How to act on inconsistent news: Ignore, resolve, or reject

Anthony Hunter

Department of Computer Science
University College London
Gower Street
London WC1E 6BT, UK

Abstract

Inconsistencies frequently occur in news about the real-world. Some of these inconsistencies may be more significant than others, and some news may contain more inconsistencies than others. This creates problems of deciding whether to act on these inconsistencies, and if so how. Possible actions on an inconsistency in a news report include ignore the inconsistency, resolve the inconsistency, and reject the report. To support this, we extend and apply a general characterization of inconsistency, based on Belnap's four-valued logic. For conflicts arising between the news and background knowledge, we analyse coherence and significance of the corresponding the four-valued models for that knowledge and show how this analysis can indicate an appropriate course of action.

Key words: Analysing inconsistency; Measuring inconsistency; Contradiction; Inconsistency tolerance; Information integration; Knowledge fusion.

1 Introduction

News reports are often inconsistent. Some tolerance of inconsistency is therefore normally necessary. However, it is not always clear how much tolerance should be excercised. Sometimes, it is necessary to reject news reports. Sometimes, it is necessary to resolve some of the inconsistencies in the news report. At other times, news reports are accepted with inconsistencies, but treated with caution. Having some understanding of the "degree of inconsistency" of a news report can help in

Email address: a.hunter@cs.ucl.ac.uk (Anthony Hunter).
URL: www.cs.ucl.ac.uk/staff/a.hunter (Anthony Hunter).

deciding how to act on it. To do this systematically calls for a formal framework for considering inconsistency.

In order to analyse and reason with information in news, we assume the news reports are provided in the form of structured news reports. These are XML documents, where the text entries are restricted to individual words or simple phrases, such as names and domain-specific terminology, and numbers and units. It is assumed structured news reports do not require natural language processing. Though we do assume structured news reports can be obtained as output from information extraction technology [1]. Much information can be represented in the form of structured news reports, including weather reports, business reports, summaries of scientific articles, results from clincal trials, market intelligence, operations reports within organisations, etc. We are therefore taking a very general view on news to include any kind of regular supply of information that conforms to some stereotypical format.

Each structured news report is isomorphic to a logical term. So structured news reports can be used, in conjunction with some domain knowledge, for logical reasoning as part of some software tool. Intelligent tasks based on structured news reports include: (1) forming a merged structured news report from a set of heterogeneously sourced structured news reports [2,3]; (2) looking for interesting news by looking for violations of expectations [4,5]; and (3) constructing arguments for and against conclusions based on conflicting news reports [6]. Central to all these intelligent tasks is the need to address inconsistencies that arise.

An important aspect of analysing an inconsistency is evaluating its significance. As an illustration of the need to evaluate significance, consider two news reports on a World Cup match, where the first report says that Brazil beat Germany 2-0, and the second report says that Germany beat Brazil 2-0. This is clearly a significant inconsistency. Now consider two news reports on the same football match, where the first report says that the referee was Pierluigi Collina and the second report says that the referee was Ubaldo Aquino. This inconsistency would normally be regarded as relatively insignificant.

Moreover, inconsistencies between information in a news report and domain knowledge can tell us important things about the news report. For this we use a significance function to give a value for each possible inconsistency that can arise in a news report in a given domain. We may then use this in various ways: (1) to reject reports that are too inconsistent; (2) to highlight unexpected news; (3) to focus on repairing significant inconsistencies; and (4) to monitor sources of information to identify sources that are unreliable.

In this paper, to analyse the significance of inconsistencies in structured news reports, we will adapt and extend an approach to measuring inconsistency based on four-valued models [7]. In this adaptation, we use Belnap's logic for finding the models for a set of inconsistent formulae. We then extend this framework to support the analysis of heterogeneously sourced information such as news from different newsfeeds.

2 Structured news reports

Syntactically, a structured news report is a data structure containing a number of grammatically simple phrases together with a tag (giving semantic information) for each phrase. As an illustration, see Example 1. Each phrase that is tagged is a textentry. The set of tags in a structured news report is meant to parameterize a stereotypical situation, and so a particular structured news report is an instance of that stereotypical situation. For example, news reports on corporate acquisitions can be represented as structured news reports using tags including `buyer`, `seller`, `acquisition`, `value`, and `date`. Each phrase in structured news report is very simple, such as a proper noun, a date, or a number with unit of measure, or a word or phrase from a prescribed lexicon. For an application, the prescribed lexicon delineates the types of states, actions, and attributes, that could be conveyed by the structured news report.

The definition for a structured news report is very general. In practice, we would expect a DTD for a given domain. So for example, we would expect that for an implemented system that merges weather reports, there would be a corresponding DTD. One of the roles of a DTD, say for weather reports, would be to specify the minimum constellation of tags that would be expected of a weather report. We may also expect integrity constraints represented in classical logic to further restrict appropriate structured news reports for a domain.

Each structured news report is isomorphic with a ground atom of classical logic where each tagname is a function or predicate symbol and each textentry is a constant symbol: The root tagname is the predicate symbol for the atom, and the other tagnames are function symbols for the atom. This isomorphism is illustrated in the following example.

Example 1 *Consider the following structured news report.*

```
⟨report⟩
    ⟨log⟩⟨station⟩Inverness⟨/station⟩⟨date⟩12/2/03⟨/date⟩⟨/log⟩
    ⟨rainfall⟩2.3cm⟨/rainfall⟩
⟨/report⟩
```

This can be represented by the following atom.

$$\texttt{report}(\texttt{log}(\texttt{station}(\texttt{Inverness}), \texttt{date}(12/3/02)), \texttt{rainfall}(2.3\texttt{cm}))$$

Representing each structured news report by a logical atom means that they can then be directly used with other information in a knowledgebase via logical inference. To show this we provide a few simple definitions. From now on, a **report** will be taken to mean an atom that represents a structured news report.

Definition 1 *A* **grounding** *is an equality predicate where the first argument is a variable and the second argument is a ground term. A* **grounding set** *is a set of groundings which can be substituted into an unground term to give a grounded term. Let α be a formula and let Φ be a grounding set.* $\mathsf{Ground}(\alpha, \Phi)$ *gives the result of substituting each variable x in α with term t where the grounding $x = t$ is in Φ.*

Example 2 *Let* $\texttt{a}(\texttt{b}(\texttt{x}), \texttt{c}(\texttt{y}), \texttt{d}(\texttt{z}))$ *be a formula where* $\texttt{x}$, $\texttt{y}$ *and* $\texttt{z}$ *are variables. Let the grounding set Φ be* $\{\texttt{y} = \texttt{john}, \texttt{z} = \texttt{betty}\}$.

$$\mathsf{Ground}(\texttt{a}(\texttt{b}(\texttt{x}), \texttt{c}(\texttt{y}), \texttt{d}(\texttt{z})), \Phi) = \texttt{a}(\texttt{b}(\texttt{x}), \texttt{c}(\texttt{john}), \texttt{d}(\texttt{betty}))$$

We use access rules, defined next, to reason with information in reports.

Definition 2 *An* **access rule** *is a first order formula of the form where: (1) $x_1, ..., x_k$ are the variables in α; (2) the variables in $\beta_1, ..., \beta_n$ are a subset of, or equal to, $x_1, .., x_k$; (3) all of $\alpha, \beta_1, ..., \beta_n$ are unquantified; and (4) the predicate symbol for α is not used as a predicate symbol for any of $\beta_1, .., \beta_n$.*

$$\forall x_1, ..., x_k; \alpha \rightarrow \beta_1 \wedge ..., \beta_n$$

If there is a grounding set Φ s.t. $\mathsf{Ground}(\alpha, \Phi)$ *is a report, then* $\mathsf{Ground}(\beta_1, \Phi)$*, and ..., and* $\mathsf{Ground}(\beta_n, \Phi)$ *are atoms that can be inferred from the report.*

In the definition of an access rule, because α is ground by a report, α is a monadic predicate (i.e. it has exactly one argument). However, the $\beta_1, .., \beta_n$ predicates are of arbitrary arity. They do not need to be restricted to monadic predicates.

