Distributed Belief Revision as Applied Within a Descriptive Model of Jury Deliberations

Aldo Franco Dragoni,* Paolo Giorgini,** Ephraim Nissan***

* Istituto di Informatica, University of Ancona, Via Brecce Bianche, 60131 Ancona, Italy. dragon@inform.unian.it
http://www.inform.unian.it/personale/~dragon/dragoni.html

** DISA, Univ. of Trento, Via Inama 5, Trento, Italy. pgiorgini@cs.unitn.it

*** School of Computing and Mathematical Sciences, University of Greenwich, Wellington Street, Woolwich, London SE18 6PF, U.K.
E.Nissan@greenwich.ac.uk http://www.gre.ac.uk/~E.Nissan
http://www.gre.ac.uk/~E.Nissan/CmsWeb/ai_and_law_group.html

Belief revision is a well-research topic within AI. We argue that the new model of distributed belief revision as discussed here is suitable for general modelling of judicial decision making, along with extant approach as known from jury research. The new approach to belief revision is of general interest, whenever attitudes to information are to be simulated within a multi-agent environment with agents holding local beliefs yet by interacting with, and influencing, other agents who are deliberating collectively. In the approach proposed, it's the entire group of agents, not an external supervisor, who integrate the different opinions. This is achieved through an election mecahnism. The principle of "priority to the incoming information" as known from AI models of belief revision are problematic, when applied to factfinding by a jury. The present approach incorporates a computable model for local belief revision, such that a principle of recoverability is adopted. By this principle, any previously held belief must belong to the current cognitive state if consistent with it. For the purposes of jury simulation such a model calls for refinement. Yet, we claim, it constitutes a valid basis for an open system where other AI functionalities (or outer stimuli) could attempt to handle other aspects of the deliberation which are more specific to legal narratives, to argumentation in court, and then to the debate among the jurors.

1 Jury Research

Reid Hastie's paper collection, *Inside the Juror* (1993), is now a classic of jury research, and the psychology of judicial decision making by lay factfinders: descriptive models of juror decision making. Already in 1983, Hastie, Penrod and Pennington had published *Inside the Jury.* This domain has eluded thus far the mainstream of AI & Law research.

It appears to be the case that the very first paper published in an AI forum in the domain was Gaines et al. (1996). It described a neural model simulating juror decision making according to one of the several approaches current in psychologists' formal modelling of juror decision making.

Disciplines contributing to the approaches presented in *Inside the Juror* include "social psychology, behavioral decision theory, cognitive psychology, and behavioral modeling" (from the blurb on the back cover), yet this list is not complete. For example, in Ch. 11, Ehud Kalai sketched a game-theoretic framework. The affiliation of the authors in that volume is with schools of Law, departments of Statistics, or Psychology, or Management, or social or political science, but none comes from computer science. The discipline will have to take notice. Ours is a step in that direction.

Hastie's long introduction to his volume is usefully detailed—an excellent overview which we summarise below by way of introduction to our own application of a new general model of distributed belief revision. "One development in traditional jurisprudential scholarship is a candidate for the role of a general theory of juror decision making; namely the utilitarian model of rational decision making that has been imported into jurisprudence from economics" (4). Optimal decision making has been modeled, in the literature, not just for the role of the juror, but for the judge, attorney, police, and perpetrators of criminal behavior as well. Optimality, or rationality, for decision making is too strong an assumption (5). The research in *Inside the Jury* "focuses on the manner in which jurors behave before they enter the social context of deliberation in criminal felony cases" (5), with "at least four competing approaches represented" among behavioral scientists' descriptive models of decision making (10), namely, such that are "based on probability theory, 'cognitive' algebra, stochastic processes, and information processing theory" (10–11). Bayes' theorem is involved, in the former, for descriptive purposes in *Inside the Juror*—being applied to the psychological processes in which a juror is engaged—rather than in prescribing how to evaluate evidence to reach a verdict, "or to evaluate and improve jurors' performance" (12).

Note that Bayesianism in legal evidence research is a controversial, hotly debated topic: Allen and Redmayne (1997) is a journal special issue with contributions from both camps, namely the Bayesio-skeptics and the so-called Bayesian enthusiasts. Yet, when it comes to descriptive models of how jurors shape their opinions, it's not obvious *prima facie* that the controversy extends into jury research. Jurors do *not* reason about the evidence according to the Bayes theorem, it may be argued, and even if they tried to apply probability explicitly, they would lack the formal skills to do so. This is beyond the point. Rather, among the descriptive models of juror decision making there are *also* probabilistic or stochastic models, to describe a process in general terms—not for a specific case at hand. This point is essential for making sense of our contribution in this paper.

The second class of approaches to juror's decision making, as enumerated in the introduction to Hastie's volume, fits among such psychological theories of mental processes that are couched in the form of algebraic equations (17), with evidence being combined according to a weighted average equation. "As in the Bayesian model, we are dealing with a single meter in which the results of all the subprocesses are summarized in a current belief and in which the ultimate 'categorical' verdict decision is based on the comparison of the final belief meter reading to a threshold to convict" (19), but belief updating in the algebraic approach is additive instead of multiplicative as in Bayesian models, and moreover extreme judgments are adjustable instead of final.

