REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Pu[illegible] ma[illegible] 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and [illegible]ation. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE February 1993	3. REPORT TYPE AND DATES COVERED Professional Paper
4. TITLE AND SUBTITLE THE ROLE OF CONDITIONAL EVENT ALGEBRA IN THE MODELING OF C^3 SYSTEMS		5. FUNDING NUMBERS PR: ZW40 PE: 0601152N WU: DN300173
6. AUTHOR(S) I. R. Goodman		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Command, Control and Ocean Surveillance Center (NCCOSC) RDT&E Division San Diego, CA 92152-5001		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Chief of Naval Research Independent Research Programs (IR) OCNR-10P Arlington, VA 22217-5000		10. SPONSORING/MONITORING AGENCY REPORT NUMBER
11. SUPPLEMENTARY NOTES		
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.		12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This paper highlights the newly-emerging discipline of conditional event algebra, showing how this area can be of use in combining conditional evidence in a C^3 setting, compatible with all possible conditional probability evaluations. Typically, such information arises from many different sources and has widely varying antecedents. Recent breakthroughs are underscored, including: (1) derivation of a number of new properties pointing to a particular conditional event algebra (GNW) as being one of the leading candidates of choice; (2) full development of non-boolean operators acting upon conditional events, including cartesian products and inverse functions, resulting in a fully-developed model of joint conditional rand variables. In addition, a recent uniqueness of representation of fuzzy sets by random sets is shown for both the unconditional and conditional cases. Finally, a previously introduced diagonalization technique for combining different stochastic information is further extended and justified, showing how multi-source linguistic-based and stochastic, conditional and unconditional, information can be combined.

Published in *Proceedings* 1992 Symposium C^2 Research, June 1992.

14. SUBJECT TERMS data fusion conditional events syntactics semantics linguistic information			15. NUMBER OF PAGES
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT SAME AS REPORT

UNCLASSIFIED

21a. NAME OF RESPONSIBLE INDIVIDUAL	21b. TELEPHONE (Include Area Code)	21c. OFFICE SYMBOL
I. R. Goodman	(619) 553-4014	Code 421

DTIC QUALITY INSPECTED 4

Accession For		
NTIS GRA&I	☑	
DTIC TAB	☐	
Unannounced	☐	
Justification		
By		
Distribution/		
Availability Codes		
Dist	Avail and/or Special	
A-1	20	

NSN 7540-01-280-5500

Standard form 298 (BACK)

UNCLASSIFIED

THE ROLE OF CONDITIONAL EVENT ALGEBRA IN THE MODELING OF C^3 SYSTEMS

Dr. I. R. Goodman

Code 421, NRaD,
Naval Command, Control
and Ocean Surveillance Center,
Research, Development, Test and
Evaluation Division,
San Diego, CA 92152-5000

Paper accepted for presentation at the 1992 Symposium on Command and Control Research, Unclassified, to be held at Naval Postgraduate School, Monterey, CA, June 8-11, 1992, sponsored by The Technical Panel for C^3, Basic Research Group, Joint Directors of Laboratories

THE ROLE OF CONDITIONAL EVENT ALGEBRA IN THE MODELING OF C^3 SYSTEMS

I.R. Goodman
Code 421, NRaD,
Naval Command, Control
and Ocean Surveillance Center,
Research, Development, Test and
Evaluation Division
San Diego, CA 92152-5000

Abstract

This paper highlights the newly-emerging discipline of conditional event algebra, showing how this area can be of use in combining conditional evidence in a C^3 setting, compatible with all possible conditional probability evaluations. Typically, such information arises from many different sources and has widely varying antecedents. Recent breakthroughs are underscored, including: (1) derivation of a number of new properties pointing to a particular conditional event algebra (GNW) as being one of the leading candidates of choice; (2) full development of non-boolean operators acting upon conditional events, including cartesian products and inverse functions, resulting in a fully-developed model of joint conditional random variables. In addition, a recent uniqueness of representation of fuzzy sets by random sets is shown for both the unconditional and conditional cases. Finally, a previously introduced diagonalization technique for combining different stochastic information is further extended and justified, showing how multi-source linguistic-based and stochastic, conditional and unconditional, information can be combined.

1 General Introduction

C^3 systems reflect, to a large extent, everyday real-world problems - in addition to the obvious pressing (possibly life-threatening) decisions entailed in a typical military scenario. Thus, in attempting to model and better understand C^3 systems, we also enhance our own daily perspectives on complex decision processes. The philosophy of approach taken here is that one needs first to develop a comprehensive mathematical model of the problem, while at the same time understanding its empirical aspects. The model should also - emphatically! - be as natural as possible, tie in with previously established results, and be derivable from a minimum of, if any, ad hoc assumptions.

This paper continues the ongoing effort of the author and others in attempting to deal with the data fusion aspect of C^3 systems. Many seemingly simple problems of combination of evidence belie much deeper mathematical and logical issues which must be addressed. For example, consider the job of a command unit in synthesizing the following information, a week before a given time deadline, the size of enemy ground forces ready for combat:

Source 1. "I (John Doe) hope that (based upon my expertise) the size is quite small - perhaps 8000 or less."

Source 2. Intelligence source Q, based on high technology devices, provides a known probability function over the interval [7000, 9000], in units of 200 troops.

Source 3. According to combat rule book R_1: "If weather condition b holds night before, then troop size a can hold", where a and b can vary jointly in sets A and B, respectively. a and b have known - updated to command unit's time - joint probabilities of occurrences. Typically, b could be the combination of factors "full moon, damp, between 60° and 70°", while a could be "between 11000 and 12000 troops".

Source 4. According to combat rule book R_2: "If political condition d holds, then troop size c can hold", where $c \in C$, $d \in D$, with known (updated) joint probabilities of occurrences, analogous to that in Source 3.

By a concensus of experts, joint probabilities of occurrences of all a,b,c,d can be estimated for Sources 3 and 4.

Clearly, Source 1 is more linguistic in nature, involving the non-indicative modal form ..hope.. (a bouletic form). Source 2 is readily seen to be essentially characterized by a single ordinary probability distribution. Sources 3 and 4, most naturally, represent tables of possible conditional probability relations. Later, it will be seen that such information sources, despite their different structures, can be reasonably treated for fusion of data.

In short, this paper continues - and extends further - previous attempts at understanding examples such as above, and is aimed at establishing a rigorous theory of data fusion that can be applied universally to all such problems. More specifically, this author has endeavored over the past several years to address the following fundamental aspect of C^3 modeling:

How can one establish a cohesive, universal, but feasible, theory for fusing multiple source, disparate information, which can be in any of several forms, including any combination of the following: narrative / linguistic / human source-oriented; stochastic / mechanical sensor-oriented; conditional or unconditional; indicative or properly modal?

In particular, in [1]-[12] an attempt has been made to treat this problem. This approach necessitated the development of new "mathematical tools" and the enhancement of established ones in dealing with these difficult issues. These included:

(1) The modeling of conditional stochastic type information and the related concept of "conditional events" and their logical relations, especially for the situations where they have differing antecedents, through *conditional event algebra*. (For background, see the monographs [13],[14] and the survey paper [15].)

(2) The modeling of natural language information via fuzzy set theory and the *one point coverage function representation by random sets*. (For background, see, e.g. [16].)

(3) The establishment of a technique for combining any of several stochastic descriptions (distributions, for example) of a common unknown parameter of interest, extending - and compatible with - classical estimation and regression theory, namely: *diagonalization*. (For background, see [3].)

(4) Derivations of procedures which utilize aspects of both (1) and (2):

(I) Conditioning of linguistic-based information naturally modeled through *conditioning of fuzzy sets*. (See [17]-[18] for a previous direction of analysis.)

(II) Modeling of *modal / non-indicative linguistic information*, utilizing conditional event representations and conditional event algebra or conditional fuzzy sets. (See [19].)

In addition, recent work should be mentioned on the development of comparison criteria via game theory for help in choosing among competing uncertainty functions, for a given situation, the most appropriate. This includes the class of all probability functions, possibility functions, and Dempster-Shafer functions, among others. (See [50].)