Definition 3 *Let Δ be a set of access rules and let ρ be a report. Let* $\mathsf{Access}(\Delta, \rho)$ *be the set of atoms obtained by exhaustively applying ρ to the access rules in Δ*

using generalized modus ponens and conjunction elimination as defined below.

$$\mathsf{Access}(\Delta, \rho) = \{\mathsf{Ground}(\beta_1, \Phi), .., \mathsf{Ground}(\beta_n, \Phi) \mid \forall x_1, .., x_k; \alpha \rightarrow \beta_1 \wedge .. \wedge \beta_n \in \Delta \text{ and } \mathsf{Ground}(\alpha, \Phi) \text{ is the atom } \rho\}$$

Example 3 *Let ρ be the following report.*

$$\texttt{report}(\texttt{forecast}(\texttt{date}(\texttt{25July01}), \texttt{city}(\texttt{Malaga}), \texttt{today}(\texttt{blizzard})))$$

Suppose the set of access rules, Δ, contains the following.

$$\forall \texttt{x}, \texttt{y}, \texttt{z};\ \texttt{report}(\texttt{forecast}(\texttt{date}(\texttt{x}), \texttt{city}(\texttt{y}), \texttt{today}(\texttt{z}))) \rightarrow \texttt{date}(\texttt{x}) \wedge \texttt{location}(\texttt{y}) \wedge \texttt{today}(\texttt{z})$$

So for ρ we get $\texttt{date}(\texttt{25July2001}), \texttt{location}(\texttt{Malaga}), \texttt{today}(\texttt{blizzard})$ *is a member of* $\mathsf{Access}(\Delta, \rho)$*. Suppose Δ also contains the following access rule.*

$$\forall \texttt{x}, \texttt{y};\ \texttt{report}(\texttt{forecast}(\texttt{date}(\texttt{x}), \texttt{city}(\texttt{y}), \texttt{today}(\texttt{blizzard}))) \rightarrow \texttt{blizzardWarning}(\texttt{x}, \texttt{y})$$

So we also get $\texttt{blizzardWarning}(\texttt{25July2001}, \texttt{Malaga}) \in \mathsf{Access}(\Delta, \rho)$

We assume that textentries in structured news reports are heterogeneous in format. For example, the format of date values is unconstrained (12/12/1974; 31st Dec 96; 12 Nov 2001 etc.) as is the format of numbers and currency values (3 million; 3, 000, 000 GBP; $4, ¥500K etc.). Various kinds of heterogeneity in ontology can also arise, such as synonyms, hypernyms, etc, can be used. Elsewhere, we have discussed how this heterogeneity can be handled in logic by various kinds of equivalence axioms that can be used in conjunction with the access rules [2,8,3].

3 Four-valued models

We now start to consider how we can analyse inconsistent information. Our starting point is the analysis of four-valued models. A four-valued model provides a natural characterization of inconsistent information.

Definition 4 *A* **four-valued** *model, denoted X^D, is such that $X^D \subseteq \{+\alpha \mid \alpha \in D\} \cup \{-\alpha \mid \alpha \in D\}$ where D is the* **domain** *of X^D. An element of a four-valued* *model is called an* **object**. *A* **positive object** *is an object with a + prefix and a* **negative object** *is an object with a − prefix. A four-valued model may contain both* $+\alpha$ *and* $-\alpha$ *for some* α*. If a four-valued model* X^D *contains both* $+\alpha$ *and* $-\alpha$ *for some* α*, then we use the shorthand* $\pm\alpha$ *for these objects, and hence* $\pm\alpha \in X^D$ *iff* $+\alpha \in X^D$ *and* $-\alpha \in X^D$*. Let* Λ *denote the set of four-valued models.*

We can view four-valued models in terms of four truth values "true", "false", "both" or "neither", which we denote by the symbols T, F, B, and N, respectively. Furthermore, we can relate these truth values to a knowledge lattice (see Figure 1 (left)). As more "information" is obtained about a formula, the truth-value "increases", going upwards in the lattice. In other words, if we know nothing about a formula, it is N. Then as we gain some information, it becomes either T or F. Finally, if we gain too much information it becomes B. The knowledge lattice gives the following ordering relation over truth values which we use to formalize the notion of a truth assignment for a four-valued model.

Definition 5 *Let* $\geq_k$ *be an ordering relation, called the* **knowledge ordering**, *such that* $B \geq_k B$, $B \geq_k T$, $B \geq_k F$, $B \geq_k N$, $T \geq_k T$, $T \geq_k N$, $F \geq_k F$, $F \geq_k N$, *and* $N \geq_k N$.

Definition 6 *Let* X^D *be a four-valued model. A* **truth assignment** *for* X^D*, denoted* t_{X^D}*, is an assignment* $t_{X^D} : D \mapsto \{B, T, F, N\}$*, such that for all* $\alpha \in D$*,*

$$t_{X^D}(\alpha) \geq_k T \textit{ iff } +\alpha \in X^D$$

$$t_{X^D}(\alpha) \geq_k F \textit{ iff } -\alpha \in X^D$$

Hence, we get the following equivalences between truth assignments and four-valued models.

$$t_{X^D}(\alpha) = B \text{ iff } +\alpha \in X^D \text{ and } -\alpha \in X^D$$

$$t_{X^D}(\alpha) = T \text{ iff } +\alpha \in X^D \text{ and } -\alpha \notin X^D$$

$$t_{X^D}(\alpha) = F \text{ iff } +\alpha \notin X^D \text{ and } -\alpha \in X^D$$

$$t_{X^D}(\alpha) = N \text{ iff } +\alpha \notin X^D \text{ and } -\alpha \notin X^D$$

Normally, four-valued models are presented directly in terms of a truth assignment. They were first presented as such by Belnap [9]. We have presented four-valued models using objects because they make it easier for us to present our definitions in subsequent sections.

Example 4 *Let* X^D *be a four-valued model where* $D = \{\alpha, \beta, \gamma, \delta\}$*, and* $X^D = \{+\alpha, \pm\beta, -\delta\}$*. So* t_{X^D} *is such that* $t_{X^D}(\alpha) = T$*,* $t_{X^D}(\beta) = B$*,* $t_{X^D}(\gamma) = N$*, and* $t_{X^D}(\delta) = F$.

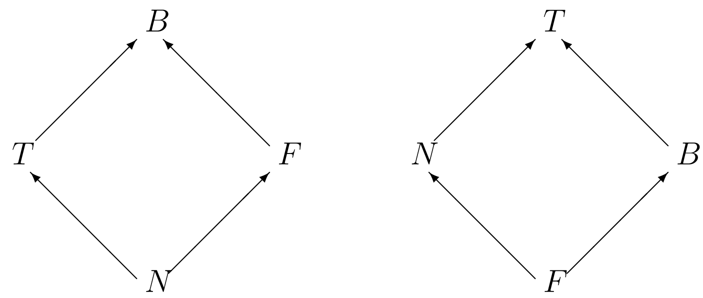

Fig. 1. The knowledge lattice (left) and the logical lattice (right)

The opinionbase of a four-valued model X^D is the set of atomic beliefs (atoms) for which there are reasons for the atom or its negation in X^D, and the conflictbase of X^D is the set of atomic beliefs with reasons for the atom and its negation in X^D.