Stochastic process models are the third family; they differ from the previous two in that the larger process is assumed to behave in a random fashion, and what is probabilistic is state transitions over time. The fourth family adopts the information processing paradigm from cognitive psychology; they are typified by the room they make for mental representations, memory activation, elementary information processes, an executive monitor, and a specific cognitive architecture.

For example, Reid Hastie's Ch. 4 in his volume is devoted to algebraic models: the basic averaging and the sequential averaging models. "To date, the most visible accomplishments have been byproducts of the algebraic application; e.g., useful individual-level numerical indices of the importance of evidence, presumptions of innocence, and standards of proof" (110). Norbert Kerr (Ch. 5) is concerned with stochastic models; David A. Schum and Anne W. Martin (Ch. 6), with probabilistic evidence probativity assessment. Schum, by background a psychologist who has also researched at the meet of computing and operations research as well as law, is indeed one of the most visible representatives of the Bayesian camp within legal evidence research, yet he has also made other important contributions—especially his adaptation of Wigmorean analysis, as well as his repertoire of basic formal op[erations for marshalling the evidence—whose value is fairly acceptable also for Bayesio-skeptics and can arguably be embraced by AI & Law research with little risk of antagonising those sceptical about Bayesianism's value for the analysis of the evidence in a given legal case.

Schum's and Martin's chapter in Hasdtie's volume has a major focus "upon inductive inference tasks, which Wigmore [1937] termed 'catenated'; the modern terms for these tasks are 'cascaded' or 'hierarchical' " (136). In contrast Ch. 7, by Schum, which eventually also applies Bayesian likelihood-ratio formulations for weighing evidence, places more emphasis on argument structuring. Arguably, this could be an entry point into the domain for such AI & Law researchers whose interests are in argumentation models. In Ch. 8, Nancy Pennington and Reid Hastie present a cognitive theory of story construction on the part of the juror, and indeed we propose (see Nissan's paper on the JAMA model in this forum) that models of narrative understanding from natural-language processing are all-important, if one is to apply, next, AI & Law to narratives of a case at hand or, perhaps preferably, to narrative patterns for siituational classification purposes within problem-solving tasks.

This is the backdrop from jury research, for our proposed application of the new model of distributed belief revision based on the principle of recoverability as explained at the start.

2 A Novel General Approach

Jurors' opinions and beliefs are destined to evolve as the trial goes on. New information and evidence integrate and corroborate the cognizance of the Court, but other testimonies might cause conflicts. In this case, it seems natural that the acquisition of the new evidence s hould be accompanied by a reduction of the credibility of the conflicting pieces of knowledge. If the juror's corpus of evidence is not a flat set of facts but contains rules, finding such conflicts and determining all the sentences involved in the contradictions can be hard. In dealing with these "changes of mind" we heavily relies on symbolic logic, since as much as it contributed to the history of "thinking", logic could as well solve the problem of "thinking over". AI reserchers call this cognitive process "belief revision".

Since the seminal, philosophical and influential works of Alchourrón, Gärdenfors and Makinson (1985) ideas on "belief revision" have been progressively refined (Gärdenfors 1988) toward normative, effective and computable paradigms (Benferhat et al. 1993; Nebel 1994). They introduced three rational principles to whom belief revision should obey:

AGM1 *Consistency*: revision must yield a consistent knowledge space.

AGM2 *Minimal Change*: revision should alter as little as possible the knowledge space.

AGM3 *Priority to the Incoming Information*: incoming information always belongs to the revised knowledge space.

They conceived a cognitive state K as a deductively closed set of sentences of a formal language L. From $AGM1 \div AGM3$ they drew up eight postulates for belief revision. Here K^*p denotes the cognitive state K revised in the light of the incoming information p, while K^+p denotes the deductive closure of $K \cup \{p\}$.

K^*1. For each p and K, K^*p is still a cognitive state
K^*2. $p \in K^*p$
K^*3. $K^*p \subseteq K^+p$
K^*4. If $\neg p \notin K$ then $K^+p \subseteq K^*p$
K^*5. K^*p is inconsistent iff p is inconsistent
K^*6. If p and q are logically equivalent then $K^*p = K^*q$
K^*7. $K^*(p \wedge q) \subseteq (K^*p)^+q$
K^*8. If $\neg q \notin K^*p$ then $(K^*p)^+q \subseteq K^*(p \wedge q)$