A good summary of the above relatively new areas of investigation was provided in [12]. More recently, several additional breakthroughs have been obtained in these directions and form the focus of this paper. In turn, these results have allowed the closing of a number of gaps in the theory of data fusion and consequently have lessened the use of ad hoc constructs and assumptions. Overviews of these breakthroughs, followed by concise technical details, are provided in the next sections. Finally, the last section provides a procedure for implementing them, suitable for real-world - and in particular - C^3 system usage.

2 Conditional Event Algebra: Introduction

Conditional event algebra has been developed in response to the current lack in the standard literature of a full algebra or syntax of logical operations and relations which play a role relative to conditional probability evaluqtions, as the way now that boolean algebra plays with respect to all unconditional probabilities. In short, it is clear that one can use boolean algebra to manipulate and help in evaluations such as

$$p(((a \vee ab)' \vee c')' \cdot c \vee ac')$$
$$= p((a'\vee c')'c \vee ac') \quad \text{(by absorption)}$$
$$= p(acc \vee ac') \quad \text{(by DeMorgan)}$$
$$= p(ac \vee ac') \quad \text{(by idempotence)}$$
$$= p(a(c\vee c')) \quad \text{(by distributivity)}$$
$$= p(a) \,. \quad \text{(by orthocomplementation)} \quad (2.1)$$

Here, R is some boolean algebra of events a,b,c,d,... with $\cdot$ or its omission between symbols denoting conjunction or intersection, $\vee$ denoting disjunction or union, $(\ )'$ denoting negation or complementation, 0 being the zero or null ($\emptyset$), 1 being the unit or universal(Ω, etc.), $\leq$ being the basic lattice order or subset relation ($\subseteq$), and $p:R \to [0,1]$ being some probability measure over R. (This notation will be employed throughout the paper.)

On the other hand, suppose one wished to evaluate the conjunction of two *conditional* sentences, such as the ones given in the example of the introduction representing Sources 3 and 4. Symbolically, we write $p((a|b)\cdot(c|d))$ as the desired evaluation, where we assume the conditional probability compatibility relations

$$p(a|b)=p(\text{"if b then a"}) \;;\; p(c|d)=p(\text{"if d then c"}), \quad (2.2)$$

with

$$p(a|b) \overset{d}{=} p(ab)/p(b);\; p(c|d) \overset{d}{=} p(cd)/p(d), \quad (2.3)$$

the usual conditional probability definitions, assuming of course that $p(b),p(d) > 0$.

Is there some computable "conditional event" $(\alpha|\beta)$, $\alpha=\alpha(a,b,c,d)\in R$, $\beta=\beta(a,b,c,d)\in R$ such that

$$(\alpha|\beta) = (a|b)\cdot(c|d) \quad (2.4)$$

has meaning, as does the extension of conjunction, also denoted as $\cdot$ on "conditional events", with the compatible evaluation

$$p((\alpha|\beta)) = p(\alpha|\beta) \quad ? \quad (2.5)$$

Or, is there a similar $\gamma,\delta \in R$, such that

$$(\gamma|\delta) = (a|b)\vee(c|d) \quad (2.6)$$

has meaning and

$$p((\gamma|\delta)) = p(\gamma|\delta) \quad ? \quad (2.6')$$

Clearly, if b=d, then a natural solution, fully compatible with traditional probability concepts is to choose in (2.4),(2.5), α=ac and β=b, yielding

$$(ac|b) = (a|b)\cdot(c|b) \quad (2.4')$$

and

$$p((ac|b)) = p(ac|b), \quad (2.5')$$

with the interpretation that $(a|b)$, $(c|b)$, $(ac|b)$ are simply a,c,ac, respectively, restricted to b, i.e., ab,cb,acb, with initial R replaced by $Rb \overset{d}{=} (xb: x\in R)$ and $p:R \to [0,1]$ by the conditional probability measure

$p_b \overset{d}{=} p(\cdot|b):Rb \to [0,1]$, etc. But, the general case where $b \neq d$ cannot be treated this way.

In order to answer the above questions, one first must be able to identify just what we mean by a "conditional event". Since $a,b \in R$, it is natural to inquire whether $(a|b) \in R$ or in some enlargement of R, which is still a boolean algebra (as Copeland and others originally thought [20], [21]). The only natural candidate for $(\ |\ ):R^2 \to R$ playing the role of a conditional event former is material implication $\Rightarrow$, which is actually also the relative pseudocomplement operation for R, with form

$$b \Rightarrow a \overset{d}{=} b' \vee a = b' \vee ab \ . \qquad (2.7)$$

However, Calabrese [22] noted as early as 1975 that $\Rightarrow$ cannot be $(\ |\ )$ and pointed out (also later in 1987 [23]) the inequality

$$p(b \Rightarrow a) = 1-p(b)+p(ab) = p(a|b) + p((b')\cdot p(a'|b) \geq p(a|b) \ , \qquad (2.8)$$

with strict inequality holding in general except for trivial cases. In addition, he also proposed the question whether conditional events could be constructed outside of R, but gave back the compatible evaluations

$$p((a|b)) = p(a|b) \ . \qquad (2.8')$$

Indeed, in [23] Calabrese demonstrated that no binary boolean operation of any of the 16 possible can be used to produce conditional events. (See also Popper's even earlier related results [24].)

Even more negative than this - and unknown to Calabrese - David Lewis in 1973 (but published in 1976 [25]) showed that *no* binary (and for that matter, n-ary) function of any kind $f:R^2 \to R$ exists such that for all prob. meas. $p:R \to [0,1]$,

$$p(f(a,b)) = p(a|b), \text{ all } a,b \in R,\ p(b) > 0. \qquad (2.9)$$

(This, thus negated Copeland's attempts, mentioned earlier.)

However, on the positive side of things, Calabrese [23], independent of Lewis, demonstrated that by going to $P(R)$ (power class of R), one could obtain well-defined conditional events, and in fact,he postulated an entire calculus of operations extending the ordinary boolean ones. Earlier, Schay [26] independently also demonstrated the existence of conditional events based upon three-valued indicator functions, extending the usual two-valued ones for ordinary unconditional sets. Furthermore, Adams [27], independent of Schay and Calabrese, though not identifying what a conditional event should mean,did define the very same operations Calabrese proposed, as well as certain of Schay's! (Schay actually proposed two sets of operations for possible candidates and one definition for conjunction from one set and the other for disjunction coincide with Adams and Calabrese' common definitions.) Hailperin [28] also independently investigated conditional events, rigorizing certain of Boole's ideas on "division of events" [29], through the use of Chevalley-Uzkov algebraic fractions. But, Hailperin did not develop any nontrivial opertaions, other than common antecedent ones. He did show his "fractions" were actually isomorphic to principal ideal cosets of R, the form conditional events must take (see Theorem 2.1 later in this section). Goodman [30] showed Calabrese' conditional events were also principal ideal cosets and proposed for the first time *deriving* conditional event operations from first principal considerations, namely, via functional image extensions of the usual unconditional boolean operations $\cdot$,v, ( )', etc. This yielded closed computable forms, later extended in [13], section 3.2. Furthermore, Goodman, Nguyen, and Walker([13], chapter 2; see also Theorem 2.1 here) derived the form that conditional events *must take*, under a minimum of assumptions. For a thorough history of the problem of defining conditional events and their operations see [15]. In [13] a number of important properties of conditional event algebra,in general, and GNW (Goodman-Nguyen-Walker) conditional event algebra, in particular, are established, including tie-ins with three-valued logic and Koopman's qualitative conditional probability.

In summary, the most significant elementary results in the development of conditional event algebra are:

For any given boolean algebra R, define the space

$$(R|R) \overset{d}{=} \{(a|b):a,b \in R\} \subseteq P(R),$$

$$\begin{aligned}(a|b) &\overset{d}{=} Rb' \vee a = Rb' \vee ab \\ &= \{xb' \vee ab : x \in R\} \\ &= \{x : x \in R \ \&\ xb = ab\} \\ &= [ab, b \Rightarrow a] \overset{d}{=} \{x : x \in R \ \&\ ab \leq x \leq b \Rightarrow a\} \in P(R),\end{aligned} \qquad (2.9)$$

principal ideal coset of R with antecedent b and consequent or residue ab.