Definition 7 *Let X^D be a four-valued model.*

$$\mathsf{Conflictbase}(X^D) = \{\alpha \in D \mid t_{X^D}(\alpha) = B\}$$
$$\mathsf{Opinionbase}(X^D) = \{\alpha \in D \mid t_{X^D}(\alpha) \neq N\}$$

If $\mathsf{Opinionbase}(X^D) = \emptyset$, then X^D has no arguments for/against any beliefs, and hence X^D has no opinions. If $\mathsf{Opinionbase}(X^D) = D$, then X^D is totally opinionated. If $\mathsf{Conflictbase}(X^D) = \emptyset$, then X^D is a conflictfree four-valued model. If $\mathsf{Opinionbase}(X^D) = D$, and $\mathsf{Conflictbase}(X^D) = \emptyset$, then we describe X^D as omniscient.

We now consider measures of inconsistency called coherence and concordance. Increasing the size of the conflictbase, with respect to the size of the opinionbase, decreases the degree of coherence, as defined below. Similarly, increasing the size of the conflictbase, with respect to the size of the domain, decreases the degree of concordance.

Definition 8 *The* Coherence *function from $\wp(\Lambda)$ into $[0,1]$, is defined below when* $\mathsf{Opinionbase}(X^D) \neq \emptyset$, *and* $\mathsf{Coherence}(X^D) = 1$ *when* $\mathsf{Opinionbase}(X^D) = \emptyset$,

$$\mathsf{Coherence}(X^D) = 1 - \frac{|\mathsf{Conflictbase}(X^D)|}{|\mathsf{Opinionbase}(X^D)|}$$

If $\mathsf{Coherence}(X^D) = 1$, then X^D is a totally coherent, and if $\mathsf{Coherence}(X^D) = 0$, then X^D is totally incoherent, otherwise, X^D is partially coherent (or equivalently partially incoherent).

Definition 9 *The* Concordance *function from $\wp(\Lambda)$ into $[0,1]$, is defined below when $D \neq \emptyset$,* *and* $\mathsf{Concordance}(X^D) = 1$ *when* $D = \emptyset$,

$$\mathsf{Concordance}(X^D) = 1 - \frac{|\mathsf{Conflictbase}(X^D)|}{|D|}$$

If $\mathsf{Concordance}(X^D) = 1$, then X^D is a totally concordant, and if $\mathsf{Concordance}(X^D)$ = 0, then X^D is totally discordant, otherwise, X^D is partially concordant (or equivalently partially discordant).

Example 5 *Let X^D be a four-valued model such that $D = \{\alpha, \beta, \gamma\}$, and $t(\alpha) = T$, $t(\beta) = B$, and $t(\gamma) = F$. Hence,* $\mathsf{Coherence}(X^D) = 2/3$ *and* $\mathsf{Concordance}(X^D)$ = $2/3$.

Example 6 *Let X^D be a four-valued such that $D = \{\alpha, \beta, \gamma, \delta\}$, and $t(\alpha) = T$, $t(\beta) = B$, $t(\gamma) = F$, and $t(\delta) = N$. Hence,* $\mathsf{Coherence}(X^D) = 2/3$ *and* $\mathsf{Concordance}(X^D) = 3/4$.

Both concordance and coherence are potentially useful measures in applications. They provide different perspectives on a model. Obviously, the more of the domain covered by the opinionbase, the closer these measures become.

Proposition 1 *For all $X^D \in \Lambda$,* $\mathsf{Concordance}(X^D) \geq \mathsf{Coherence}(X^D)$.

So far, we have assumed we have models to analyse. In the next section, we consider a logic for identifying models that satisfy a set of formulae.

4 Analysing inconsistent knowledgebases

To extend concordance and coherence measures for formulae, and hence for inconsistent knowledgebases, we need a satisfaction relation for a set of formulae (i.e. a knowledgebase). For this, we use Belnap's logic which provides a natural and intuitive form of paraconsistent reasoning [9,10].

Definition 10 *The* **language** *L for the logics that we will consider in this paper is the usual classical language that can be formed from a set of atoms and the set of logical symbols $\{\neg, \vee, \wedge\}$. We call any subset of L a* **knowledgebase**. *If $\Gamma \in \wp(L)$, then* $\mathsf{Atoms}(\Gamma)$ *returns the set of atom symbols used in Γ.*

The truth assignment (Definition 6) is extended to any formula in L by induction on the structure of the formula using a distributive lattice called the logical lattice (see Figure 1 (right)). We also assume an involution operator $\bullet$ satisfying the conditions (1) $\alpha = \alpha^{\bullet\bullet}$, and (2) if $\alpha \geq \beta$ then $\beta^\bullet \geq_l \alpha^\bullet$, where $\geq_l$ is the ordering relation for the logical lattice. The truth assignment for formulae observes monotonicity and

α	N	F	T	B
$\neg\alpha$	N	T	F	B

Table 1
Truth table for negation for Belnap semantics

$\wedge$	N	F	T	B
N	N	F	N	F
F	F	F	F	F
T	N	F	T	B
B	F	F	B	B

Table 2
Truth table for conjunction for Belnap semantics

$\vee$	N	F	T	B
N	N	N	T	T
F	N	F	T	B
T	T	T	T	T
B	T	B	T	B

Table 3
Truth table for disjunction for Belnap semantics

complementation, in the logical lattice, so $x \wedge y$ is the meet of $\{x, y\}$ and $x \vee y$ is the join of $\{x, y\}$. The **truth assignment for formulae** is summarized in the truth tables (Tables 1 – 3) for the $\neg$, $\wedge$, and $\vee$ connectives.

Example 7 *Continuing Example 4, we have assignments using t_{X^D} including the following, where $\alpha \wedge \neg\beta \wedge \beta, \alpha \vee \gamma, \beta \vee \gamma, \gamma \vee \neg\gamma \in L$.*

$$t_{X^D}(\alpha \wedge \neg\beta \wedge \beta) = B \quad t_{X^D}(\beta \vee \gamma) = T$$
$$t_{X^D}(\alpha \vee \gamma) = T \qquad t_{X^D}(\gamma \vee \neg\gamma) = N$$

There is no $\alpha \in L$ such that the semantic assignment function always assigns the value T. However, there are formulae that never take the value F, for example $\alpha \vee \neg\alpha$. Also, the set of formulae that never take the value F is not closed under conjunction. For example, consider $(\alpha \vee \neg\alpha) \wedge (\beta \vee \neg\beta)$ when α is N and β is B.

In the following, we define $\mathsf{B}(\Delta)$ as the set of four-valued models that satisfy Δ by the Belnap satisfaction relation and the domain is exactly the set of atoms occuring in Δ.

Definition 11 *The* **satisfaction relation** *for Belnap's four-valued logic, denoted $\models_b$ is defined as follows, where X^D is a four-valued model and $\alpha \in L$.*

$$X^D \models_b \alpha \text{ iff } t(\alpha) \geq_k T$$

For $\Delta \in \wp(L)$, let $X^D \models_b \Delta$ denote that $X^D \models_b \alpha$ for every $\alpha \in \Delta$. Let $\mathsf{B}(\Delta)$ be the set of **Belnap models** *for Δ defined as follows.*

$$\mathsf{B}(\Delta) = \{X^D \mid X^D \models_b \Delta \text{ and } \mathsf{Atoms}(\Delta) = D\}$$

Example 8 *Let $\Delta = \{\neg\alpha, \alpha \vee \beta\}$. So $D = \{\alpha, \beta\}$ and $\mathsf{B}(\Delta) = \{X_1^D, X_2^D, X_3^D, X_4^D, X_5^D, X_6^D\}$ where the models are defined as follows.*

X_i^D	$t_{X_i}^D(\alpha)$	$t_{X_i}^D(\beta)$	Concordance	Coherence
X_1^D	F	T	1	1
X_2^D	F	B	$1/2$	$1/2$
X_3^D	B	N	$1/2$	0
X_4^D	B	F	$1/2$	$1/2$
X_5^D	B	T	$1/2$	$1/2$
X_6^D	B	B	0	0

Example 9 *Let $\Delta = \{\neg\alpha, \alpha \vee \beta, \neg\beta, \gamma\}$. So $D = \{\alpha, \beta, \gamma\}$ and $\mathsf{B}(\Delta) = \{X_1^D, X_2^D, X_3^D, X_4^D, X_5^D, X_6^D\}$ where the models are defined as follows.*

X_i^D	$t_{X_i}^D(\alpha)$	$t_{X_i}^D(\beta)$	$t_{X_i}^D(\gamma)$	Concordance	Coherence
X_1^D	B	F	T	$2/3$	$2/3$
X_2^D	F	B	T	$2/3$	$2/3$
X_3^D	B	B	T	$1/3$	$1/3$
X_4^D	B	F	B	$1/3$	$1/3$
X_5^D	F	B	B	$1/3$	$1/3$
X_6^D	B	B	B	0	0

From Examples 8 and 9, we see that it is not the case in general that all models in $\mathsf{B}(\Delta)$ have the same coherence or concordance. So we extend coherence and concordance to knowledgebases as follows.