These axioms describe the rational properties to which revision should obey, but they do not suggest how to perform it. An obvious way is that of deleting $\neg p$ from K (reducing in some way K at a point that $\neg p$ is no longer derivable), adding p and making the deductive closure The deletion of $\neg p$ from K, $K^-\neg p$, is called contraction, and can be defined in terms of "Epistemic Entrenchment" (Gärdenfors 1988), which is an ordering $\leq$, that envisages the logical dependencies of the formulae in K; it depends on K but it applies to all the formulae of L. $p \leq q$ means that p is less entrenched (i.e., more exposed to eventual changes) than q. $\leq$ satisfies the following postulates:

EE1. $\leq$ is transitive
EE2. For all p, $q \in L$, if $p \vdash q$ then $p \leq q$
EE3. For all p, $q \in L$, either $p \leq p \wedge q$ or $q \leq p \wedge q$
EE4. If K is consistent, then $p \notin K$ iff for all $q \in L$, $p \leq q$
EE5. If for all q of L, it holds $q \leq p$, then p is a tautology

Contraction could be defined from the Epistemic Entrenchment as follows: $q \in K^-p$ iff $q \in K$ and, either $p < q \vee p$, or p is a tautology. K^-p contains only the formulae of K that have a greater degree of epistemic entrenchment than p. There are three problems with such a kind of revision:

1. it deals with infinite sets of sentences
2. $\leq$ depends on K, so it is difficult to iterate the revision because the ordering defined on K^*p could be different from the one defined on K
3. the choice of a particular ordering $\leq$ satisfying the postulates EE1$\div$ EE5 is arbitrary; as Gärdenfors (1988) wrote: "[the postulates] leave the main problem unsolved: what is a reasonable metric for comparing different epistemic states?".

Indeed, regarding the latter problem, one of the claim of this paper is that, *such computable and reasonable metric can be provided only by numerical approaches.* The AGM approach to belief revision do respect Dalal's (1988) "principle of irrelevance of the syntax" by which, syntactically different but logically equivalent formulae represent the same knowledge space. The partisans of *syntax-dependent* belief revision consider knowledge spaces made up of a limited number of sentences. They claim that asserting facts is more important than deriving others from them. Nebel's (1994) *epistemic relevance ordering* stratifies a base B into n priority classes $B_1, ..., B_n$. Epistemic relevance does not respect the logical contents of the sentences as epistemic and partial entrenchment do. A justification seems to rely on the logical paradoxes of the material implication: a rule $q \rightarrow p$ should not necessarily be considered more important than p just because $p \vdash q \rightarrow p$. Let $B{\downarrow}p$ denote the set of the subsets of B that fail to imply p. Nebel defines $B{\Downarrow}p$ as the subset of $B{\downarrow}p$ made of the elements that *contain as many sentences of the highest priority as possible.*

The corresponding revision is defined as:

$$B \oplus p = Th\left(\left(\bigcap_{B' \in (B \Downarrow \neg p)} Th(B') \right) \cup \{p\} \right)$$

where $Th(B')$ denotes the deductive closure of B'. There are two problems with this revision: =

- it does not satisfy all the AGM postulates
- it is still computationally hard.

We could adopt various criteria to sort and select the elements of $B{\downarrow}p$. Let $B' = B_1^{'} \cup ... \cup B_n^{'}$ and $B'' = B_1^{''} \cup ... \cup B_n^{''}$ two consistent subsets of B where $B_i^{'} = B^{'} \cap B_i$ and $B_i^{''} = B^{''} \cap B_i$. Benferhat et al. (1993) [cf. Dubois & Prade (1992)] suggest (implicitly) three ways

to translate the epistemic relevance into a preference relation $\leq$ on $B \downarrow p$.

- *best-out ordering.* $B^{''} \leq B^{'}$ iff the most credible of the sentences in $B \setminus B^{''}$ is less credible than the most credible of the sentences in $B \setminus B^{'}$.
- *inclusion-based ordering.* $B^{''} \leq B^{'}$ iff there exists a stratum i such that $B_i^{'} \supset B_i^{''}$ and for any $j < i$, $B_j^{'} = B_j^{''}$. This preordering is strict but partial; its maximal consistent elements are also maximal for the best-out ordering.
- *lexicographic ordering.* $B^{''} \leq B^{'}$ iff there exists a stratum i such that $|B_i^{'}| > |B_i^{''}|$ and for any $j < i$ $|B_j^{'}| = |B_j^{''}|$, and $B^{''} = B^{'}$ iff for any j, $|B_j^{'}| = |B_j^{''}|$.

$B{\Downarrow}p$ contains the elements of $B{\downarrow}p$ maximal w.r.t. inclusion-based ordering.

A juror's cognitive state does not suffer only from inconsistency; it can also be affected by uncertainty. Numerical distributions of credibility over the sentences of L or over the set Ω of the models of L, play the same role that "epistemic entrenchment", p.e.r. and epistemic relevance play in the symbolic frameworks. Generally, numerical approaches do not respect logical dependencies among the sentences. Logics of uncertainty often represent a cognitive state K and the incoming information p in terms of their sets of models (also said "possible worlds"), respectively, $[K]$ and $[p]$. A cognitive state is

represented not simply by $[K]$, but by an assignment function $d(\omega) : \Omega \rightarrow [0, 1]$ such that $d(\omega^{'}) = 0$ for each $\omega^{'} \notin [K]$. The arrival of p generally means that the real world belongs to $[p]$. This event changes d into a new assignment (new prioritization) d'. Imposing the priority to the incoming information (AGM3) means assigning $d'(\omega) = 0$ to each $\omega \notin [p]$. Minimizing this change (AGM2) means minimizing some kind of distance between d and d'.