Conversely, given any $\alpha \leq \beta \in R$, the *closed interval of events*

$$[\alpha,\beta] \overset{d}{=} \{x : x \in R \ \&\ \alpha \leq x \leq \beta\} = (\alpha|\beta \Rightarrow \alpha) \in (R|R), \qquad (2.10)$$

so that also

$$(R|R) = \{[\alpha,\beta] : \alpha \leq \beta \in R\}. \qquad (2.11)$$

Also, note for any $a,b,c,d \in R$,

$$(a|1) = a \ , \qquad (2.12)$$

identifying singletons with their elements, yielding

$$R \subseteq (R|R) \subseteq P(R) \ ; \qquad (2.13)$$

$$(a|b) = (ab|b) \ ; \qquad (2.14)$$

$$\begin{aligned}(a|b) = (c|d) &\text{ iff } ab = cd \ \&\ b \Rightarrow a = d \Rightarrow c \\ &\text{ iff } ab = cd \ \&\ b = d \ ;\end{aligned} \qquad (2.15)$$

$$(a|0) = (0|0) = R \ ; \qquad (2.16)$$

$$(0|b) = (b'|b) = Rb' = [0,b'];\ (1|b) = (b|b) = Rb' \vee b = Rvb = [b,1]. \qquad (2.17)$$

That (R|R) should be chosen as the conditional event extension of R and each (a|b) as a conditional event with antecedent b and consequent a or as "if b then a" or "a given b", is justified by :

<u>Theorem</u> 2.1 ([13], chapter 2)

Let R be any boolean algebra, $S \supseteq R$ any space and

$f:R^2 \to S$ any surjective function such that for all prob. meas. $p:R \to [0,1]$, p can be extended in a well-defined way to, using same symbol, $p:S \to [0,1]$ such that for all $a,b \in R$, with $p(b) > 0$,

$$p(f(a,b)) = p(a|b). \tag{2.18}$$

Suppose also that for all $a,b,c,d \in R$,

$$\left.\begin{array}{l} f(a,b) = f(ab,b) \\ \text{and} \quad f(a,b) = f(c,d) \text{ implies } b = d \end{array}\right\}. \tag{2.19}$$

Then, there is a bijection $\tau:S \to (R|R)$ such that for all $a,b \in R$,

$$\tau(f(a,b)) = (a|b). \tag{2.20}$$

■

Next, for any boolean algebra $R \subseteq P(D)$, for some set D (by the Stone Representation Theorem), define associated 3-valued indicator function transform $\phi:(R|R) \to \{0,u,1\}^D$, where

$$0 \le u \le 1 \tag{2.20'}$$

represents an indeterminate value: For any $a,b \in R$, and hence any $(a|b) \in (R|R)$, $\phi(a|b):D \to \{0,u,1\}$ is given by

$$\phi(a|b)(x) = \begin{cases} 1, & \text{if } x \in ab, \\ 0, & \text{if } x \in a'b, \\ u, & \text{if } x \in b'. \end{cases} \tag{2.21}$$

Note also that

$$\{g: g:D \to [0,1] \ \& \ g^{-1}(i) \in R, i \in \{0,u,1\}\} = \{\phi(a|b): a,b \in R\}, \tag{2.22}$$

extending the classical indicator-set relations.

Theorem 2.2 (remark, [13], p.31)

Let $R \subseteq P(D)$ be a boolean algebra. Then, there is a natural bijection between $\{\phi(a|b): a,b \in R\}$ and $(R|R)$ via ϕ.

■

The GNW conditional event algebra operations are summarized as follows:

Theorem 2.3 ([13], Theorem 2, p. 62 extended)

Let R be any boolean algebra. Then, the functional image extensions of the usual boolean operations $\cdot, v, (\ )', \Rightarrow, \Leftrightarrow, +$ are, for any $a,b,c,d \in R$

(i)

$$(a|b)\cdot(c|d) \overset{d}{=} \{x\cdot y: x \in (a|b), y \in (c|d)\} = (abcd|Q_\cdot) \in (R|R), \tag{2.23}$$

where

$$Q_\cdot \overset{d}{=} a'b \vee c'd \vee bd = a'b \vee c'd \vee abcd$$
$$= a'bd' \vee c'db' \vee bd = (b \vee c'd)(d \vee a'b), \tag{2.24}$$

$$(a|b)v(c|d) \overset{d}{=} \{x v y: x \in (a|b), y \in (c|d)\} = (ab \vee cd\,|Q_v) \in (R|R), \tag{2.25}$$

where

$$Q_v \overset{d}{=} ab \vee cd \vee bd = ab \vee cd \vee a'bc'd$$
$$= abd' \vee cdb' \vee bd = (b \vee cd)\,(d \vee ab), \tag{2.26}$$

$$(a|b)' \overset{d}{=} \{x':x \in (a\ b)\} = (a'|b) = (a'b|b), \tag{2.27}$$

$$(c|d) \Rightarrow (a|b) \overset{d}{=} \{x \Rightarrow y : x \in (c|d), y \in (a|b)\}$$
$$= (c|d)' \vee (a|b) = (c'd \vee ab|c'd \vee ab \vee bd), \tag{2.28}$$

$$(a|b) \Leftrightarrow (c|d) \overset{d}{=} \{x \Leftrightarrow y: x \in (a|b), y \in (c|d)\}$$
$$= ((c|d) \Rightarrow (a|b))\cdot((a|b) \Rightarrow (c|d))$$
$$= (a|b)'\cdot(c|d)' \vee (a|b)\cdot(c|d)$$
$$= (ab \Leftrightarrow cd|bd), \tag{2.29}$$

$$(a|b) + (c|d) \overset{d}{=} \{x+y: x \in (a|b), y \in (c|d)\}$$
$$= (ab + cd|bd)$$
$$= ((a|b) \Leftrightarrow (c|d))' \tag{2.30}$$

(ii) For all $a_j, b_j \in R$,

$$(a_1|b_1)\cdot\cdots\cdot(a_n|b_n) = (a_1b_1 \cdots a_nb_n|Q_{\cdot,n}), \tag{2.31}$$

$$(a_1|b_1)v\cdots v(a_n|b_n) = (a_1b_1 \vee \cdots \vee a_nb_n|Q_{v,n}), \tag{2.32}$$

$$Q_{\cdot,n} \overset{d}{=} a_1'b_1 \vee \cdots \vee a_n'b_n \vee b_1\cdots\cdots b_n, \tag{2.33}$$

$$Q_{v,n} \overset{d}{=} a_1b_1 \vee \cdots \vee a_nb_n \vee b_1\cdots\cdots b_n. \tag{2.34}$$

(iii) Define for all $a,b,c,d \in R$,

$$(a|b) \le (c|d) \text{ iff } (a|b) = (a|b)\cdot(c|d). \tag{2.35}$$

Then,

$(a|b) \le (c|d)$ iff $(c|d) = (a|b)v(c|d)$ iff

$ab \le cd$ & $b \Rightarrow a \le d \Rightarrow c$ iff $ab \le cd$ & $c'd \le a'b$ iff

the corresponding corners of $[ab, b \Rightarrow a] \le$
" " " " " $[cd, d \Rightarrow c]$ (2.36)

(iv) Bayes' theorem; chaining relations hold, etc.:

$$(ac|b) = (a|cb)\cdot(c|b). \tag{2.36'}$$

■

The common SAC (Schay-Adams-Calabrese) extensions of unconditional boolean operations are defined as:

$$(a|b)^{\circlearrowleft} \overset{d}{=} (a'|b) = (a|b)', \tag{2.37}$$

$$(a|b) \textcircled{v} (c|d) \overset{d}{=} (ab \vee cd|b \vee d), \tag{2.38}$$

$$(a|b) \odot (c|d) \overset{d}{=} ((a|b)' \textcircled{v} (c|d)')'$$
$$= ((a'b \vee c'd)'|b \vee d)$$
$$= ((b \Rightarrow a)(d \Rightarrow c)|b \vee d). \tag{2.39}$$

