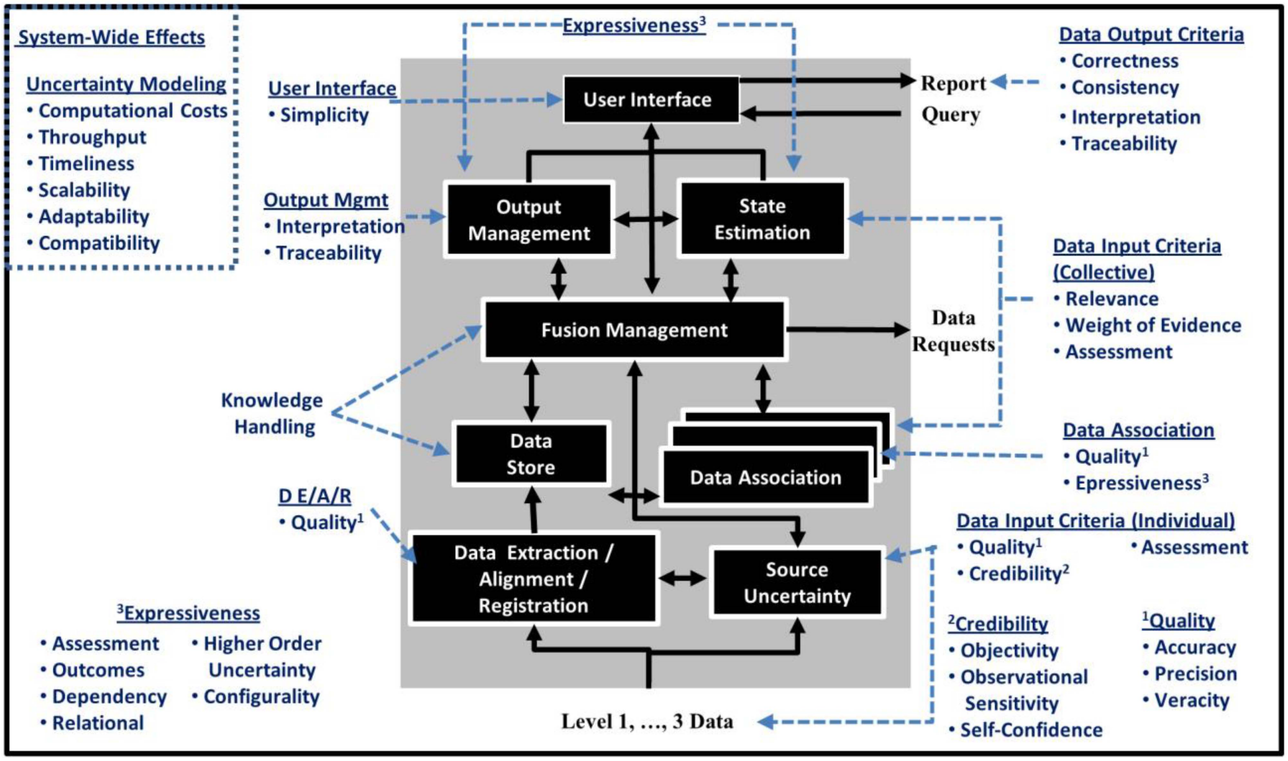


Fig. 10. Application of URREF evaluation criteria to different components of the generic fusion system model

output (see Traceability in section 4.4.2). Understanding how the system provides default values is important in these cases.

Inconsistency occurs when two or more inputs are contrary (they support different outputs in the fusion process) or contradictory. This is also called conflict and is a common issue in fusion systems. Conflict generally decreases the overall evidential force, as the conflicting items favor different outcomes. Conflict also increases uncertainty about source credibility and veracity, especially when one item of evidence favors an outcome significantly different that the remaining relevant evidence from different sources. Conflict has multiple causes, including non-source-initiated deception, source credibility/veracity issues, world segment model mismatches, and incomplete or uncertain model specification. Subject to available time, the desired approach is for the fusion process to alert the users to conflicts and allow them to conduct the necessary investigations to identify and resolve the root cause of the conflict. If resolution is not possible, then the system must be able to form a judgment based on the credibility of the evidence. Conflict can result in a significant amount of uncertainty that hinders decision making. Conflict modeling is usually addressed via probabilistic [59] or a belief function-type approach [62] [33] [36] [65].