In the probabilistic approach (Pearl 1988) a cognitive state is characterized by a *probability measure* P on 2^{Ω}, whose fundamental property is *additivity*: $\forall A, B \subseteq \Omega, A \cap B \neq \oslash \Rightarrow P(A \cup B) = P(A) + P(B)$. $P(\Omega) = 1$, so if $\bar{A} = \Omega - A$ then $P(A) + P(\bar{A}) = 1$. We might also consider the *probability distribution* $pr(\omega)$ that assignes a probability degree to each world in Ω, where

$$P(A) = \sum_{\omega \in A} pr(\omega) \ .$$

$pr(\omega) = 0$ means that ω is not a possible world. $pr(\omega) = 1$ means that ω is surely the real world. An incoming information p changes the probability measure of any sentence q of L through the very famous Bayes' Conditioning Rule:

$$P(q|p) = \frac{P([q] \cap [p])}{P([p])} = \frac{P([p]|[q]) \cdot P([q])}{P([p])}$$

which can also be expressed in terms of probability distribution:

$$pr(\omega|p) = \begin{cases} \dfrac{pr(\omega)}{P([p])} & \text{if } \omega \in [p] \\ 0 & \text{otherwise} \end{cases}$$

This modification is defined only for $P([p]) > 0$, hence it is not applicable when p is judged impossible by the previously determined probability measure P. =

Bayesian conditioning obeys the principle of priority to incoming information (AGM3); it increases the probability of the not-impossible worlds belonging to $[p]$ to the prejudice of those external to $[p]$ which become all impossible. =

In the probabilistic framework the probability of a sentence p is simply the probability measure $P([p])$. Thus, probability measures order the sentences of L, but, unfortunately, they do not generate epistemic entrenchments. In effect, probability measures satisfy EE1 since they are, obviously, transitive (if $P([p]) \leq P([q])$ and $P([q]) \leq P([r])$ then $P([p]) \leq P([r])$. EE2 too is verified since $p \vdash q$ means $[p] \subseteq [q]$ hence $P([p]) \leq P([q])$ (it is always easier to retract p than q). Even EE4 is verified; in fact, $p \notin K$ means $P([p]) = 0$, and if K is consistent then there are sentences q such that $P([q]) = 0$, hence $P([p]) = 0$ iff $\forall q \in L$ $(P([p]) \leq P([q]))$. Finally, EE5 is verified since if $\forall q \in L$ $P([q]) \leq P([p])$ then $[p] = \Omega$ which means that p is a tautology. Unfortunately, EE3 is generally unsatisfied since $[p \wedge q] \subseteq [p]$ and $[p \wedge q] \subseteq [q]$ so that $P([p \wedge q]) \leq P([p])$ and $P([p \wedge q]) \leq P([q])$; normally it is easier to retract a conjunction than any of its conjuncts.

Also the belief function framework (Shafer 1990; Shafer & Srivastava 1990) assigns a probability P to the subsets of $= \Omega$, with the constraints $P(\oslash) = 0$ and $\sum_{A \subseteq \Omega} P(A) = 1$. If $P(A) > 0$ then A is said to be a *focal element*. The *belief function* on the subsets of Ω is defined as

$$Bel(A) = \sum_{X \subseteq A} P(X)$$

$Bel(A)$ measures the persuasion that the real world is inside A; maybe that there is no evidence that directly support A but it cannot be excluded because there is evidence that supports some of its subsets. This function is not additive: $Bel(A) + Bel(\bar{A}) \leq 1$. The knowledge is:

- *certain* and *precise* if there exists a $\omega \in \Omega$ such that $P(\{\omega\}) = 1$
- *certain* and *imprecise* as if there exists an $A \subset \Omega$ such that $P(A) = 1$ but A is not singleton
- *consistent* if all the focal elements are nested
- *inconsistent* if all the focal elements are disjoint
- *void* if $P(\Omega) = 1$ and for all $A \subseteq \Omega$, $P(A) = 0$.