In addition, it should be noted that the two conditional event algebras proposed by Schay are

$$(a|b)'^{1} \overset{d}{=} (a\ b)'^{2} \overset{d}{=} (a|b)', \tag{2.40}$$

$$(a|b) \textcircled{v}_1 (c|d) \overset{d}{=} (a|b) \vee (c|d), \tag{2.41}$$

$$(a|b) \odot_1 (c|d) \overset{d}{=} (abcd|bd), \tag{2.42}$$

$$(a|b) \textcircled{v}_2 (c|d) \overset{d}{=} (ab \vee cd|bd), \tag{2.43}$$

$$(a|b) \odot_2 (c|d) \overset{d}{=} (a|b) \odot (c|d). \tag{2.44}$$

3 Conditional Event Algebra: Key Past Results and Some Recent Developments

It is obvious that even restricting possible choices to only GNW and SAC algebras, there is still a wide discrpency in forms and evaluations relative to conjunction and disjunction. One advantage for using SAC is the "smoothing" property when antecedents are disjoint :

For $b\cdot d = 0$, and hence $abcd = 0$,

$$(a|b) \odot (c|d) = (a|b) \textcircled{v} (c|d) = (ab \vee cd|b \vee d), \tag{3.1}$$

while on the other hand GNW yields the trivial results

$$(a|b)\cdot(c|d) = (0|a'b \vee c'd), \tag{3.2}$$

$$(a|b)v(c|d) = (ab \vee cd\,|\,ab \vee cd), \tag{3.3}$$

which are obviously zero and unity in value for all well-defined probability evaluations. On the other hand, note the following (GNW)

$$ab \vee (d|d) = (ab \vee d | ab \vee d) \ , \qquad (3.4)$$

with obvious unit probability evaluations, compatible with the monotonic increasing property of v and the fact that (d|d) is also an obviously unit probability event, but for SAC the same expression becomes

$$ab \circledvee (d|d) = ab \vee d \ , \qquad (3.5)$$

which if a probability measure p is chosen so that ab is small in value as is d, then despite (d|d) being a unity event, $p(ab \circledvee (d|d))$ is also clearly small in value, showing the SAC v is not at all monotonic increasing relative to probability evaluations. However both GNW and SAC "anomolies can be accounted for through their connection with logic, as will be seen. The problem of choosing the most appropriate conditional event algebra for a given situation must rely on both empirical and theoretical guidelines. In this section other significant past-derived results are shown for GNW, SAC, and Schay 1,2, together with the most recent discoveries.

Theorem 3.1

(i) Both of Schay's algebras, SAC, and GNW all reduce to the usual fixed common antecedent coset relations

$$(a|b) \vee (c|b) = (a \vee c|b); (a|b) \cdot (c|b) = (ac\ b), \qquad (3.6)$$

etc., for all $a,b,c \in R$.

(ii) Both of Schay's algebras form full (meet-join) lattices and can be algebraically characterized, leading to an extension of the Stone Representation Theorem.

(iii) GNW forms not only a full relatively pseudo-complemented lattice, but is also a Stone lattice (or Stone algebra), where the rel. pseudo. is

$$(c|d) \to (a|b) = (a|b) \vee c'd \vee b'd' \qquad (3.7)$$

Moreover, GNW can also be completely algebraically characterized, leading to an extension of the Stone Representation Theorem.

(iv) SAC forms separate semilattices relative to $\odot$ and $\circledvee$ (which are DeMorgan as is GNW), but, in general is not a full lattice.

Proofs: (i) is immediate from inspection of the defining equations.
(ii) follows from [26]
(iii) follows from [13], chapter 4
(iv) follows from (2.38),(2.39) readily. ∎

Remarks In [13], Theorem 2, p.55, it was pointed out that SAC disjunction $\circledvee$ can be derived via ordinary class intersection relative to $P(R)$: for any $a,b,c,d \in R$ (boolean algebra, as usual),

$$(a|b) \cap (c|d) = [ab, b \Rightarrow a] \cap [cd, d \Rightarrow c]$$

$$= \delta_{ab \vee cd \le (b \Rightarrow a)(d \Rightarrow c)} \cap (a|b) \circledvee (c|d) \ , \qquad (3.8)$$

where here δ is the Krönecker delta function. (But no relation with class union holds for SAC or others.)
It has also been verified that in the literature of *interval algebra* (see, e.g. [31],[32]), the above related version of $\circledvee$ is a popular choice for *conjunction*. It is important to persue these connections between interval algebra and conditional event algebra further. In fact, some results in this direction have already been obtained in showing, more generally than the boolean case, and including the property of GNW algebras, that Stone algebras with certain additional properties propagate the same structure to all their higher order interval algebras [37]. ∎

The next critical result identifies a large class of conditional event algebras with all three-valued logics (truth-functionally defined)!

Call an operation $f:(R|R)^n \to (R|R)$ *boolean-like* if there exist ordinary boolean operations $f_j:R^q \to R$ such that, using the multivariable notation

$$(\underline{a}|\underline{b}) \stackrel{d}{=} ((a_1|b_1),\ldots,(a_n|b_n)) \in (R|R)^n \ , \qquad (3.8')$$

$$f(\underline{a}|\underline{b}) = (f_1(a_1b_1,b_1,\ldots,a_nb_n,b_n) | f_2(a_1b_1,\ldots,a_nb_n)) . \qquad (3.9)$$

All of Schay's algebras as well as SAC and GNW are clearly boolean-like in all of their basic operations.

Theorem 3.2 ([33];[13], section 3.4)

There is a natural bijection, via the 3-valued indicator function transform φ, between all boolean-like conditional event algebras and all truth-functional (i.e., table-defined) 3-valued logics, such that the isomorphisms hold, for any boolean-like $f:(R|R)^n \to (R|R)$, assuming $R \subseteq P(D)$:

$$\phi(f(\underline{a}|\underline{b}))(x) = \phi(f)(\phi(\underline{a}|\underline{b})(x)) , \qquad (3.10)$$

for all $x \in D$. $\phi(f):\{0,u,1\}$ can be explicitly constructed. ∎

Remarks

1. An immediate consequence of Theorem 3.2 is that GNW algebra corresponds uniquely to Lukasiewicz' 3-valued logic (min,max,1-()), SAC algebra uniquely to Sobocinski's logic, and Schay's algebras to combinations of Bochvar's and Sobocinski's logics [13]. (See also Rescher[51] for further details of these logics.)

2. Theorem 3.2 also shows that any properties of a given conditional event algebra can be re-interpreted through an appropriately corresponding 3-valued logic and vice-versa. This can be used to explain, for example, the behavior of SAC and GNW in (3.1)-(3.5), where it is seen the Sobocinski interpretation of u is "undefined" or "not applic--ble, while the Lukasiewicz is as an actual intermediate level of truth between 0 and 1. Another approach for GNW, incidently, which avoids the trivial reductions for disjoint antecedents - explained also by referring to inconsistent data - is through the use of jointness, i.e., cartesian products and sums. (See Theorem 3.8.) Conversely, Theorem 3.2 also shows conditional event algebras supply concrete representations for any given 3-valued logics. Thus, results in the areas of both fields can be used to gain insight into the other.

3. Some examples of characterizations of conditional

event algebras carried out via use of Theorem 3.2 are (see [34] and [13], section 3.5):

The only conditional event algebra(s) which extend (unconditional) boolean conjunction, disjunction, negation (with the extension $(a|b)'=(a'|b)$) and are:
(i) DeMorgan and mutually distributive is GNW,
(ii) Stone algebra is GNW,
(iii) DeMorgan, commutative, monotone increasing for conjunction and disjunction, and continuous is GNW (continuity refers to the corresponding truth evaluations never having, e.g. 0 in a conjunction going to anything but 0 and u values not going to 1, etc.),
(iv) Non-DeMorgan full lattices are only Schay's two algebras,
(v) DeMorgan, commutative, associative, idempotent and smooth (generalization of the property enjoyed by SAC because of its antecedents forms) is SAC.