Relevance, as an assessment of the force of an individual piece of true (from a credible and truth-telling source) evidence, is often dependent upon the related pieces of evidence. In many cases, evidence can be synergistic, either positively or negatively; its force is greater or lesser than its force when considered individually. Evidential relevance for additional like evidence tends to decrease if the multiple items provide limited or no additional new information. The synergy needs to be accounted for in the modeling.

Weight of Evidence (WOE) is an assessment of the totality of the available data and its effects on the output of the fusion. It is a holistic measure. It assesses the completeness of the evidence in supporting the fusion system output. It involves both the input evidence and the reasoning processes within the fusion system. There are multiple approaches to establishing the weight of evidence [77] [6]. Consider a physics analogy—weight is a function of the force of gravity and the mass of an object. Here, we will use effective force of evidence. This force results from the collective effects of Credibility and Quality on Relevance for each piece of evidence.

$$\text{WOE} = f(\text{Credibility, Quality, Relevance, Mass}) \quad (3)$$

The first three have already been discussed. Mass as used here is a measure of the comprehensiveness of the evidence—how many possible outputs are ruled out by the data. This makes Mass more than a simple count of how many items of evidence the system has. Rather, it focuses on the reasoning process in the fusion. A fusion process making a situation or impact assessment works as much by eliminating possible outputs as by supporting a specific output. Outputs that are neither positively or negatively supported remain as doubt in the system. WOE is also a useful tool in explaining the fusion system's output results.

4.4.2. Fusion Outputs Uncertainty Assessment Criteria

The Reasoning criteria of Correctness and Consistency are the core criteria for assessing uncertainty in the fusion system's output. How well the output mirrors the reality of the world segment it models is the primary measure of goodness of a fusion system. This makes output Correctness the central URREF criterion. However, this criterion is different than the Accuracy criterion for input information. Fusion system outputs normally come with an uncertainty hedge. Most often, this hedge is presented probabilistically—"There is a 90% chance this ship is the ship of interest." If a Correctness measures does not account for the probabilistic nature of the output, it will provide an incorrect view of the system's performance. Correctness can be assessed quantitatively using scoring rules [51], [29]. The original scoring rule is the Brier score. There are several versions; the most common applies to cases where the predicted outcome occurred or did not occur.

$$\mathrm{BS} = \frac{1}{N}\sum_{i=1}^{n}(f_i - a_i)^2 \tag{4}$$

Where N is the total number of outputs for which both a forecast probability (f_i) and an actual outcome (a_i) are available [10].

Closely following is the Consistency criterion. There are two considerations in this criterion:

- How repeatable are the results, when the same kind of evidence is provided?
- How sensitive is the output to minor changes in the input conditions?

Within the Brier score, there is a measure of the consistency of the forecasts. This assesses whether something predicted to be true 80% time actually occurs 80% of the time. It is also called reliability or calibration in the literature. It is

$$\frac{1}{N}\sum_{k=1}^{J} n_k(\mathbf{f}_k - \bar{\mathbf{o}}_k)^2 \tag{5}$$

Where N is the total number of outputs for which both a forecast probability (f_k) and an actual outcome are available, J is the number of forecast probabilities (assumed finite), n_k is the number of forecast probabilities in bin k, $\mathbf{f}_k$ is the forecast probability of bin k, and is the observed frequency of the outcomes predicted to occur in bin k. Both $\mathbf{f}_k$ and $\bar{\mathbf{o}}_k$ are vector quantities.

As with the Accuracy criterion, in those cases where there is no ground truth to establish a correct answer (including a simulated ground truth), the reasoning process can still be evaluated in terms of how its answers correspond to a gold standard (e.g. SMEs, documentation, etc.) [34].

In addition to providing the users with correct and consistent outputs, users benefit from understanding how and why the fusion system generated those outputs [74]. The data handling criteria of Interpretation and Traceability qualitatively assess this capability. Interpretation is the ability of a fusion system to support a coherent explanation of its conclusions. This is a summary explanation of the key evidence and reasoning process that supports the output. It is a justification for using the output in decision making. Interpretation can be assessed in at least two ways:

- Operationally via a user/stakeholder assessment that a representative range of output interpretations satisfy their information needs.
- Developmentally via fusion system experts' assessment that the interpretation captures the essential information input into or created by the system

Traceability is a diagnostic capability allowing users to follow the system's processes. It assesses the ability of a fusion system to provide an accurate and unbroken historical record of its inputs and the chain of operations that led to its conclusions. It is useful when the user wants an in-depth understanding of how the system came to its conclusions, or when the user suspects something is wrong or out of the ordinary in the output and its interpretation and wants to investigate further. Few fusion systems log intermediate results. But if the system records all inputs, including user requests, and the initial system states, and allows access to intermediate products during execution, system traceability can be conducted off-line. Traceability also applies to knowing exactly what evidence was used. Some sources occasionally find they need to retract evidence that turns out to be in error. Tracing what evidence items exist in one's data base supports this retraction process.