Definition 12 *Let* $\Delta \in \wp(L)$.

$$\mathsf{Coherence}(\Delta) = \mathsf{Max}(\{\mathsf{Coherence}(X^D) \mid X^D \in \mathsf{B}(\Delta)\})$$
$$\mathsf{Concordance}(\Delta) = \mathsf{Max}(\{\mathsf{Concordance}(X^D) \mid X^D \in \mathsf{B}(\Delta)\})$$

Example 10 *Continuing Example 8, we have* $\Delta = \{\neg\alpha, \alpha \vee \beta\}$*, and so we get the following.*

$$\mathsf{Coherence}(\Delta) = \mathsf{Concordance}(\Delta) = 1$$

Example 11 *Continuing Example 9, we have* $\Delta = \{\neg\alpha, \alpha \vee \beta, \neg\beta, \gamma\}$*, and so we get the following.*

$$\mathsf{Coherence}(\Delta) = \mathsf{Concordance}(\Delta) = 2/3$$

The definition of coherence (and similarly concordance) for knowledgebases can be considered as an optimistic view on the nature of inconsistency in the knowledgebase, in contrast to a pessimistic view where coherence for a knowledgebase Δ would be the minimum in $\{\mathsf{Coherence}(X^D) \mid X^D \in \mathsf{B}(\Delta)\}$. Other possibilities include using a leximax function or an average function. For example, using leximax can allow more fine grained discrimination between a knowledgebase that has all its models having coherence equal to 1, and a knowledgebase that has some models equal to 1, and some less than 1.

For some applications, the coherence or concordance measures are useful to analyse the quality of information. For example, sources of structured news reports can be monitored over time. If a source, has an average degree of coherence or concordance below a certain level, then some investigation of that source could be undertaken to see if the average quality could be improved.

However, in general for deciding how to act on structured news reports, we appear to need more than just the coherence and concordance measures. As we saw in the introduction, it is clear that some inconsistencies in news reports are very significant whereas others are insignificant. If we are to consider, for example, rejecting news reports on the basis of the inconsistencies arising, we will want to consider more than just the number of inconsistencies arising, we will want to take their significance into account.

5 Evaluating significance of inconsistencies

We now evaluate the significance of inconsistencies in four-valued models, and thereby in sets of formulae. The approach is based on specifying the relative significance of incoherent models using the notion of a mass assignment which is defined below.

Definition 13 *Let X^D be a four-vaued model. If* $\mathsf{Coherence}(X^D) = 0$*, then X^D is a* **frame**. *If X_1^D and X_2^D are frames, and $X_1^D \subseteq X_2^D$, then X_1^D is a subframe of X_2^D.*

Definition 14 *Let X^D be a frame. A* **mass assignment** *m for X^D is a function from $\wp(X^D)$ into $[0, 1]$ such that:*

(1) *If* $Y \subseteq X^D$ *and* $\mathsf{Coherence}(Y) = 1,$ *then* $m(Y) = 0$

(2) $\Sigma_{Y \subseteq X^D}\ m(Y) = 1$

In the definition for mass assignment, we have the constraint $\mathsf{Coherence}(X^D) = 0$ to ensure that for all $\alpha \in D$, we have $\pm\alpha \in X^D$. Condition 1 ensures mass is only assigned to models that contain conflicts and condition 2 ensures the total mass distributed sums to 1.

Given a frame X^D, a mass assignment can be localized on small subsets of X^D, spread over many subsets of X^D, or limited to large subsets of X^D. A mass assignment can be regarded as a form of metaknowledge, and so it needs to be specified for a domain, where the domain is characterized by X^D, and so the possible models of the domain are subsets of X^D.

Example 12 *Let $X^D = \{\pm\alpha, \pm\beta\}$. A mass assignment m is given by $m(\{\pm\alpha\}) = 0.2$ and $m(\{\pm\beta\} = 0.8$. Another mass assignment m' is given by $m'(\{\pm\alpha\}) = 0.2$, $m'(\{\pm\alpha, -\beta\}) = 0.6$, and $m'(\{\pm\alpha, \pm\beta\}) = 0.2$.*

A significance function gives an evaluation of the significance of the conflicts in a QC model. This evaluation is in the range $[0, 1]$ with 0 as least significant and 1 as most significant.

Definition 15 *Let m be a mass assignment for X^D. A* **significance function** *for X^D, denoted S_{X^D}, is a function from $\wp(X^D)$ into $[0, 1]$. A* **mass-based significance function** *for m, denoted $S^m_{X^D}$, is a significance function defined as follows for each $Y \in \wp(X^D)$.*

$$S^m_{X^D}(Y) = \Sigma_{Z \subseteq Y}\ m(Z)$$

We will only consider mass-based significance functions in the rest of this paper. So to ease reading in the following, we drop the superscript and subscript for significance functions.

Example 13 *Let $X^D = \{\pm\alpha, \pm\beta, \pm\gamma\}$, and let m be a mass assignment for X^D,* *where* $m(\{\pm\alpha\}) = 0.4$*,* $m(\{\pm\alpha, -\beta\}) = 0.2$*, and* $m(\{-\beta, \pm\gamma\}) = 0.4$*. Below we consider the significance for some subsets of* X^D*.*

$$S(\{\pm\alpha\}) = 0.4 \quad S(\{\pm\alpha, -\beta\}) = 0.6 \quad S(\{\pm\alpha, \pm\beta\}) = 0.6$$
$$S(\{\pm\alpha + \gamma\}) = 0.4 \; S(\{\pm\beta, \pm\gamma\}) = 0.4 \; S(\{\pm\alpha, \pm\beta, \pm\gamma\}) = 1.0$$

The definitions for mass assignment and mass-based significance correspond to mass assignment and belief functions (respectively) in Demspter-Shafer theory [11]. However, here they are used to formalise significance rather than uncertainty.

In the following result, we see that mass-based significance is not additive. Also the remaining significance need not be for the complement of Z (ie, Z^c). Some may be assigned to models not disjoint from Z.

Proposition 2 *Let* X^D *be a frame. If* m *is a mass assignment for* X^D*, and* S *is a significance function for* m*, then for all* $Z, Y \in \wp(X^D)$*, the following hold, where* Y^c *is the set complement of* Y *(i.e.* $Y^c = X^D \setminus Y$*).*

(1) $Z \subseteq Y$ *implies* $S(Z) \leq S(Y)$

(2) $S(Z \cup Y) \geq (S(Z) + S(Y) - S(Z \cap Y))$

(3) $S(Y) + S(Y^c) \leq 1$

We now extend the significance functions to knowledgebases. Since $\mathsf{B}(\Delta)$ is not necessarily a singleton, the significance for a set of formulae Δ is the lowest significance obtained for an $X \in \mathsf{B}(\Delta)$. This means we treat the information in Δ as a "disjunction" of four-valued models, and we regard each of those models as equally acceptable, or equivalently we regard each of those models as equally representative of the information in Δ.