4. Recently [35], it has been shown that a full probability evaluation of (3.10) - with suitable modifications - holds for all probability measures p and r.v $V:\Omega\to D$ acting as an identity relative to p ($p(a) = p(V$ in $a)$, all a in R), and replacing u in $\phi(\underline{a}|\underline{b})$ by $p(\underline{a}|\underline{b})$, and u occurring in the domain of $\phi(f)$ by $p(\underline{a}|\underline{b})$ and in its range by $p(f(\underline{a}|\underline{b}))$, denoting these substitutions by the subscript:

$$E_V(\phi_p(f(\underline{a}|\underline{b}))(V))=E_V(\phi_p(f)(\phi_p(\underline{a}|\underline{b})(V)))=p(f(\underline{a}|\underline{b})). \quad (3.11)$$

This allows for the development of a sampling / frequency theory for conditional event algebras. ■

The following characterization of GNW algebra holds relative to the compatibility of its algebraic order (≤), and the numerical ordering for indicator functions and probabilities:

Theorem 3.3 (new result combining [13], Lemma 2, p. 48 and Theorem 1,p.154 with the definition in (2.21) here)

For all $0<a<b<1$, $0<c<d$; $a,b,c,d \in R\subseteq P(D)$:

$\phi(a|b) \le \phi(c|d)$ pointwise over D

iff $p(a|b) \le p(c|d)$, for all prob. meas. $p:D\to[0,1]$ with $p(b),p(d)>0$

iff $(a|b) \le (c|d)$ in the sense of GNW

iff $ab\le cd$ & $c'd \le a'b$. ■

By exploiting the closed interval form of conditional events, higher order ones can be obtained using also the relative pseudocomplement property of $(R|R)_{GNW}$ (see (3.7):

Theorem 3.4 ([13], section 8.1)

(i) For all $a,b,c,d \in R$, relative to GNW

$$((a|b)|(c|d)) \overset{d}{=} \{(x|y): x,y\in R \text{ \& } (c|d)\cdot(x|y)=(a|b)\cdot(c|d)\}$$
$$=((a|b)\cdot(c|d) \mid (c|d))$$
$$=[(a|b)\cdot(c|d)\ ,\ (c|d)\to(a|b)]\in PP(R) \quad (3.12)$$

(ii) Letting $\bar{u}:PP(R)\to P(R)$ be the class union operation, where, for any $A\in PP(R)$,

$$\bar{u}(A) \overset{d}{=} \bigcup\{C: C\in A\}. \quad (3.13)$$

$\bar{u}$ is a class homomorphism between $PP(R)$ and $P(R)$. In particular, for any $\alpha,\beta,\gamma,\delta \in R$ with $\alpha\le\beta$, $\gamma\le\delta$, $\alpha\le\gamma$, $\beta\le\delta$,

$$\bar{u}[[\alpha,\beta],[\gamma,\delta]] = [\alpha,\delta]. \quad (3.14)$$

Then, for all $a,b,c,d\in R$,

$$\bar{u}((a|b)|(c|d)) = (abcd|b\cdot(cd \vee a'd')), \quad (3.15)$$

$$\bar{u}((a|b)|(c))=\bar{u}((a|b)|(c|b)) = (abc|c) \quad (3.16)$$ ■

Remark. By inspection of (3.12), $((a|b)|(c|d))$ has the probability tautology form $(\alpha|\alpha)$ for some $\alpha\in(R|R)$ iff $(c|d)\le(a|b)$, (3.17)
whence

$$\bar{u}((a|b)|(c|d)) = (cd|cd) \quad (3.18)$$ ■

Next, consider connecting directly the interval form of a conditional event with its probability evaluation as a conditional probability. First, for any weight $\lambda\in[0,1]$ define the following sequence of iterated weighted averages of s and t, for any $s<t\in[0,1]$,

$$w(\lambda,1)[s,t] \overset{d}{=} (1-\lambda)s+\lambda t; \quad (3.19)$$

for any integer $n\ge2$,

$$w(\lambda,n)[s,t] \overset{d}{=} (1-w(\lambda,n-1)[s,t])\cdot s + w(\lambda,n-1)[s,t]\cdot t. \quad (3.20)$$

Lemma 3.1 ([13], pp. 151,152)

For any $s<t\in[0,1]$,

$$\lim w(\lambda,n)[s,t] = \lambda_o[s,t] \overset{d}{=} s/(1-t+s) \quad (3.21)$$

and sequence

$(w(\lambda,n)[s,t])_{n=1,2,..}$ is:

decreasing down to $\lambda_o[s,t]$ iff $\lambda>\lambda_o[s,t]$,

increasing up to $\lambda_o[s,t]$ iff $\lambda<\lambda_o[s,t]$,

identically to $\lambda_o[s,t]$ iff $\lambda=\lambda_o[s,t]$. ■

In light of the above result, call $\lambda_o[s,t]$ the *fixed point weighted average of* [s,t].

Theorem 3.5 ([13], pp. 151-152)

For any boolean algebra R and any prob. meas. $p:R\to[0,1]$ which is surjective, then for any $a,b\in R$, the functional image extension of p acting on conditional event $(a|b)=[ab,b\Rightarrow a]$ is

$$p((a|b)) = [p(ab),p(b\Rightarrow a)]=[p(ab),1-p(b)+p(ab)] \quad (3.22)$$

and

$$\lambda_o(p((a|b))) = p(a|b). \quad (3.23)$$ ■

We conclude this section with the important recent full development of non-boolean function extensions to conditional events with applications to cartesian products and sums and joint conditional r.v.'s relative to distinct antecedents.

First, let R and S be any two boolean algebras. Call $f:R\to S$ monotone increasing if

$$\text{for all } a\le b\in R,\ f(a)\le f(b)\in S. \quad (3.24)$$

In turn, the functional image extension of f (also denoted by the same symbol) $f: P(R) \to P(S)$, noting for any interval $[a,b] \subseteq P(R)$, $a \le b \in R$,

$$f[a,b] \subseteq [f(a),f(b)] \quad , \qquad (3.25)$$

in general with, however, the end events $f(a), f(b)$ achieved. With this as motivation, **define**, then the natural approximation

$$\hat{f}[a,b] \overset{d}{=} [f(a),f(b)], \qquad (3.26)$$

omitting in practise, for convenience, the hat notation. This yields

Theorem 3.6 (easily proven new result)

Let $f: R \to \mathcal{S}$ be a monotone increasing function between two boolean algebras. Then, the functional image extension (à la (3.26)) $f:(R|R) \to (S|S)$ is well-defined with, for any $a,b \subseteq R$,

$$f(a|b) = [f(ab), f(b \Rightarrow a)] = (f(ab) \,|\, f(b \Rightarrow a) \to f(ab)) \qquad (3.27)$$

■

Theorem 3.7 (corollary to Theorem 3.6 [36])

Let $R \subseteq P(D)$ and $S \subseteq P(F)$ be two boolean algebras and $f: D \to F$ be any function and $f: P(D) \to P(F)$ its functional image extension and $f^{-1}: P(F) \to P(D)$ its inverse image extension. Then the above f and f^{-1} both satisfy the hypothesis of Theorem 3.6, with (3.27) thus valid and $f^{-1}:(S|S) \to (R|R)$ well-defined, where for any $c,d \in S$,

$$\begin{aligned} f^{-1}(c|d) &= f^{-1}[cd, d \Rightarrow c] = [f^{-1}(cd), f^{-1}(d \Rightarrow c)] \\ &= [f^{-1}(c) \cdot f^{-1}(d), f^{-1}(d) \Rightarrow f^{-1}(c)] \\ &= (f^{-1}(c) \,|\, f^{-1}(d))) \; . \end{aligned} \qquad (3.28)$$

■

Next, define the product boolean algebra $R \otimes S$ of boolean algebras R and S as above by simply first forming

$$R \times S \overset{d}{=} \{a \times b : a \in R,\ b \in S\} \qquad (3.29)$$

and then letting $R \otimes S$ be the smallest boolean algebra $\subseteq P(D \times F)$ with $R \times S \subseteq R \otimes S$, where, as usual,

$$(a \times b) \cdot (c \times d) = ac \times bd \, , \qquad (3.30)$$

$$a \times b = (a \times 1) \cdot (1 \times b) \, , \qquad (3.31)$$

$$(a \times b)' = (1 \times b') \vee (a' \times 1)' \qquad (3.32)$$

$$a \times 0 = 0 \times a = 0 \, , \qquad (3.33)$$

and cartesian sum is given by

$$a \dagger b \overset{d}{=} (a' \times b')' = (1 \times b) \vee (a \times 1). \qquad (3.34)$$