4.4.3. Effects of Fusion System Processing Uncertainty Assessment Criteria

In assessing the uncertainty representation within a fusion system, one must consider the overall ability of the system in reducing the total uncertainty on the reported outputs, the errors introduced by the fusion process, and the cost and fusion limitations imposed by the selected uncertainty representation approach. This is an area where significant work is required to fully understand where and how uncertainty is generated and propagated through the various fusion processes.

4.4.3.1. Uncertainty reduction and introduction of errors and uncertainty reduction

A fusion system is designed to reduce uncertainty by integrating the evidence, using one piece of evidence to reduce uncertainty in another. This requires (at least) conditional independence between the evidences. That is, the only dependencies between the evidence are mediated by the output whose uncertainty one wants to reduce. There are no other causes of correlation between the evidence items. With increasing efforts to increase the degree of L1/L2 fusion at source systems, such as the US Air Force's Distributed Common Ground System [75], It is important for fusion system designers to understand the possibility of multiple counting of common source evidence.

The fusion process can also introduce errors, which can increase the uncertainty in the output. Common errors are in the data extraction/alignment/registration process through incorrect classification/assignments, rounding, and misalignment [56] [20]. Information development processes can introduce errors through mis-association, misclassification, or unwarranted elimination of embedded uncertainty in the source evidence. The uncertainty representational scheme used plays a significant role in establishing the kinds of uncertainties that can be assessed in the information development process. For example, if the incoming data is heavily ambiguous, but the process has no mechanism for representing that ambiguity, the evidence output may be specified as being more definitive than the data warrants. Fusion reasoning elements need to account for possible accuracy, precision and veracity errors in extracted information [38]. For the reasoning processes, expressiveness of the chosen representations is an important consideration. These are:

- Assessment: Establishes what kinds of uncertainties can be addressed in the fusion system.
- Outcomes: Determines whether the outputs can incorporate the residuals of the types of uncertainties in the input data and created by the fusion process.
- Configurality: Determines the range over which a particular uncertainty representation needs to operate.
- Dependency: Determines whether the world segment model and source models incorporate all the dependencies necessary for the fusion model to correctly represent the uncertainties in the world segment and the sources.
- Higher order uncertainty: Determines if the uncertainty representation can include uncertainty about one's uncertainties. This is especially the case for uncertainty about probabilities that are used in reasoning models.

4.4.3.2. Effects on Fusion System Resources

In addition to assessing the range of needed uncertainty representation capabilities, there are a set of criteria to evaluate the effects of the uncertainty representation capabilities on the resource costs and range of capabilities for the fusion system. The first set of criteria identify the effects of different uncertainty representation approaches on the design of the fusion system. They are:

- Computational costs. Different representation schemes have varying demands on the fusion system's computational resources. Truth-functional approaches of possibilistic representations or probabilistic approaches that use canonical models [19] generally have the lowest cost, while random set approaches [50] have the highest. The computational cost will also depend on whether exact or approximate techniques are used, which have their own effects on output uncertainty.
- Performance (sub criteria—throughput and timeliness): Assesses the upper limit on system volume and velocity, determining if the selected uncertainty representation schemes significantly affect the ability of the system to meet the users' needs. These two sub-criteria are intertwined with the computational cost criteria.
- Scalability: Effect of the representation to scale the model used. This is of especial interest when the world segment model allows for a significantly varying number of entities with different relationships between them.

The second set look at the constraints the uncertainty representation models place on the use of the system:

- Adaptability: Degree of change allowed to the configuration of the uncertainty representation, allowing it to model variations in the world segment or source models.
- Compatibility: how well the representation allows the use of common data standards within the domain within which the fusion system works (e.g. STANAGS for NATO systems, NIST IT standards for US systems, etc.).
- Knowledge handling: The effect of a particular uncertainty representation on the fusion system's information management capabilities.
- Simplicity: the degree of complexity of the user interface, especially with regards to the system's output explanation capabilities.