Definition 16 *Let* $\Delta \in \wp(L)$*. We extend the definition for a significance function* S *to knowledgebases as follows.*

$$S(\Delta) = \mathsf{min}(\{S(X) \mid X \in \mathsf{B}(\Delta)\})$$

Some knowledgebases have zero significance. If $\Delta \not\vdash \bot$, and hence $\mathsf{Coherence}(\Delta) = 1$, then $S(\Delta) = 0$. Also significance is monotonic. So for $\Delta \in \wp(L)$ and $\alpha \in L$, $S(\Delta) \leq S(\Delta \cup \{\alpha\})$.

Example 14 *Let* $X^D = \{\pm\alpha, \pm\beta, \pm\gamma\}$*. Let* $m(\{\pm\alpha\}) = 0.6$*,* $m(\{\pm\alpha, +\beta\}) = 0.3$*, and* $m(\{\pm\beta, +\gamma\}) = 0.1$*. So* $S(\{\alpha \wedge \neg\alpha, \beta \vee \gamma\}) = 0.6$*.*

In order to determine the frame for which a mass function is defined, we can use the delineation function as follows.

Definition 17 *For $\Delta \in \wp(L)$, $\mathsf{Delineation}(\Delta) = \{\pm\alpha \mid \alpha \in \mathsf{Atoms}(\Delta)\}$.*

Example 15 *Let $\Delta_1 = \{\neg\alpha, \alpha \wedge \beta, \neg\beta\}$, $\Delta_2 = \{\alpha \vee \beta, \neg\alpha \wedge \alpha\}$, and $\Delta_3 = \{\beta, \neg\alpha \vee \neg\beta\}$. Let the frame $X^D = \mathsf{Delineation}(\Delta_1 \cup \Delta_2 \cup \Delta_3) = \{\pm\alpha, \pm\beta\}$. Also let $m(\{\pm\alpha, \pm\beta\}) = 0.2$ and $m(\{\pm\alpha\}) = 0.8$. So $S(\Delta_1) = 1$, $S(\Delta_2) = 0.8$, and $S(\Delta_3) = 0$.*

A mass assignment can be regarded as transforming the four-valued semantics of Belnap logic into a many-valued logic where the value B has been split into a chain of truth values $B_0, .., B_1$. If we equate the truth values $\{N, T, F, B_0\}$ with the Belnap truth values $\{N, T, F, B\}$, respectively, then Belnap's four-valued lattice is a sublattice of this lattice.

6 Modular mass assignments

In practice, we may need to consider a very large frame for monitoring coherence. This may therefore involve considering many different subsets to which mass has to be assigned. One solution is to take a modular approach by defining separate mass assignments to disjoint subsets of the frame, and then combining these mass assignments to give a mass assignment for the overall frame.

Definition 18 *Let $X_1^{D_1}$ be a frame with mass m_1, and let $X_2^{D_2}$ be a frame with mass m_2. m_1 is* **embedded** *in m_2 iff (1) $D_1 \subseteq D_2$; (2) for all $Y \in \wp(X_1^D)$ $m_1(Y) \geq m_2(Y)$; and (3) for all $Z, Y \in \wp(X_1^D)$, if $m_1(Z) \geq m_1(Y)$, then $m_2(Z) \geq m_2(Y)$.*

Embedding is not necessarily unique as illustrated by the following example.

Example 16 *Let $X_1^D = \{\pm\alpha\}$ and let $X_2^D = \{\pm\alpha, \pm\beta\}$. If $m_1(\{\pm\alpha\}) = 1.0$, then two possible embeddings are given by m_2 and m_2' as follows:*

$$m_2(\{\pm\alpha\}) = 0.2 \;\; m_2(\{\pm\beta\}) = 0.8$$

$$m_2'(\{\pm\alpha\}) = 0.7 \;\; m_2'(\{\pm\beta\}) = 0.3$$

The following definition gives a non-unique splitting of a frame into a set of subframes called a framesplit. A framesplit together with a mass assignment for each subframe in the framesplit is called a framebox.

Definition 19 *Let X^D be a frame and let $\Pi \subseteq \wp(X^D)$. $\Pi = \{X_1^{D_1}, ..., X_n^{D_n}\}$ is a* **framesplit** *for X^D iff (1) $D = D_1 \cup ... \cup D_n$; and (2) for all $X_i^{D_i}, X_j^{D_j} \in \Pi$,*

$D_i \cap D_j = \emptyset$*; and (3) for all* $X_i^{D_i} \in \Pi$, $\mathsf{Coherence}(X_i^{D_i}) = 0$. Ψ *is a* **massclass** *for* Π *iff for each* $X_i^{D_i} \in \Pi$ *there is exactly one mass assignment* $m_i \in \Psi$ *such that* m_i *is a mass assignment for* $X_i^{D_i}$. *A* **framebox** *for* X^D *is a tuple* $\langle \Pi, \Psi \rangle$ *where* Π *is a framesplit for* X^D *and* Ψ *is a massclass for* Π.

For a framebox, the following defines the embed function which gives the set of all mass assignments such that each is an embedding of all of the mass assignments in the framebox.

Definition 20 *Let* $\langle \Pi, \Psi \rangle$ *be a framebox for* X^D. *The embed function, denoted* Embed*, is defined as follows:*

$$\mathsf{Embed}(\Pi, \Psi) = \{m_i \mid \text{ for all } m_j \in \Psi \ \ (m_j \text{ is embedded in } m_i)\}$$

We adopt the following type of mass assignment which only assigns mass to disjoint models that are totally incoherent.

Definition 21 *Let* m *be a mass assignment for* X^D. m *is* **acute** *iff (1) for all* $Y \in \wp(X^D)$ *if* $\mathsf{Coherence}(Y) = 0$ *then* $m(Y) \geq 0$*; and (2) for all* $Y \in \wp(X^D)$ *if* $\mathsf{Coherence}(Y) > 0$ *then* $m(Y) = 0$*; and (3) for all* $Z, Y \in \wp(X^D)$ *if* $m(Z) > 0$ *and* $m(Y) > 0$ *then* $Z \cap Y = \emptyset$.

Significance is additive for totally incoherent models when the mass assignment is acute.

Proposition 3 *Let* m *be an acute mass assignment for* X^D. *Let* S *be a mass-based significance function for* m *and let* $Y \in \wp(X^D)$. *If* $\mathsf{Coherence}(Y) = 0$, *then* $S(Y) + S(Y^c) = 1$.

A useful feature of an acute mass-based significance function is that as the number of conflicts rises in a model, then the significance of the model rises. This is formalized by the following notion of conflict cumulativity. It does not hold in general (see Example 17).

Proposition 4 *Let* m *be an acute mass assignment for* X^D. *If* S *is a significance function for* m, *then the following property of conflict cumulativity holds for all* $Z^D, Y^D \in \wp(X^D)$.

$$\mathsf{Concordance}(Z^D) \geq \mathsf{Concordance}(Y^D) \text{ implies } S(Z^D) \leq S(Y^D)$$

Example 17 *Let* $X^D = \{\pm\alpha, \pm\beta, \pm\gamma\}$, $Z^D = \{\pm\alpha, +\gamma\}$ *and* $Y^D = \{\pm\alpha, \pm\beta\}$ *Let* m *be an mass assignment for* X^D *such that* $m(Z^D) = 0.6$ *and* $m(Y^D) = 0.4$. *So we have* $\mathsf{Concordance}(Z^D) \geq \mathsf{Concordance}(Y^D)$, *but* $S(Z^D) \not\leq S(Y^D)$.

We now consider how we can combine the modular mass assignments in a framebox. For this we use the following definition of a fair assignment. It is perhaps the simplest and least biased of the possible ways to combine the mass assignments.