Theorem 3.8 (new result [36])

Given R and S as in Theorem 3.7, it follows that the operations $\times, \dagger: (R,S) \to R \otimes S$ are both monotone increasing and an obvious modication of Theorem 3.6 is valid here, yielding, for any $a,b \in R$, $c,d \in S$,

$$\begin{aligned} (a|b) \times (c|d) &= (ab \times cd \,|\, ((b \Rightarrow a) \times (d \Rightarrow c)) \Rightarrow (ab \times cd) \\ &= (ab \times cd \,|\, (a'b \times 1) \vee (1 \times c'd) \vee (b \times d)) \\ &= (ab \times cd \,|\, (a'b \times 1) \vee (1 \times c'd) \vee (ab \times cd)) \\ &= ((a|b) \times (d|d)) \cdot ((b|b) \times (c|d)) \\ &= ((a|b) \times 1) \cdot (1 \times (c|d)) \end{aligned} \qquad (3.35)$$

and

$$\begin{aligned} (a|b) \dagger (c|d) &= (ab \dagger cd \,|\, ((b \Rightarrow a) \dagger (d \Rightarrow c)) \Rightarrow (ab \dagger cd)) \\ &= (ab \dagger cd \,|\, (ab \times 1) \vee (1 \times cd) \vee (b \times d)) \\ &= ((a|b)' \times (c|d)')' = ((a|b) \times 1) \vee (1 \times (c|d)) \end{aligned} \qquad (3.36)$$

noting the complete analogue of eq. (3.35) to conjunction (2.23), (2.24) and of (3.36) to disjunction (2.25), (2.26), for the GNW algebra.

■

Next, apply the above results to the development of conditional random variables. (This rigorizes the previous ad hoc involved approach given in [13], section 5.3.)

Let (Ω, A, p) be a fixed probability space, $(\mathbb{R}^k, \mathbb{B}^k)$ the usual k-dimension real measurable space, $V: \Omega \to \mathbb{R}^m$ and $W: \Omega \to \mathbb{R}^n$ any two r.v.'s. Also, define for any $a \subseteq \mathbb{B}^m$, $b \subseteq \mathbb{B}^n$, the product form conditional event

$$[a|b] \overset{d}{=} (a \times b \,|\, \mathbb{R}^m \times b) \, . \qquad (3.37)$$

Denote $[\mathbb{B}^m|\mathbb{B}^n] \overset{d}{=} ([a|b] : a \subseteq \mathbb{B}^m,\ b \subseteq \mathbb{B}^n)$. (3.37')

Consider the joint or cartesian product mapping $V \times W: \Omega \to \mathbb{R}^m \times \mathbb{R}^n$, where for any $x \in \Omega$,

$$(V \times W)(x) \overset{d}{=} (V(x), W(x)), \qquad (3.38)$$

and its functional image extension $V \times W: P(\Omega) \to P(\mathbb{R}^m) \times P(\mathbb{R}^n)$, appropriately restricted and assumed range-measurable, so that $V \times W: A \to \mathbb{B}^m \times \mathbb{B}^n$ is well-defined. In turn, using Theorem 3.7, $V \times W: (A|A) \to (\mathbb{B}^m \times \mathbb{B}^n | \mathbb{B}^m \times \mathbb{B}^n)$ is also well-defined.

Also, define $[V|W]^{-1}: [\mathbb{B}^m|\mathbb{B}^n] \to (A|A)$ as the restriction $(V \times W)^{-1}: [\mathbb{B}^m|\mathbb{B}^n] \to (A|A)$, noting for any $a \in \mathbb{B}^m$, $b \in \mathbb{B}^n$, the identifications from Theorem 3.7,

$$(V \times W)^{-1}(a \times b) = V^{-1}(a) \cdot W^{-1}(b), \qquad (3.39)$$

and as in (3.28),

$$\begin{aligned} [V|W]^{-1}[a|b] &= ((V \times W)^{-1}(a \times b) \,|\, (V \times W)^{-1}(\mathbb{R}^m \times b)) \\ &= (V^{-1}(a) \cdot W^{-1}(b) \,|\, W^{-1}(b)) = (V^{-1}(a) \,|\, W^{-1}(b)). \end{aligned} \qquad (3.40)$$

Thus:

Theorem 3.9 (new result [36])

With the same assumptions as above,

$$\begin{aligned} p([V|W] \text{ is in } [a|b]) &\overset{d}{=} p([V|W]^{-1}[a|b]) \\ &= p(V^{-1}(a) \,|\, W^{-1}(b)) \\ &= p(V \text{ is in } a \,|\, W \text{ is in } b). \end{aligned} \qquad (3.41)$$

■

Thus, there is full compatibility between the traditional approach to conditioning and the use of conditional event algebra to extend. Indeed, analogous to the way an unconditional r.v. or a fixed antecedent conditional r.v. ($V|W$ in b) induces a probability space, $[V|W]$ induces the space $(\mathbb{R}^{m+n}, [\mathbb{B}^m|\mathbb{B}^n], p \circ [V|W]^{-1})$, where $p \circ [V|W]^{-1}: [\mathbb{B}^m|\mathbb{B}^n] \to (A|A)$ produces conditional probability as in (3.41) and preserves all GNW operations. For example, for any $a,c \in \mathbb{B}^m$, $b,d \in \mathbb{B}^n$, for GNW conjunction

$$\begin{aligned} & p([V|W]^{-1}([a|b] \cdot [c|d]) \\ &= p([V|W]^{-1}(ac \times bd \,|\, (ac \times bd) \vee (a' \times b) \vee (c' \times d))) \\ &= p(V^{-1}(ac) \cdot W^{-1}(bd) \,|\, (V^{-1}(ac) \cdot W^{-1}(bd)) \vee (V^{-1}(a)' W^{-1}(b)) \\ &\qquad \vee (V^{-1}(c)' W^{-1}(d))) \\ &= p((V^{-1}(a) \,|\, W^{-1}(b)) \cdot (V^{-1}(c) \,|\, W^{-1}(d))) \\ &= p([V|W]^{-1}[a|b] \cdot [V|W]^{-1}[c|d]) \; . \end{aligned} \qquad (3.42)$$

Finally, consider the problem of modeling joint conditional r.v.'s, especially those with no common antecedent restriction.

Theorem 3.10 (new result [36])

As before, (Ω,A,p) is a fixed probability space. Let $V_j:\Omega\to\mathbb{R}^{m_j}$, $W_j:\Omega\to\mathbb{R}^{n_j}$ be r.v.'s, with all of the development in eqs.(3.38)-(3.42) valid for j=1, 2. Then, for any $a_j \in \mathbf{B}^{m_j}$, $b_j\in\mathbf{B}^{q_j}$, j=1,2, relative to GNW,

$$\begin{aligned} &p([V_1|W_1] \text{ is in } [a_1|b_1] \;\&\; [V_2|W_2] \text{ is in } [a_2|b_2]) \\ &\overset{d}{=} p([V_1|W_1]^{-1}[a_1|b_1]\cdot[V_2|W_2]^{-1}[a_2|b_2]) \\ &= p(V^{-1}(a_1)|W_1^{-1}(b_1))\cdot(V_2^{-1}(a_2)|W_2^{-1}(b_2))) \\ &= p(V_1\times W_1\times V_2\times W_2 \text{ is in } [a_1|b_1]\times[a_2|b_2]) \qquad (3.43)\end{aligned}$$

■

The following result shows that for fixed, but not necessary equal antecedents, the cartesian product and conjunction forms of conditional r.v.'s yields legitimate joint probability distributions!