The assessment results on the effects on system resources should be incorporated into a larger system performance analysis. This enables a proper trade-off analysis between resource demand and uncertainty handling representation with the context of the overall system requirements.

5. CONCLUSION

This paper provided a broad examination of how the URREF uncertainty handling criteria can be applied to typical HLIF applications. We ground the discussion with a Fusion Process Environment Model to identify where the criteria should be applied. The application of the Framework's criteria to the evaluation of the uncertainties and their representations in a fusion system is shown from different perspectives. As noted, as an uncertainty evaluation framework, URREF must be seen in its current state as a first output of an effort to better understand the representation and effects of uncertainties in a HLIF system. As HLIF technologies advance, understanding and correctly addressing uncertainties will play an important part. Based on the points raised in this paper, we forecast two major directions for this effort in the future. First, comprehensive quantitative and qualitative comparisons among different representation approaches are important to better understand the appropriate applicability of each approach and guide HLIF developers in their design decisions. As probabilistic, possibilistic, and evidential approaches evolve, they gain new capabilities and provide new insights that can be shared across approaches. Second, a deeper understanding of real-world fusion processes is required to select and apply the most appropriate fusion models and systems for each specific situation.

REFERENCES

[1] R. G. Almond
Graphical Belief Modeling
CRC Press, 1995.

[2] M. Bedworth and J. O'Brien
The Omnibus model: a new model of data fusion?
IEEE Aerospace and Electronic Systems Magazine, 15 (2000), 30–36.

[3] P. L. Bernstein
Against the Gods: The Remarkable Story of Risk.
New York: John Wiley & Sons, 1985.

[4] M. Black
Vagueness: An Exercise in Logical Analysis.
Philosophy of Science, 4 (1937), 427–455.

[5] E. Blasch, I. Kadar, J. Salerno, M. M. Kokar, S. Das, G. M. Powell, D. D. Corkill, and E. H. Ruspini
Issues and challenges of knowledge representation and reasoning methods in situation assessment (Level 2 Fusion).
in Proceedings: Signal Processing, Sensor Fusion, and Target Recognition XV, 6235 (2006).

[6] E. Blasch, K. B. Laskey, A. L. Jousselme, V. Dragos, P. C. Costa, and J. Dezert
URREF reliability versus credibility in information fusion (STANAG 2511).
in Proceedings of the 16th International Conference On Information Fusion, 2013, 1600–1607.

[7] E. P. Blasch and S. Plano
Level 5: user refinement to aid the fusion process.
in Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications. International Society for Optics and Photonics, 2003, 288–298.

[8] H. Boström, S. F. Andler, M. Brohede, R. Johansson, A. Karlsson, J. van Laere, L. Niklasson, M. Nilsson, A, Persson, and T. Ziemke
On the Definition of Information Fusion as a Field of Research.
Report No. HS-IKI-TR-07-006, University of Skovde, School of Humanities and Informatics, Skovde, Sweden, 2007.

[9] R. J. Brachman, and H. J. Levesque
Knowledge Representation and Reasoning.
Amsterdam: Morgan Kaufmann, 2004.

[10] G. Brier
Verification of forecasts expressed in terms of probability.
Monthly Weather Review, 28 (1950), 1–3.

[11] H. Carvalho, W. Heinzelman, A. Murphy, and C. Coelho
A general data fusion architecture.
in Proceedings of the 6th International Conference on Information Fusion, 2003, 1465–1472.

[12] F. Castanedo
A review of data fusion techniques.
The Scientific World Journal, 2013.

[13] P. C. Costa, K. B. Laskey, E. Blasch, and A. L. Jousselme
Towards unbiased evaluation of uncertainty reasoning: the URREF ontology.
in Proceedings of the 15th International Conference on Information Fusion, 2012, 2301–8.

[14] P. C. G. Costa and D. M. Buede
Dynamic decision making: a comparison of approaches.
Journal of Multi-Criteria Decision Analysis, 9 (2000), 243–262.

[15] S. K. Das
High-Level Data Fusion.
Artech House, 2008.

[16] B. V. Dasarathy
Sensor fusion potential exploitation-innovative architectures and illustrative applications.
Proceedings of the IEEE, 85 (1997), 24–38.