This framework deals also with uncertain inputs. They are treated as new probability assignments on 2^Ω. The change consists of merging the two evidences (the prior P_1 and the new P_2) through the Dempster's Rule of Combination:

$$P(A) = \frac{\sum_{X_1 \cap X_2 = A} P_1(X_1) \cdot P_2(X_2)}{\sum_{X_1 \cap X_2 = \oslash} P_1(X_1) \cdot P_2(X_2)}$$

for all $A \subset \Omega$. This rule, easily extensible to combine n probability assignments, reinforces concordant evidence and weakens conflicting ones. It can be applied only if evidences are independent and referred to the same Ω. Because of the commutativity of the product, the rule is independent from the sequence $P_1 ... P_n$ so *it violates the principle of priority to the incoming information*! From a knowledge engineering point of view, the worst problem with the Dempster's Rule of Combination is its computational complexity. One should generate a frame of $2^{|\Omega|}$ elements to calculate it! However, much work has been spent in reducing the complexity of that rule. Such methods range from "efficient implementations" (Kennes 1992) to "qualitative approaches" (Parson 1994) through "approximate techniques" with statistical methods as the Montecarlo sampling algorithm (Wilson 1991; Moral & Wilson 1996).

3 Requirements for a Belief Revision Framework in a Multi Source Environment

We think that to revise beliefs in a Multi-Agent scenario, where many sources give information about a same static situation, the framework should satisfy some requisites.

- Ability to reject incoming information

Jurors should not obey the principle of "priority to the incoming information" which is not acceptable since there is no strict correlation between the chronology of the informative acts and the credibility of their contents (Dragoni, Mascaretti & Puliti 1995); it seems more reasonable to treat all the available pieces of information as they had been collected at the same time.

- Ability to recover previously discarded beliefs

Jurors should be able to recover previously discarded pieces of knowledge after that new evidence redeems them. The point is that this should be done not only when the new information directly "supports" a previously rejected belief, but also when the incoming information indirectly supports it, by disclaiming the beliefs that contradicted it, causing its ostracism. More formally, for each cognitive state K, and sentences p and q such that $K \vdash p$ and $K^*q \nvdash p$, there can always be another piece of information r such that $(K^*q)^*r \vdash p$, even if $r \nvdash p$. An obvious case should be $r = \neg q$. We elsewhere called this rule *principle of recoverability*: "any previously held piece of knowledge must belong to the current knowledge space if consistent with it" (Dragoni, Mascaretti & Puliti 1995; Dragoni 1997; Dragoni & Giorgini 1997a).

The rationale for this principle is that, if someone gave us a piece of information (sometime in the past) and currently there is no reason to reject it, then we should accept it! This is stronger than the traditional "coherence" spirit of belief revision, since the piece of knowledge to accept is not a *generic* sentence of the language but a *generated* piece of information; somewhere there is an utilitarian intelligent information source that guarantees for it. Of course, this principle does not hold for updating, where changes may be irrevocable. This feature could also be subtitled: "revocable treatment of consistency" . We remember of Minsky's lection: "I do not believe that consistency is necessary or even desirable in a developing intelligent system ... What is important is how one handles paradoxes or conflicts ... *Enforcing consistency produces limitations.* As we will see in a moment, we overcome this problem by defining a single global, never forgetting, eventually inconsistent *Knowledge Background*, upon which act multiple specific, competitive, ever changing, *consistent* cognitive states.

- Ability to *combine* contradictory and concomitant evidences

The notion of *beliefs integration* should blend that of revision (Dragoni & Giorgini 1997b). Every incoming information changes the cogniti= ve state. Rejecting the incoming information does not mean leaving beliefs unchanged since, in general, incoming information alters the distribution of the weights. Surely the last incoming information decreased the credibility of the beliefs with whom it got in contradiction, even in the case that it has been rejected. The same when receiving a piece of information which we were already aware of; it is not the case that nothing happened (as AGM K^*4 states) since we are now, in general, more sure about that belief. More generally, there is no reason to limit the changes introduced by the new information to an insertion into a pre-established relative order with consequent rearrangement of the ranking to accomplish the logical relations between beliefs (as Williams' transmutation does). If it is true that new incoming information affects the old one, it is likewise true that the latter affects the former. In fact, an autonomous agent (where "autonomous" means that his cognitive state is not *determined* by other agents) judges the credibility of new information on the basis of its previous cognitive state. "Revising beliefs" should simple mean "dealing with a new broader set of pieces of information".

- Ability to deal with couples <source,information> rather than with information alone

The way the credibility ordering is generated and revised must reflect the fact that beliefs come from different sources of information, since the reliability and the number of independent informants affect the credibility of the information and vice versa (Dragoni 1992).

- Ability to maintain and compare multiple candidate *cognitive states*

This ability is part of humans intelligence which does not limit its action to comparing single pieces of information but goes on trying to reconstruct alternative cognitive scenarios as far as it is possible.

- Sensibility of the syntax

Despite Dalal's (1988) aforementioned principle, syntax plays an important role in everyday life. The way we pack (and unpack) pieces of information reflects the way we organize thinking and judge credibility, importance, relevance and even truthfulness. A testimony of the form $\alpha\wedge\delta\wedge...\wedge\zeta\wedge\neg\alpha$ from a defendant A in a trial has the same semantic truth value than the testimony $\beta \wedge \neg\beta$ from defendant B, but we remember many cases in which B has been condemned while A has been absolved, being regarded his/her testimony "partially true", contrasting with the B's one regarded as "absolutely contradictory". A set of sentences seems not to be logically equivalent to their conjunction and we could change a cognitive state by simply clustering the same beliefs in a different way.