Definition 22 *Let $\langle \Pi, \Psi \rangle$ be a framebox for X^D where each $m_i \in \Psi$ is an acute mass assignment. A* **fair assignment***, denoted m, for $\langle \Pi, \Psi \rangle$ is defined as follows: For each $X_i^{D_i} \in \Pi$, and for each $m_i \in \Psi$, such that m_i is a mass assignment for $X_i^{D_i}$, if $Y \subseteq X_i^{D_i}$, then*

$$m(Y) = m_i(Y) \times \frac{|D_i|}{|D|}$$

Essentially, the fair mass assignment is obtained from a normalized version of each of the mass assignments in the framebox.

Example 18 *Let $\langle \{X_1^{D_1}, X_2^{D_2}\}, \{m_1, m_2\} \rangle$ be a framebox for $X_3^{D_3}$ where $X_1^{D_1} = \{\pm\alpha\}$, $X_2^{D_2} = \{\pm\beta\}$, $X_3^{D_3} = \{\pm\alpha, \pm\beta\}$, and m_1 and m_2 are defined below. For this framebox, the fair assignment is m_3.*

$$m_1(\{\pm\alpha\}) = 1.0 \; m_2(\{\pm\beta\}) = 1.0$$
$$m_3(\{\pm\alpha\}) = 0.5 \; m_3(\{\pm\beta\}) = 0.5$$

Example 19 *Let $\langle \{X_1^{D_1}, X_2^{D_2}\}\{m_1, m_2\} \rangle$ be a framebox for $X_3^{D_3}$, where $X_1^{D_1} = \{\pm\alpha, \pm\gamma\}$, $X_2^{D_2} = \{\pm\beta\}$, and $X_3^{D_3} = \{\pm\alpha, \pm\beta, \pm\gamma\}$, and m_1 and m_2 are defined below. For this framebox, the fair assignment is m_3.*

$$m_1(\{\pm\alpha\}) = 0.4 \;\; m_1(\{\pm\gamma\}) = 0.6 \;\; m_2(\{\pm\beta\}) = 1.0$$
$$m_3(\{\pm\alpha\}) = 0.27 \; m_3(\{\pm\gamma\}) = 0.4 \; m_3(\{\pm\beta\}) = 0.33$$

Proposition 5 *Let $\langle \Pi, \Psi \rangle$ be a framebox for X^D. If m is a fair assignment for $\langle \Pi, \Psi \rangle$, then m is a mass assignment for X^D.*

Proof: Let m be a fair assignment. So Condition (1) of Definition 14 is trivially satisfied. For each frame $X_i^{D_i} \in \Pi$, with mass assignment m_i,

$$\sum_{Y \subseteq X_i^{D_i}} m_i(Y) = 1$$

Hence, for each frame $X_i^{D_i} \in \Pi$, with mass assignment m_i,

$$\sum_{Y \subseteq X_i^{D_i}} (m_i(Y) \times \tfrac{|D_i|}{|D|}) = (\sum_{Y \subseteq X_i^{D_i}} \; m_i(Y)) \times \tfrac{|D_i|}{|D|} = \tfrac{|D_i|}{|D|}$$

Since, the frames in Π are disjoint, i.e. Conditions (1) and (2) of Definition 19 hold.

$$\sum_{Y \subseteq X^D} \; m(Y) = \sum_{i \text{ such that } X_i^{D_i} \in \Pi} \; \tfrac{|D_i|}{|D|} = 1$$

So Condition (2) of Definition 14 is satisfied. Hence, m is a mass assignment for X^D. $\Box$

We can obtain a fair embedding using the embed function.

Proposition 6 *Let $\langle\Pi, \Psi\rangle$ be a framebox for X^D. If m is a fair assignment for $\langle\Pi, \Psi\rangle$, then $m \in \mathsf{Embed}(\Pi, \Psi)$.*

Proof: Let m be a fair assignment for $\langle\Pi, \Psi\rangle$. Let $\langle\Pi, \Psi\rangle$ be a framebox for X^D such that $\Pi = \{X_1^{D_1}, ..., X_n^{D_n}\}$. From these assumptions, we get the following: (1) $D = D_1 \cup .. \cup D_n$, and for each D_i, $D_j \subseteq D$, $D_i \cap D_j = \emptyset$; (2) for all $Y \in \wp(X_i^D)$ $m(Y) = m_i(Y) \times \frac{|D_i|}{|D|}$, and so $m_i(Y) \geq m(Y)$; and (3) for all $Z, Y \in \wp(X_i^D)$, if $m_i(Z) \geq m_i(Y)$, then $m(Z) \geq m(Y)$. Therefore, by Definition 18, (1), (2), and (3) imply that for all i, where $1 \geq i \geq n$, we have m_i is embedded in m. Hence, by Definition 20, $m \in \mathsf{Embed}(\Pi, \Psi)$. $\Box$

As a direct consequence of Definition 22, a fair assignment is a unique embedding.

Proposition 7 *Let $\langle\Pi, \Psi\rangle$ be a framebox for X^D. If m is a fair assignment for $\langle\Pi, \Psi\rangle$, and m' is a fair assignment for $\langle\Pi, \Psi\rangle$, then $m = m'$.*

A fair assignment is not necessarily an acute mass assignment. However, if a framebox only contains acute mass assignments, then the fair mass assignment for the framebox is acute.

Proposition 8 *Let $\langle\Pi, \Psi\rangle$ be a framebox for X^D. If all $m_i \in \Psi$ are acute mass assignments, and m is a fair assignment for $\langle\Pi, \Psi\rangle$, then m is an acute mass assignment.*

Proof: Let $\langle\Pi, \Psi\rangle$ be a framebox for X^D such that all $m_i \in \Psi$ are acute mass assignments. Also let m be a fair assignment for $\langle\Pi, \Psi\rangle$. (1) For all $Y \in \wp(X^D)$, if $\mathsf{Coherence}(Y) = 0$, then there is a $m_i \in \Pi$ such that $m_i(Y) \geq 0$, and hence $m(Y) \geq 0$; (2) For all $Y \in \wp(X^D)$, if $\mathsf{Coherence}(Y) > 0$, then there is not a $m_i \in \Pi$ such that $m_i(Y) > 0$, and hence $m(Y) = 0$; (3) For all $m_i, m_j \in \Pi$, for all $Z, Y \in \wp(X^D)$, if $m_i(Z) > 0$, and $m_j(Y) > 0$, then $Z \cap Y = \emptyset$, hence if $m(Z) > 0$, and $m(Y) > 0$, then $Z \cap Y = \emptyset$. From (1), (2), and (3), by Definition 21, m is an acute mass assignment. $\Box$

So a framebox for a frame X^D with a fair assignment provides a simple and effective modular approach to mass assignment for larger frames.

Example 20 *Consider a framebox $\langle\Pi, \Psi\rangle$ where m_1 and m_2 are defined as follows. For this example,* `t` *and* `r` *are abbreviations for elements in the domain where* `t` *denotes* temperature *and* r *denotes* rainfall.

$$m_1(\pm\texttt{t(1C)}) = 0.02, m_1(\pm\texttt{t(2C)}) = 0.02, .., m_1(\pm\texttt{t(50C)}) = 0.02$$
$$m_2(\pm\texttt{r(1cm)}) = 0.02, m_2(\pm\texttt{r(1cm)}) = 0.02, .., m_2(\pm\texttt{r(50cm)}) = 0.02$$

A fair assignment gives m_3 as follows.

$$m_3(\pm\texttt{t(1C)}) = 0.01, m_3(\pm\texttt{t(2C)}) = 0.01, .., m_3(\pm\texttt{t(50C)}) = 0.01$$
$$m_3(\pm\texttt{r(1cm)}) = 0.01, m_3(\pm\texttt{r(1cm)}) = 0.01, .., m_3(\pm\texttt{r(50cm)}) = 0.01$$

Hence, significance for some subsets are as follows.

$$S_1(\{\pm\texttt{t(1C)}, \pm\texttt{t(2C)}, .., \pm\texttt{t(7C)}\}) = 0.14$$
$$S_1(\{\pm\texttt{t(1C)}, \pm\texttt{t(2C)}, .., \pm\texttt{t(15C)}\}) = 0.30$$
$$S_1(\{\pm\texttt{t(21C)}, \pm\texttt{t(22C)}, .., \pm\texttt{t(35C)}\}) = 0.30$$

$$S_3(\{\pm\texttt{t(1C)}, \pm\texttt{t(2C)}, .., \pm\texttt{t(7C)}\}) = 0.07$$
$$S_3(\{\pm\texttt{t(1C)}, \pm\texttt{t(2C)}, .., \pm\texttt{t(15C)}\}) = 0.15$$
$$S_3(\{\pm\texttt{t(1C)}, \pm\texttt{t(2C)}, .., \pm\texttt{t(45C)}, \pm\texttt{r(5cm)}, \pm\texttt{t(6cm)}, .., \pm\texttt{t(9cm)}\}) = 0.50$$

For domains with some numerical data, such as sport reports or weather reports, mass can be assigned to particular models, and then mass for further models is obtained by interpolation.