[17] J. P. de Villiers, K. Laskey, A. L. Jousselme, E. Blasch, A. de Waal, G. Pavlin, and P. Costa
Uncertainty representation, quantification and evaluation for data and information fusion.
in Proceedings of the 18th International Conference on Information Fusion, 2015, 50–57.

[18] J. Dezert and F. Smarandache
Presentation of DSmT.
in J. Dezert and F. Smarandache (Eds.). Advances and Applications of DSmT for Information Fusion. American Research Press, 2004, 3–35.

[19] F. J. Díez and M. J. Druzdzel
Canonical probabilistic models for knowledge engineering.
Report No. CISIAD-06-01, Madrid, Spain: Universidad Nacional de Educación a Distancia, 2006.

[20] X. L. Dong, E. Gabrilovich, G. Heitz, W. Horn, K. Murphy, S. Sun, and W. Zhang
From data fusion to knowledge fusion.
Proceedings of the Very Large Database Endowment, 7 (2014), 881–892.

[21] E. Elles and B. Fitelson
Measuring confirmation and evidence.
The Journal of Philosophy, 97 (2000), 663–672.

[22] M. R. Endsley
Toward a Theory of Situation Awareness in Dynamic Systems.
Human Factors, 37 (1995), 32–64.

[23] M. R. Endsley
Theoretical underpinnings of situation awareness: A critical review.
in M. R. Endsley and D. J. Garland (Eds.) Situation Awareness Analysis and Measurement. Mahwah, NJ: Lawrence Earlbaum Associates, 2000, pp. 3–32.

[24] J. Esteban, A. Starr, R. Willetts, P. Hannah, and P. Bryanston-Cross
A review of data fusion models and architectures: towards engineering guidelines.
Neural Computing & Applications, 14 (2005), 273–281.

[25] Evaluation of Technologies for Uncertainty Reasoning Working Group (ETURWG)
Homepage: http://eturwg.c4i.gmu.edu.

[26] P. H. Foo and G. W. Ng
High-level information fusion: An overview.
Journel of Advances in Information Fusion, 8 (2013), 33–72.

[27] J. García, L. Snidaro, and J. Llinas
Architectural aspects for context exploitation in information fusion.
in Context-Enhanced Information Fusion. Springer, 2016, 185–203.

[28] L. Getoor, N. Friedman, D. Koller, A. Pfeffer, and B. Taskar
Probabilistic relational models.
in L. Getoor and B. Taskar (Eds.) Introduction to Relational Statistical Models. The MIT Press, 2007.

[29] T. Gneiting and A. E. Raftery
Strictly proper scoring rules, prediction, and estimation.
Journal of the American Statistical Association, 102 (2007), 359–378.

[30] D. L. Hall and J. Llinas
An introduction to multisensor data fusion.
Proceedings of the IEEE, 85 (1997), 6–23.

[31] D. L. Hall and J. Llinas
Multi-sensor data fusion.
in M. E. Liggins, D. L. Hall and J. Llinas (Eds.) Handbook of Multisensor Data Fusion: Theory and Practice. Boca Raton, FL: CRC Press, 2009, 1–14.

[32] E. K. Jones, N. Denis, and D. Hunter
Hypothesis management for information fusion.
in Proceedings of the Fifth International Conference On Information Fusion, 2002, 894–899.

[33] A. Jøsang
Subjective Logic.
Springer, 2016.

[34] A. Jøsang, P. C. Costa, and E. Blasch
Determining model correctness for situations of belief fusion.
in Proceedings of the 16th International Conference On Information Fusion, 2013, 1886–1893.

[35] A. Jousselme, P. Maupin, E. Bosse, and D. Lambert
Knowledge, belief and uncertainty.
in E. Bossé, J. Roy, and S. Wark (Eds.) Concepts, Models, and Tools for Information Fusion. Artech House, 2007, 141–150.

[36] A. L. Jousselme
Semantic criteria for the assessment of uncertainty handling fusion models.
in Proceedings of the 19th International Conference On Information Fusion, 2016, 488–495.

[37] A. L. Jousselme, A. C. Boury-Brisset, B. Debaque, and D. Prevost
Characterization of hard and soft sources of information: a practical illustration.
in Proceedings of the 17th International Conference On Information Fusion, 2014, 1–8.

[38] A. L. Jousselme and G. Pallotta
Dissecting uncertainty-based fusion techniques for maritime anomaly detection.
in Proceedings of the 18th International Conference On Information Fusion, 2015, 34–41.