4 A Computable Model for Belief Revision

Our sentence-based approach for belief revision (Dragoni 1997) envisages two knowledge repositories:

1. the *knowledge background* KB, which is the set of all the propositional sentences available to the reasoning agent (as assumptions); it can be inconsistent

2. the *knowledge base* $B \subseteq KB$, which is the maximally consistent, currently preferred piece of knowledge that should be used for reasoning and decision supporting

Computationally, our way to belief revision consists of five steps (Dragoni & Giorgini 1997a,b):

S1. detection of the minimally inconsistent subsets of $KB \cup \{p\}$ (*nogoods*)
S2. generation of the maximally consistent subsets of $KB \cup \{p\}$ (*goods*)
S3. revision of the credibility weights of the sentences in $KB \cup \{p\}$
S4. choice of a preferred *good* as the new revised base B'

S5. selection of the derived sentences which are derivable from B'

The incoming information p, with its weight of evidence, is confronted not just within the current base B, but within the overall knowledge background KB. Doing so, the degrees of credibility of the sentences in $KB \cup \{p\}$ are reviewed on a broader and less prejudicial basis (S3). As already explained, the main advantage is that we can rescue sentences from KB by virtue of the maximal consistency of B'. If we'd revise only B by p, we could not recover information from KB. For instance, Nebel's revision would select some B'$\in B\Downarrow\neg p$, but it will be always possible to find out some B"$\in KB\Downarrow\neg p$ such that B'$\subseteq B$".

S4 might choose a new base B' syntactically equal to the previous B (meaning that p has been rejected) but, in general, B' will have a different credibility distribution than B. p might be rejected even if S4 chooses a base B' different from B, but that still containing sentences incompatible with p.

When p is consistent with B, not necessarily B'$=B\cup\{p\}$, since S3 may yield a totally different choice at S4. Previously rejected pieces of knowledge $R\subset KB$ can be rescued simply by determining some upsetting between the credibility of a set $S\subset B$ and the credibility of R, this may happen if p *supports* R against S. The rejection of the priority to the incoming information principle implies that K^*4 and and K^*5 hold no longer (if p is inconsistent it will be part of none of the goods produced at S2, so it will never be part of a base).

S1, S2 and S5 deal with consistency and derivation, and act on the symbolic part of the information. Operations are in ATMS style; to find out nogoods and goods, we adopt (and adapt) the most efficient set-covering algorithm that we are aware of Reiter (1987). Notwithstanding this, even in the propositional case, determining all the minimal inconsistencies can be very hard. However, such condition can be relaxed (the consequence is that some of the goods are not really consistent) and in practical applications dealing with commonsense knowledge (see e.g. Dragoni & Di Manzo 1995), such minimal inconsistencies could be provided interactively from the outside by the user.

S3 and S4 deal with uncertainty and work with the numerical weight of the information. Both contribute to the choice of the revised knowledge space so their reasonableness should be evaluated as a couple. Numerical formalisms are able to perform both of them since the credibility of a single sentence p is determined in the same way as the credibility of a set of sentences B by the weights attached to $[p]$ and $[B]$, respectively. Flexibility is an advantage in separating the two steps; for instance, depending on the characteristics of the knowledge domain under consideration and the kind of task and/or decision that should be taken on the basis of the revision outcome, the selection function could consider also one (or a combination) of the methods described in Benferhat et al. (1993).

Probabilistic methods with uncertain inputs seem inadequate for the strong dependence that they impose on the credibility of a sentence and that of its negation. We see that the belief-function formalism, in the special guise in which Shafer and Srivastava (1990) apply it to auditing, could work well because it treats all the pieces of information as they had been provided at the same time.

The method has the following I/O (see Dragoni & Giorgini 1997a):

Input:

list of pairs <source,piece of information>
list of pairs <source,reliability>

Output:

list of pairs <piece of information,credibility>
list of pairs <source,reliability>

Let S=$\{s_1, ..., s_n\}$ be the set of the sources, and let kb_i be the subset of KB received from s_i. Each source s_i is associated with a *reliability* $R(s_i)$, that is regarded as the *probability* that the source is faithful. The main idea with this multi-source version of the belief function framework is that a reliable source cannot give false information, while an unreliable source can give correct information; the hypothesis that s_i is reliable is compatible only with the models of kb_i, while the hypothesis that s_i is unreliable is compatible with the overall Ω. Each source s_i is an evidence for KB and generates the following *bpa* $m_i(\cdot)$ on 2^Ω:

$$m_i(X) = \begin{cases} R(s_i) & \text{if } X = [kb_i] \\ 1 - R(s_i) & \text{if } X = \Omega \\ 0 & \text{otherwise} \end{cases}$$