7 Classification of inconsistencies using thresholds

We can classify structured news reports on the basis of the significance of the inconsistencies that arise in them. For any set of formulae Ψ, we can set a threshold T^r for rejection of the formulae where $T^r \in [0, 1]$. So if $T^r < S(\Psi)$, then Ψ should be rejected, whereas if $T^r \geq S(\Psi)$, then Ψ should be accepted, perhaps after some resolution of inconsistencies.

Further thresholds can be adopted for finer grained selection of actions. For example, we could adopt a threshold T^i for ignoring inconsistencies. So if $S(\Psi) < T^i$, then the inconsistencies in Ψ are ignored, whereas if $S(\Psi) \geq T^i$, then the inconsistencies in Ψ should be amended. Intermediate thresholds between T^i and T^r could be used to select the resources committed to resolve the inconsistencies. In the rest of this paper we will just consider rejection thresholds. The definitions and results can be easily extended to handle any number of thresholds.

In this section, we will consider thresholds for models, and in the next section for formulae. A threshold should change with size/structure of frame/mass assignment. As the frame increases, and hence as the mass is distributed more widely, the action and rejection thresholds should fall.

Definition 23 *Let T_1^r be a rejection threshold for $X_1^{D_1}$ with mass m_1, and let T_2^r be a rejection threshold for $X_2^{D_2}$ with mass m_2, where $D_1 \subseteq D_2$ and m_1 is embedded in m_2. Let S_1 be a mass-based significance function for m_1 and let S_2 be a mass-based significance function for m_2. If the condition below holds for all $Y \in \wp(X_1^{D_1})$, then T_2^r is* **careful relaxation** *of T_1^r.*

$$S_1(Y) > T_1^r \text{ iff } S_2(Y) > T_2^r$$

If T_2^r is careful relaxation of T_1^r, then any model Y, classified for rejection using m_1 and T_1^r, is also classified for rejection using m_2 and T_2^r.

Proposition 9 *Let T_1^r be a rejection threshold for X_1^D with mass m_1, and T_2^r be a rejection threshold for X_2^D with mass m_2, and m_1 is embedded in m_2. If T_2^r is a careful relaxation of T_1^r then $T_1^r \geq T_2^r$.*

Proof: Let m_1 be embedded in m_2. Therefore, for all $Y \in X^D$, $m_1(Y) \geq m_2(Y)$. Hence, for all $Y \in X^D$, $S_1(Y) \geq S_2(Y)$. So, if T_2^r is a careful relaxation of T_1^r, then $T_1^r \geq T_2^r$. □

Example 21 *We continue Example 20. Let $T_1^r = 0.2$ be a rejection threshold for X_1^D with mass m_1, and $T_3^r = 0.1$ be a rejection threshold for X_3^D with mass m_3. Let S_1 be a mass-based significance function for m_1 and let S_3 be a mass-based significance function for m_3. So for selected models, we have the following.*

$$S_1(\{\pm\texttt{t(1C)}, \pm\texttt{t(2C)}, .., \pm\texttt{t(7C)}\}) < T_1^r$$

$$S_1(\{\pm\texttt{t(1C)}, \pm\texttt{t(2C)}, .., \pm\texttt{t(15C)}\}) > T_1^r$$

$$S_3(\{\pm\texttt{t(1C)}, \pm\texttt{t(2C)}, .., \pm\texttt{t(7C)}\}) < T_3^r$$

$$S_3(\{\pm\texttt{t(1C)}, \pm\texttt{t(2C)}, .., \pm\texttt{t(15C)}\}) > T_3^r$$

In effect, the absolute values for significance evaluations and thresholds are not important per se. However the combination of them allows us to classify inconsistencies. Furthermore, the thresholds can be relaxed for larger frames whilst preserving the same classification given to inconsistencies in subframes.

8 Significance of inconsistencies in news

We now return to how we can analyse inconsistent news. We assume news is input in the form of structured news reports and that the corresponding atom representing the structured news report can be used with a set of access rules to give a set of inferences as explained in Section 2. As a simple illustration, some atoms representing structured news reports, denoted $\{\rho_1, .., \rho_5\}$ are given in Example 22.

Example 22 *Each of $\rho_1 - \rho_5$ is an atom representing a structured news report.*

$$\mathsf{Access}(\Delta, \rho_1) = \{\texttt{temp}(\texttt{30C}), \texttt{pptn}(\texttt{snow}), \texttt{pollen}(\texttt{high})\}$$
$$\mathsf{Access}(\Delta, \rho_2) = \{\texttt{temp}(\texttt{30C}), \texttt{pptn}(\texttt{snow}), \texttt{pollen}(\texttt{low})\}$$
$$\mathsf{Access}(\Delta, \rho_3) = \{\texttt{temp}(\texttt{10C}), \texttt{pptn}(\texttt{snow}), \texttt{pollen}(\texttt{high})\}$$
$$\mathsf{Access}(\Delta, \rho_4) = \{\texttt{temp}(\texttt{10C}), \texttt{pptn}(\texttt{snow}), \texttt{pollen}(\texttt{low})\}$$
$$\mathsf{Access}(\Delta, \rho_5) = \{\texttt{temp}(\texttt{0C}), \texttt{pptn}(\texttt{snow}), \texttt{pollen}(\texttt{high})\}$$

Given a structured news report, we want to consider its consistency with respect to domain knowledge that we may possess. We will represent this domain knowledge by a set of propositional formulae. A potential inconsistency that can arise between the information obtained from a structured news report is any set of atoms that may be rebutted by the domain knowledge.

Definition 24 *Let ρ be a report and let Γ be domain knowledge. For $\{\phi_1, .., \phi_n\} \subseteq$ $\mathsf{Access}(\Delta, \rho)$, $\{\phi_1, .., \phi_n\}$ is* **rebutted** *by Γ iff $\Gamma \vdash \neg(\phi_1 \wedge \ldots \wedge \phi_n)$.*

So, if $\{\phi_1, .., \phi_n\}$ is **rebutted** by Γ, then $\Gamma \cup \{\phi_1, .., \phi_n\}$ is inconsistent.

Example 23 *Domain knowledge for weather reports may include the following three clauses.*

(clause 1) $\neg\texttt{temp}(\texttt{30C}) \vee \neg\texttt{pptn}(\texttt{snow})$

(clause 2) $\neg\texttt{temp}(\texttt{10C}) \vee \neg\texttt{pptn}(\texttt{snow})$

(clause 3) $\neg\texttt{pollen}(\texttt{high}) \vee \neg\texttt{pptn}(\texttt{snow})$

So with the report ρ_1 in Example 22, we get that $\{\texttt{temp}(\texttt{30C}), \texttt{pptn}(\texttt{snow})\}$ is rebutted by clause 1, and that $\{\texttt{pollen}(\texttt{high}), \texttt{pptn}(\texttt{snow})\}$ is rebutted by clause 3.