Theorem 3.11 (new result [36])

Let $t\in\mathbf{R}$ be arbitrary and denote infinite left ray

$$a(t) \overset{d}{=} (-\infty,t]\,. \qquad (3.43')$$

For any given probability space (Ω,R,p) ($R\le P(\Omega)$ certainly a boolean algebra), choose any $b_j \in R$, $j=1,\dots,n$, with $p(b_1\cdots b_n)>0$. Consider then, the function $F_{\underline{b}}:\mathbf{R}^n\to[0,1]$, where, for any $\underline{t} = (t_1,\dots,t_n) \in \mathbf{R}^n$,

$$F_{\underline{b}}(\underline{t}) \overset{d}{=} p((a(t_1)|b_1)\cdots(a(t_n)|b_n)), \qquad (3.44)$$

with conjunction in the GNW sense. Then, $F_{\underline{b}}$ is a legitimate joint cdf over $\mathbf{R}^n$. ■

Remarks. For n=2, one can write, using eqs. (2.23), (2.24),

$$F_{\underline{b}}(\underline{t}) = \alpha(\underline{t})/(c-\beta(t_1)-\gamma(t_2)), \qquad (3.45)$$

where

$$\alpha(\underline{t})\overset{d}{=}p(a(t_1)\cdot a(t_2)\cdot b_1\cdot b_2);\ \beta(t_1)\overset{d}{=}p(a(t_1)\cdot b_1\cdot b_2');$$
$$\gamma(t_2)\overset{d}{=}p(a(t_2)\cdot b_1'\cdot b_2)\ :\ c\overset{d}{=}p(b_1\vee b_2)\,. \qquad (3.46)$$

Assuming probability density $f_{\underline{b}}$ exists for $F_{\underline{b}}$, it follows from standard relations,

$$f_b(\underline{t}) = \partial^2 F_{\underline{b}}(\underline{t})/\partial t_1\partial t_2 = A(\underline{t})/B(\underline{t})^4\,, \qquad (3.47)$$

where $$B(\underline{t}) \overset{d}{=} c-\beta(t_1)-\gamma(t_2)\,, \qquad (3.48)$$

$$\begin{aligned} A(\underline{t}) \overset{d}{=}\; &2B(\underline{t})\cdot\alpha(\underline{t})\cdot\beta'(t_1)\cdot\gamma'(t_2) + \\ &B(\underline{t})^2\cdot(\partial\alpha(\underline{t})/\partial t_1)\cdot\gamma'(t_2) + (\partial\alpha(\underline{t})/\partial t_2)\cdot\beta'(t_1)) \\ &+ B(\underline{t})^3\cdot\partial^2\alpha(\underline{t})/\partial t_1\partial t_2. \qquad (3.48)\end{aligned}$$

Thus, it is clear, by inspection, even if the $p(a(t_j)|b_j)$ are gaussian distributed in t_j, j=1,2, $f_b(\underline{t})$ will not in general take any joint gaussian distribution form. Of course, one notes, on the other hand, the reduction of all of the above, when $b_1=b_2=b$, in which case $F_b(\underline{t})$ reduces simply to the function $p(a(t_1)\cdot a(t_2)|b)$ and $f_b(\underline{t})$ accordingly, which is compatible with gaussian forms. ■

Other recent developments in conditional event algebra include: A full extension of GNW algebra to the consideration of certain types of Stone algebras and the higher order interval algebras generated from them - which can be shown also to have a certain Stone algebra structure. This structure is useful in addressing the higher order conditioning problem, since the union of the nested higher order interval algebras is also a certain type of Stone algebra and closed with respect to conditioning. (See [37].); Conditional event extension of the classical relation between material implication being a tautology and its antecedent being dominated by its consequence relative to the basic order (≤) [13], pp. 191-193, [38]; An extension of the normal disjunctive expansion of all boolean operations over boolean algebra R to all boolean-like operations over (R|R) [39]; A semantic extended division of indicator approach, with applications to an alternate approach to conditional events, the higher order conditioning problem, and the definition of conditional fuzzy sets [40]. Due to severe space limitations, only the last-mentioned result will be briefly outlined in the second part of this section, following an exposition on a recent breakthrough in the representation of fuzzy sets by random sets.

4 Modeling Natural Language Information Through Fuzzy Set Theory and Relations with Random Sets and Conditional Events

4.1 A Recent Result on the Uniqueness of Represenation of Fuzzy Sets by Random Sets

The first premise here is that essentially all natural language descriptions can be modeled in a straightforward way by formal logical combinations of membership relations between population elements, or measurement variables connected with them, and modifiers or attributes - called *fuzzy sets*. For example, the sentence

$s_1 \overset{d}{=}$ "The ship is rather close to us, but is still going very fast" (4.1)

can be naturally interpreted as

$$s_1 = (\text{dist(ship,us)} \in \text{rather[close]})\ \&\ (\text{speed(ship)} \in \text{very[fast]})\,. \qquad (4.2)$$

The second premise is that the semantic or truth evaluation of the sentence(s) can be carried out truth-functionally, i.e., in a homomorphic way-preserving completely all formal relations, where the formal relation between element and fuzzy set is converted to a corresponding membership function over a domain of values of the attribute and with range values in the unit interval of truth possibilities. In particular, all ordinary (unconditional) sets and the ordinary membership relation (ε) are special cases of fuzzy sets and their membership functions. Thus, as an example of evaluations, s_1 becomes (4.3)

$$tr(s_1) = \phi_\&(\phi_{rath}(\phi_{clos}(d)),\phi_{very}(\phi_{fast}(s)))\,,$$

where $\phi_\&:[0,1]^2\to[0,1]$ is a binary operation over the unit interval squared, representing conjunction. This is often put in the form of min or more generally, a *t-norm* or *copula*. A t-norm $t:[0,1]^2\to[0,1]$ is a nondecreasing - usually continuous - function with boundary conditions (4.4)

for all $x\in[0,1]$, $t(x,0)=t(0,x)=0; t(x,1)=t(1,x)=x$,

and being an associative, commutative function.

The latter conditions allow the t-norm to be extended to any finite number of arguments unambiguously. Usually, the t-norm chosen and its t-conorm partner are in a DeMorgan relation, with the t-conorm satisfying dual conditions as for t-norms and with negation being represented by the membership function

$$\phi_{not} = 1-(\ ). \qquad (4.5)$$

Thus, t-conorms, representing disjunction, if in a DeMorgan relation to their t-norm partners, can be written

$$\bar{t}(x,y) = 1-t(1-x,1-y)\ ,\ \text{all } x,y \in [0,1].$$

(See [41] or [16] for background.)

Also, a copula is a function which is formally the same as the cdf of any n-ary joint r.v. all of whose marginals are distributed uniformly over [0,1]. In connection with this, the following result is critical:

Theorem 4.1 (A.Sklar, 1973 [42])

(i) For any positive integer n and any joint cdf $F:\mathbb{R}^n \to [0,1]$, denoting its one-dimensional marginals as $F_j:\mathbb{R} \to [0,1]$, $j=1,\ldots,n$, there is always some copula $cop:[0,1]^n \to [0,1]$ such that

$$F = cop \circ (F_1, \ldots F_n). \qquad (4.7)$$

(ii) Conversely, the right hand side of (4.7), for any choice of cop and one-dimensional marginal cdf's yields F as a legitimate cdf over $\mathbb{R}^n$. ■

Returning to the example in eq.(4.3), $\phi_{rath}, \phi_{very}: [0,1] \to [0,1]$ are natural interpretations, with ϕ_{very} being nondecreasing, 0 at 0, 1 at 1. $\phi_{clos}: \mathbb{R}^+ \to [0,1]$ represents possible distances from "us" and is 1 at 0 and nonincreasing approaching 0 at large values. Similar comments hold for ϕ_{fast}.