All these *bpas* will be then combined through the Dempster Rule of Combination. From the combined *bpa* $m(\cdot)$, the credibility of a sentence p of L is given, as usual, by:

$$Bel(p) = \sum_{X\subseteq[p]} m(X)$$

From this mechanism we obtained an easy way to calculate the new reliabilities of the sources. Let Φ be an element of 2^S. If the sources are independent, the reliability of Φ is

$$R(\Phi) = \prod_{S\in\Phi} R(s) \cdot \prod_{S\in\Phi} (1 - R(s))$$

It holds that

$$\sum_{\Phi\in 2^S} R(\Phi) = 1$$

It may be that some source fall in contradiction, so that some elements of 2^S are impossible. The remaining elements are subjected to Bayesian conditioning so that their reliabilities sum up again to 1. The revised reliability $R^*(s)$ of a source s is the sum of the new reliabilities of the surviving elements of 2^S that contain s. If a source has been involved in some contradictions, then $R^*(s) \leq R(s)$, otherwise $R^*(s) = R(s)$.

S4 translates such ordering on the *sentences* in $KB \cup \{p\}$ into an ordering on the *goods* of $KB \cup \{p\}$. The best classified good is selected as the preferred revised knowledge base. If the ordering on $KB \cup \{p\}$ is not strict, then there can be multiple preferred goods. In this case we could take their intersection as revised knowledge base (Benferhat et al. 1993). Yet, the intersection is not maximally consistent and this means that all the conflicting pieces of knowledge with the same credibility will be rejected.

Another question is: S4 should consider only the *qualitative* ordering of the sentences in $KB \cup \{p\}$ (relative classification without the numerical weights) or could it take advantage of the *explicit* ordering (numerical weights). The first approach seems closer to the human cognitive behavior (which normally refrains from numerical calculus). The second one seems more informative (it takes into account not only relative positions but also the gaps between the items). In our model we do not use the "best-out" ordering for its "drowning effect" (Benferhat et al. 1993). The lexicographic one could be justified in some particular application domains (e.g. diagnosys). The inclusion-based method seems the most reasonable since it eliminates always the least credible one among conflicting pieces of knowledge.

As an example of a numerical way to perform S4, ordering the goods according to their average credibility seems reasonable and easy to calculate. With this method the preferred good may not contain the most credible sentence.

In the belief function framework, a "good" g is an element of Ω, precisely the one in which all the sentences in g are considered "true" and all the sentences out of g are considered "false" . This implies that the belief-function formalism is able to attach directly a degree of credibility to g, bypassing S4 in our framework. Unfortunately, when a good contains only part of the information supplied by a source, the belief-function formalism puts at zero its degree of credibility. This is unreasonable and, unluckily, the event is all but infrequent, so that often the credibility of all the goods is null.

A final step in our revision mechanism is the selection of the derived sentences which are still derivable from B' since the assumptions on which they rely are all contained in B'. Theoretically, it simply consists in applying classical entailment on the preferred good to deduce plausible conclusion from it. We adopted an ATMS and we stored each sentence derived by the Theorem Prover with an *origin set* (Martins & Shapiro 1988), i.e., a set of basic assumptions which are all *necessary* to derive it. Practically, this step consists in selecting from the derived sentences, all those whose origin set is subset of the preferred good. We could relax the definition of origin set to that of a set of basic assumptions used to derive the sentence. This is easier to compute and does not have harmful consequences; the worst it can happen is that, being this relaxed origin set a superset of the real one, it is not certain that it will be a subset of the preferred good as the real one is, and so some derived logical consequences of the preferred good may be not recognized (at first).

Besides recoverability, this computational model for belief revision overcomes various limitations of other classic approaches, in particular:

- the revision can be iterated
- inconsistent incoming information does not yield inconsistent revised knowledge spaces
- the numerical revision is performed on a broader base (the overall KB)
- the revision is more flexible;
- the complete numerical ordering renders the revision as least drastic as possible
- the splitting between the symbolic treatment of the inconsistencies and the numerical revision of the credibility weights, provides a clear understanding of what is going on and lucid explanations for the choices.

Dragoni and Giorgini are currently applying this conception of belief revision in a distributed monitoring system (Dragoni & Giorgini 1998) and in the police inquiry domain (Dragoni, Ceresi & Pasquali 1996).

Within jury research, such a model of deliberative negotiation on opinion are not to be adopted "as is", as the model is likely to require fine-tuning to the specifics of trial contexts, let alone taking account of the exclusionary rules of evidence as reflected in the judge's instructions to the jury. Yet, arguably we have here an important approach that could eventually stand at least on a par with the approaches (especially the probabilistic or stochastic ones) represented in Hastie's volume.

References

C.E. Alchourrón, P. Gärdenfors, and D. Makinson (1985) "On the logic of theory change: Partial meet contraction and revision functions". *The Journal of Symbolic Logic*, 50: pp. 510–530.