[39] J. A. Karakowski
Towards Visual Data Fusion.
in Proceedings of the National Symposium of Sensor and Data Fusion, 1998, Lexington, MA, USA.

[40] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi
Multisensor data fusion: A review of the state-of-the-art.
Information Fusion, 14 (2013), 28–44.

[41] F. H. Knight
Risk, Uncertainty, and Profit.
Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Co., 1921.

[42] D. Koller, A. Levy, and A. Pfeffer
P-CLASSIC: a tractable probabilistic description logic.
Proceedings of the Fourteenth National Conference on Artificial Intelligence and Ninth Conference on Innovative Applications of Artificial Intelligence, 1997, 390–397.

[43] P. Krause and D. Clark
Representing Uncertain Knowledge: An Artificial Intelligence Approach.
Kluwer Academic Publishers, Dordrecht, Boston, 1993.

[44] D. A. Lambert
Grand challenges of information fusion.
in Proceedings of the Sixth International Conference on Information Fusion, 2003, 213–220.

[45] D. A. Lambert
The state transition data fusion model.
in E. Blasch, E. Bossé, and D. A. Lambert. (Eds.) High-Level Information Fusion Management and Systems Design. Artech House, 2012, 33–79.

[46] K. B. Laskey, P. C. Costa, and T. Janssen
Probabilistic ontologies for knowledge fusion.
in Proceedings of the 11th International Conference On Information Fusion, 2008, 1–8.

[47] K. J. Laskey, K. B. Laskey, P. C. Costa, M. M. Kokar, T. Martin, and T. Lukasiewicz
Uncertainty Reasoning for the World Wide Web.
W3C Incubator Group Report 31, 2008.

[48] D. Lee (Ed.)
Law of Armed Conflict Deskbook, 5th ed.
The Judge Advocate General's Legal Center and School, U.S. Army, Charlottesville, Virginia, USA, 2015.

[49] T. C. Ly, S. Greenhill, S. Venkatesh, and A. Pearce
Multiple hypotheses situation assessment.
in Proceedings of the Sixth International Conference of Information Fusion, 2003, 972–978.

[50] R. P. Mahler
Statistical Multisource-Multitarget Information Fusion.
Artech House, Inc, 2007.

[51] J. E. Matheson and R. L. Winkler
Scoring rules for continuous probability distributions.
Management Science 22 (1976), 1087–1096.

[52] B. Milch, B. Marthi, S. Russell, D. Sontag, D. L. Ong, and A. Kolobov
BLOG: Probabilistic Models with Unknown Objects.
in L. Getoor and B. Taskar (Eds.) Statistical Relational Learning. The MIT Press, 2007, 373–398.

[53] Opinion
Oxford English Dictionary (Online version).
https://en.oxforddictionaries.com/definition/opinion.

[54] H. Pasula, B. Marthi, B. Milch, S. Russell, and I. Shpitser
Identity uncertainty and citation matching.
in Advances in Neural Information Processing Systems, 2002, 1401–1408.

[55] L. F. Pau
Sensor data fusion.
Journal of Intelligent and Robotic Systems, 1 (1988), 103–116.

[56] J. Piskorski, R. Yangarber
Information extraction: past, present and future.
in Multi-Source, Multilingual Information Extraction and Summarization. Springer, 2013, 23–49.

[57] J. Roy, D. Lambert, F. White, J. Salerno, J. Karakowski, and M. Hinman
Data- and Information-fusion models.
in E. Bosse, J. Roy and S. Wark, (Eds.) Concepts, Models, and Tools for Information Fusion. Artech House, Boston, 2007, 69–118.

[58] J. Salerno
Information fusion: a high-level architecture overview.
in Proceedings of the Fifth International Conference On Information Fusion, 2002, 680–686.

[59] D. A. Schum
The Evidential Foundations of Probabilistic Reasoning.
Northwestern University Press, 1994.

[60] D. A. Schum
Evidence marshaling for imaginative fact investigation.
Artificial Intelligence and Law, 9 (2001), 165–188.

[61] D. A. Schum
A science of evidence: contributions from law and probability.
Law, Probability & Risk, 8 (2009), 197–231.

[62] G. Shafer
A Mathematical Theory of Evidence.
Princeton University Press, 1976.

[63] E. Shahbazian, D. E. Blodgett, and P. Labbé
The extended OODA model for data fusion systems.
Proceedings of the Fourth International Conference On Information Fusion, 2001.