Now we need to consider how we can handle these inconsistencies. First, we apply our mass assignment to evaluate the significance of the inconsistencies.

Example 24 *Continuing Example 22 and 23, we suppose we have a mass assignment as follows.*

$$m(\{\pm\texttt{temp(30C)}, \pm\texttt{pptn(snow)}\}) = \texttt{0.6}$$

$$m(\{\pm\texttt{temp(10C)}, \pm\texttt{pptn(snow)}\}) = \texttt{0.3}$$

$$m(\{\pm\texttt{pollen(high)}, \pm\texttt{pptn(snow)}\}) = \texttt{0.1}$$

So the reports ρ_1 to ρ_5 in Example 22, together with the domain knowledge in Example 23, denoted here by Π, gives the following significance evaluations. $S(\Pi \cup \rho_1) = 0.7$, $S(\Pi \cup \rho_2) = 0.6$, $S(\Pi \cup \rho_3) = 0.4$, $S(\Pi \cup \rho_4) = 0.3$, $S(\Pi \cup \rho_5) = 0.1$. If we set the threshold of acceptability for a news report ρ_i at a significance evaluation of 0.3, then only ρ_4 and ρ_5 would be acceptable, the others would be rejected.

This example illustrates how we may find some inconsistencies acceptable and others unacceptable, and thereby select some news reports in preference to others. In practice, to harness this approach, we need to adopt more sophsticated (first-order) domain knowledge, and then ground subsets of the domain knowledge (and so create sets of propositional formulae) that are inconsistent with the (propositional) news reports.

To illustrate, consider a news report on a football match, where an inconsistency in the final score of the match could be assigned significance 0.6, an inconsistency in the name of the player who scored the most goals in less significant and could be assigned a value say 0.3, and inconsistency in the name of the referee is least significant and could be assigned a value say 0.1. Here, we could represent the domain knowledge by schema that would be instantiated by constant symbols from a news report. For example, $\texttt{referee(X)} \wedge \texttt{referee(Y)} \rightarrow \texttt{X} = \texttt{Y}$ is schema that would be ground by constant symbols in a particular football news report. Inconsistency analysis would then be undertaken on the ground propositional formulae rather than the original schema.

Similarly, for anything other than a very restricted domain, we are likely to want the mass assignment to be done for schema. For example, instead of a mass assignment for the following frame we adopt a schema representation of the form $m(\{\pm\texttt{referee(X)}\}) = \texttt{0.6}$ which says the mass assigned to any possible grounding for $\texttt{X}$ is 0.6.

$$\{\pm\texttt{referee(MikeDean)}, \pm\texttt{referee(SteveBennett)},$$
$$\pm\texttt{referee(MartinAtkinson)}, \pm\texttt{referee(MikeRiley)}, ...\}$$

To do this in practice, we would also need some simple integrity constraints to ensure that we maintain some highly desirable conditions of mass assignments. We leave the development of schema and associated integrity constraints to a subsequent paper.

9 Discussion

The traditional view in logic is that a set of formulae is either consistent or inconsistent. However, for artificial intelligence, we need more than this binary classification. We need to be able to better describe the nature of inconsistency in a set of formulae. Some proposals for measuring inconsistency have been put forward (see for example [12,13,7,14]), but the field needs further development including in terms of evaluating the significance of inconsistency [15].

In this paper, we have developed an approach to measuring of inconsistency based on four-valued logic. In order to evaluate the significance of inconsistency, we have adopted the only proposal (as far as we know) for this [7]. In [7], QC logic was used for finding the four-valued models for a set of formulae. In this paper, we have adopted a simpler and better-known four-valued logic that was proposed by Belnap [9,10].

In addition, we have gone beyond the proposal in [7] by considering in more detail how evaluating inconsistency can be used to decide on a course of action for dealing with inconsistent structured news reports. In this paper, we have introduced the notion of modular mass assignments, and the usage of thresholds for deciding courses of action.

As a result of the framework in this paper, we now have a potentially valuable way of filtering structured news reports before making them available in a database or newsfeed, or processing them further by for example merging potentially conflicting reports from heterogeneous sources to create more comprensive, better confirmed, less redundant and less uncertain aggregated news reports (see for example developments in fusion rule technology www.cs.ucl.ac.uk/staff/a.hunter/frt).

Another framework for merging potentially conflicting knowledge, by Cholvy and Garion [16], has been implemented in Prolog. It appears to be an ideal platform for future work for investigating the role of measures of degree and significance of inconsistency in knowledge fusion. In this, mass assignments could be used as meta-knowledge for providing a finer grained approach to aggregation than those based on social choice theory.

Whilst the motivation and examples in this paper have been about news reports, we have already suggested that we can take a broad view of news reports. Many different sources of information can be described as providing news reports. Furthermore, much information can be represented in the form of structured news reports, including weather reports, business reports, summaries of scientific articles, results from clincal trials, market intelligence, operations reports within organisations, etc. We are therefore considering any kind of regular supply of information that conforms to some stereotypical format. This means that the import of inconsistency, and the associated problems of inconsistency, are wide-spread. The framework in this paper, offers a solution to some of these problems. Furthermore, it may indeed form a valuable part of inconsistency tolerance in data and knowledge systems, more generally.

References

[1] J. Cowie, W. Lehnert, Information extraction, Communications of the ACM 39 (1996) 81–91.

[2] A. Hunter, Logical fusion rules for merging structured news reports, Data and Knowledge Engineering 42 (2002) 23–56.

[3] A. Hunter, R. Summerton, Fusion rules for context-dependent aggregation of structured news reports, Journal of Applied Non-classical Logic 14 (3) (2004) 329–366.

[4] E. Byrne, A. Hunter, Man bites dog: Looking for interesting inconsistencies in structured news reports, Data and Knowledge Engineering 48 (3) (2004) 265–285.

[5] E. Byrne, A. Hunter, Evaluating violations of expectations to find exceptional information, Data and Knowledge Engineering(in press).

[6] P. Besnard, A. Hunter, A logic-based theory of deductive arguments, Artificial Intelligence 128 (2001) 203–235.

[7] A. Hunter, Evaluating the significance of inconsistency, in: Proceedings of the International Joint Conference on AI (IJCAI'03), Morgan Kaufmann, 2003, pp. 468–473.

[8] A. Hunter, Merging structured text using temporal knowledge, Data and Knowledge Engineering 41 (2002) 29–66.

[9] N. Belnap, How a computer should think, in: G. Ryle (Ed.), Contemporary Aspects of Philosophy, Oriel Press, 1976, pp. 30–55.

[10] N. Belnap, A useful four-valued logic, in: G. Epstein (Ed.), Modern Uses of Multiple-valued Logic, Reidel, 1977, pp. 8–37.

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

[12] E. Lozinskii, Information and evidence in logic systems, Journal of Experimental and Theoretical Artificial Intelligence 6 (1994) 163–193.

[13] A. Hunter, Measuring inconsistency in knowledge via quasi-classical models, in: Proceedings of the 18th National Conference on Artificial Intelligence (AAAI'2002), MIT Press, 2002, pp. 68–73.

[14] S. Konieczny, J. Lang, P. Marquis, Quantifying information and contradiction in propositional logic through epistemic actions, in: Proceedings of the 18th International Joint Conference on Artificial Intellignce (IJCAI'03), Morgan Kaufmann, 2003, pp. 106–111.

[15] A. Hunter, S. Konieczny, Approaches to measuring inconsistent information, in: Inconsistency Tolerance, Vol. 3300 of Lecture Notes in Computer Science, Springer, 2004, pp. 189–234.

[16] L. Cholvy, C. Garion, Querying several conflicting databases, Journal Of Applied Non-Classical Logics 14 (3) (2004) 295–327.