R. Allen and M. Redmayne, eds. (1997) *Bayesianism and Juridical Proof.* Special issue of *The International Journal of Evidence and Proof,* 1. London: Blackstone. [Vol. 1 includes 4 regular, numbered issues and one thematic issue, unnumbered.]

S. Benferhat, C. Cayrol, D. Dubois D., J. Lang, and H. Prade (1993) "Inconsistency management and prioritized syntax-based entailment". *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, pp. 640–645.

M. Dalal (1988) "Investigations into a theory of knowledge base revision". *Proceedings of the 7th National Conference on Artificial Intelligence*, pp. 475–479.

A.F. Dragoni (1992) "A model for belief revision in a multi-agent environment". In: Werner E. and Demazeau Y., eds., *Decentralized AI, 3.* Amsterdam: North Holland / Elsevier Science.

A.F. Dragoni (1997) "Belief revision: from theory to practice". *The Knowledge Engineering Review*, 12(2).

A.F. Dragoni, C. Ceresi, and V. Pasquali (1996) "A system to support complex detective inquiries". *Proceedings of the Fifth Iberoamerican Conference on Computer Science and Law*, La Habana.

A.F. Dragoni and P. Giorgini (1997a) "Belief revision through the belief function formalism in a multi-agent environment". In: Wooldridge M., Jennings N.R., and Muller J., eds., *Intelligent Agents, III* (LNCS, vol. 1193.) Heidelberg: Springer-Verlag.

A.F. Dragoni and P. Giorgini (1997b) "Distributed knowledge revision-integration". *Proceedings of the Sixth ACM International Conference on Information Technlogy and Management* . New York: ACM Press.

A.F. Dragoni and P. Giorgini (1998) "Sensor data validation for nuclear power plants through bayesian conditioning and dempster's rule of combination". *Computers and Artificial Intelligence*, 17(2/3): pp. 151–168.

A.F. Dragoni and M. Di Manzo (1995) "Supporting complex inquiries". *International Journal of Intelligent Systems*, 10: pp. 959–986.

A.F. Dragoni, F. Mascaretti, and P. Puliti (1995) "A generalized approach to consistency-based belief revision". In: M. Gori and G. Soda, eds., *Proc. of the Conference of the Italian Association for Artificial Intelligence* (LNAI, vol. 992.) Heidelberg: Springer Verlag.

D. Dubois and H. Prade (1992) "Belief change and possibility theory". In: P. Gärdenfors, ed., *Belief Revision.* Cambridge University Press.

D.M. Gaines (1994) "Juror Simulation". BSc Project Report (Proj. No. CS-DCB-9320), Computer Science Dept., Worcester Polytechnic Institute.

D.M. Gaines, D.C. Brown and J.K. Doyle (1996) "A Computer Simulation Model of Juror Decision Making". *Expert Systems With Applications* 11(1): pp. 13–28.

P. Gärdenfors (1988) *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* Cambridge, MA: MIT Press.

R. Hastie, ed. (1993) *Inside the Juror: The Psychology of Juror Decision Making* (Cambridge Series on Judgment and Decision Making.) Cambridge, U.K.: Cambridge University Press, 1993 (hc), 1994 (pb).

R. Kennes (1992) "Computational aspects of the Möbius transform of a graph". *IEEE Transactions in Systems, Man and Cybernetics*, 22: pp. 201–223.

J. De Kleer (1986) "An assumption based truth maintenance system". *Artificial Intelligence*, 28: pp. 127–162.

J.P. Martins and S.C. Shapiro (1988) "A model for belief revision". *Artificial Intelligence*, 35: pp. 25–97.

S. Moral and N. Wilson (1996) "Importance sampling monte-carlo algorithms for calculation of Dempster-Shafer belief". *Proceedings of IPMU'96,* Granada, Spain.

B. Nebel (1994) "Base revision operations and schemes: Semantics, representation, and complexity". In: A.G. Cohn, ed., *Proceedings of the 11th European Conference on Artificial Intelligence.* John Wiley & Sons.

S. Parson (1994) "Some qualitative approaches to applying the Dempster-Shafer theory". *Information and Decision Technologies*, 19: pp. 321–337.

J. Pearl (1988) *Probabilistic Reasoning for Intelligent Systems.* San Mateo, CA: Morgan Kaufmann.

R. Reiter (1987) "A theory of diagnosis from first principles". *Artificial Intelligence*, 53.

G. Shafer (1990) "Belief functions". In: G. Shafer and J. Pearl, eds., *Readings in Uncertain Reasoning.* San Mateo, CA: Morgan Kaufmann.

G. Shafer and R. Srivastava (1990) "The Bayesian and belief-function formalisms: a general perspective for auditing". In: G. Shafer and J. Pearl, eds., *Readings in Uncertain Reasoning.* San Mateo, CA: Morgan Kaufmann.

N. Wilson (1991) "A Monte-Carlo algorithm for Dempster-Shafer belief". In: P. Smets, B.D. D'Ambrosio and P. Bonissone, eds., *Proceedings of the Seventh Conference,* pp. 414–417.