[64] P. P. Shenoy
Valuation-based systems for Bayesian decision analysis.
Operations Research, 40 (1992), 463–484.

[65] F. Smarandache and J. Dezert
Advances and Applications of DSmT for Information Fusion, Vol. IV: Collected Works.
Infinite Study, 2015.

[66] P. Smets
The combination of evidence in the transferable belief model.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 12 (1990), 447–458.

[67] A. Smirnov, T. Levashova, and N. Shilov
Patterns for context-based knowledge fusion in decision support systems.
Information Fusion, 21 (2015), 114–129.

[68] M. Smithson
Ignorance and Uncertainty: Emerging Paradigms.
New York: Cognitive Science. Springer-Verlag, 1989.

[69] L. Snidaro, L. Vati, J. García, E. D. Martí, A. L. Jousselme, K. Bryan, D. D. Bloisi, and D. Nardi
A framework for dynamic context exploitation.
in Proceedings of the 18th International Conference On Information Fusion, 2015, 1160–1167.

[70] J. F. Sowa
Knowledge Representation: Logical, Philosophical, and Computational Foundations.
Pacific Grove: Brooks/Cole, 2000.

[71] A. N. Steinberg
Foundations of situation and threat assessment.
in M. E. Liggins, D. L. Hall, and J. Llinas (Eds.) Handbook of Multisensor Data Fusion: Theory and Practice. Boca Raton, FL: CRC Press, 2009, 691–699.

[72] A. N. Steinberg, C. L. Bowman, and F. E. White
Revisions to the JDL data fusion model.
in Sensor Fusion: Architectures, Algorithms, and Applications III. International Society for Optics and Photonics, 1999, 430–442.

[73] J. Steinhauer, A. Karlsson, and S. F. Andler
Traceable uncertainty.
in Proceedings of the 16th International Conference On Information Fusion, 2013, 1582–1589.

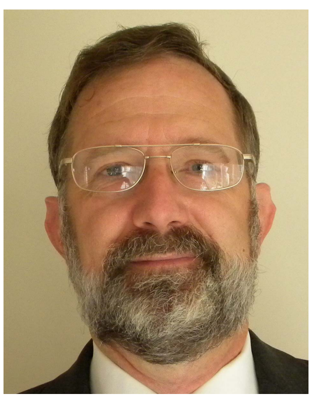
[74] S. C. Thomopoulos
Sensor integration and data fusion.
Journal of Robotic Systems, 7 (1990), 337–372.

[75] U.S. Air Force
Air Force Distributed Common Ground System.
Website: http://www.af.mil/About-Us/Fact-Sheets/Display/Article/104525/air-force-distributed-common-ground-system.

[76] U.S. Congress
Rule 401.
Federal Rules of Evidence, 1975.

[77] D. L. Weed
Weight of evidence: a review of concept and methods.
Risk Analysis, 25 (2005), 1545–1557.

[78] P. D. Whitney, C. Posse, and X. C. Lei
Toward a Unified Approach to Information Integration: a Review Paper on Data/Information Fusion.
Report No. PNNL-15428, Pacific Northwest National Laboratory, Richland, WA, USA, 2005.



Mark Locher is currently a PhD candidate in System Engineering at George Mason University, Fairfax Virginia. He received a BS in Operations Research and Aviation Science in 1979 from the US Air Force Academy, an MBA from St Mary's University, San Antonio TX, in 1982, and an MS in System Engineering from George Mason University in 2012. He completed a 20-year career with the US Air Force and then 15 years as an industry consultant to the US Department of Defense. His primary focus was developing and integrating Intelligence Surveillance Reconnaissance systems and platforms in multiple domains. His research interests include uncertain reasoning techniques, knowledge representation and logical reasoning, and decision problem aiding. He is a member of the International Council of Systems Engineering.

Paulo Costa is an Associate Professor of Systems Engineering and Operations Research at George Mason University. His teaching and research interests comprise the areas of probabilistic ontologies, multi-sensor information fusion, Bayesian reasoning, predictive analysis, cyber-security and Decision Theory. He is a former fighter pilot with extensive experience in tactical and operational planning, and an expert in requirements engineering for complex systems, such as intelligent transportation and health-care support systems. Dr. Costa is an IEEE senior member, Member of the Board of Directors of the International Society for Information fusion, and member of the International Council of Systems Engineering.