Despite the direction of the majority of the fuzzy set community - and that of the larger probability community - in creating a separation of perspectives in modeling information, a relatively small group of individuals have observed connections between the two areas (Höhle, Goodman, Nguyen, etc.). See, e.g. [13], chapter 5 and recent comments in [44]. Specifically, it is now fairly well known that fuzzy set membership functions correspond to the weakest basic way random sets can be specified, analogous to the situation in classical r.v. use where one knows only the mean of the r.v.'s involved, not the entire distributions. Quantitatively, this can be interpreted as follows: For any given function $f:D \to [0,1]$, there exist - in general infinitely many - random subsets $S:\Omega \to P(D)$, for some probability space (Ω,A,p) such that S *is one point coverage function equivalent to* f, i.e.,

$$p(x \in S) = f(x)\ ,\ \text{all } x \in D. \qquad (4.8)$$

One such random set is easily constructable : the *canonical nested random set* or *random level set* (or *random cut set*) $S_f(U):\Omega \to [0,1]$, where $U:\Omega \to [0,1]$ is a uniformly distributed r.v. (surjective) and

$$S_f(U) \overset{d}{=} f^{-1}[U,1] = \{x: x \in D \ \&\ f(x) \geq U\}. \qquad (4.9)$$

A major stumbling block in the utilization of this relationship has been which one(s) of the one point coverage equivalent random subsets of D are the most appropriate to replace the original fuzzy set membership function they represent? In addition, can the problem of determining the most appropriate fuzzy set operations to choose be related to the one point coverage problem ? Although Zadeh originally proposed from an ad hoc viewpoint the system (min,max,1-( )) for conjunction, disjunction, and negation, respectively, a plethora of more general classes of systems have since been proposed, including the t-norms, t-conorms mentioned earlier. These issues have really remained open now for the past dozen or more years since they have been proposed. (See, however, e.g. [13] for some progress in this area.) Finally, with some recent discoveries, it appears that these issues are now close to being resolved:

In the following, let (Ω,A,p) be some conveniently chosen probability space, $f:D \to [0,1]$ any given function, $R(D)$ the class of all random subsets $S:\Omega \to P(D)$, which are considered as distinct only if they induce distinct distributions via p, and $R_f(D)$, the class of all one point coverage equivalent random subsets of D to f. Also, let $V(D)$ denote the class of all stochastic 0-1 valued processes of the form $V=(V_x)_{x \in D}$, $V_x:\Omega \to \{0,1\}$ a r.v., for all $x \in D$, such that $p(V_x=1) = f(x)$, all $x \in D$. In addition, denote the corresponding cdf of each V_x as $F(V_x)$ and call $F(V)$ the class of all joint cdf's over the set $\{0,1\}^D$ for which the marginals $F(V_x)$ are fixed, $x \in D$.

Theorem 4.2 (originally shown in [12], Theorem 2.1.1 in a modified form; [45]-[47])

(i) Using the above notation, the following diagram holds, for any given $f:D \to [0,1]$, ϕ being indicator function,

$$\begin{array}{ccccc} R(D) & \xrightarrow[\text{bijection}]{\phi} & V(D) & \xrightarrow[\text{bijection}]{F} & F(V(D)) \\ \cup\!| & & \cup\!| & & \cup\!| \\ R_f(D) & \xrightarrow[\text{bijection}]{\phi} & V_f(D) & \xrightarrow[\text{bijection}]{F} & F(V_f(D)) \end{array}$$

(ii) By Sklar's Theorem,

$$F(V_f(D)) = \{cop \circ ((F(V_x))_{x \in D}: \text{cop arb over } [0,1]^D\}$$

$$S(cop,f) \overset{d}{=} \text{typical 1-pt equiv.r.set to f}$$

$$= \{\phi^{-1} \circ F(\ )^{-1} \circ (cop \circ ((F(V_x))_{x \in D})\}, \qquad (4.10)$$

$$R_f(D) = \{S(cop,f): \text{cop arb over } [0,1]^D\}. \qquad (4.11)$$

(iii) All of the above results can be extended to the following typical joint situation, where D is replaced by $\underline{D} = (D_{ij})_{i \in I, j \in J}$, f by $\underline{f} = (f_{ij})_{i \in I, j \in J}$, $f_{ij}:D_{ij} \to [0,1]$, with $R_f(D)$ replaced by $R_{\underline{f}}(\underline{D})$, $V_f(D)$ replaced by $V_{\underline{f}}(\underline{D})$, etc. ■

Thus, eqs.(4.10),(4.11) show that by choosing arbitrary copulas, one can construct all solutions to the one point coverage equivalence problem. Theorem 4.2(iii) provides the setting for the new characterization of fuzzy set operations compatible with probability interpretations:

Theorem 4.3 (new result [44],[45])

Fix probability space (Ω,A,p) and consider any operator pair for conjunction disjunction in the form (cop,cocop) which is also a (DeMorgan, continuous) t-norm,t-conorm pair. Also, for any choice of finite collection of finite sets $\underline{D}$, any functions $\underline{f}$, as above, any $\underline{x} \overset{d}{=} (x_{ij})_{i \in I, j \in J}$, $x_{ij} \in D_{ij}$,

$\underline{f}(\underline{x}) \overset{d}{=} (f_{ij}(x_{ij}))_{i\epsilon I, j\epsilon J}$, and any random sets $\underline{S} \overset{d}{=} (S_{ij})_{i\epsilon I, j\epsilon J} \epsilon R_{\underline{f}}(\underline{D})$, i.e., $S_{ij} \epsilon R_{f_{ij}}(D_{ij})$, such that it is generated by cop – $S_{ij} = S_{ij}(\text{cop}, f_{ij})$ as in (4.10), <u>the following two statements are equivalent</u>:

(1) For all possible $f_{ij}: D_{ij} \to [0,1]$, all possible corresponding S_{ij}, all possible x_{ij}, cop, cocop and S_{ij} are such that the following homomorphism holds for all possible combinations of cop over cocop and cocop over cop relative to corresponding probabilities of combinations of ordinary conjunctions over disjunctions and disjunctions over conjunctions of one point coverage relations of x_{ij} by S_{ij}. That is,

$$\text{tr}(\underset{j\epsilon J}{\&} \underset{i\epsilon I}{\text{or}} (x_{ij} \epsilon \text{ fuzzy set corresponding to } f_{ij})$$

$$= \text{cop}((\text{cocop}((f_{ij}(x_{ij}))_{i\epsilon I}))_{j\epsilon J})$$

$$= p(\underset{j\epsilon J}{\&} \underset{i\epsilon I}{\text{or}} (x_{ij} \epsilon S_{ij})) \qquad (4.12)$$

and

$$\text{tr}(\underset{i\epsilon I}{\text{or}} \underset{j\epsilon J}{\&} (x_{ij} \epsilon \text{ fuzzy set corresponding to } f_{ij})$$

$$= \text{cocop}((\text{cop}((f_{ij}(x_{ij}))_{j\epsilon J}))_{i\epsilon I})$$

$$= p(\underset{i\epsilon I}{\text{or}} \underset{j\epsilon J}{\&} (x_{ij} \epsilon S_{ij})) . \qquad (4.13)$$

(2) $(\text{cop}, \text{cocop}) = (\min, \max) \qquad (4.14a)$

and

$$S_{ij} = S_{f_{ij}}(U) \text{ , all } i\epsilon I, j\epsilon J \text{ ,} \qquad (4.14b)$$

where $U: \Omega \to [0,1]$ is a fixed uniformly distributed r.v., as before.

<u>Proof</u>: The proof is rather long and detailed and is provided in [46]. However, an outline of it is presented below, since the key steps are of interest in themselves:

<u>Lemma</u> 4.1 If the hypothesis of Theorem 4.3 holds (prior to statements (1) and (2)), then extending the above multivariable notation in the obvious way, the following two statements are equivalent:

(i) $\text{cocop}(\text{cop}(\underline{f}(\underline{x}))) = p(\text{or } \& \ (\underline{x} \epsilon \underline{S})) \qquad (4.15)$

(ii) $\text{cocop}(\underline{f}(\underline{x})) = \sum_{\emptyset \neq K \subseteq J} (-1)^{\text{card}(K)+1} \cdot \text{cop}((g_{ij}(x_{ij}))_{i\epsilon I, j\epsilon K}) . \qquad (4.16)$ ∎

The next result modifies Frank's well known theorem ([48], Theorem 5.1, pp. 220-222) by replacing his modularity assumption (essentially, the relation expressed in (4.16) for car(K)=2) by a DeMorgan one and his conclusion, which includes the class of all possible *ordinal sums* (certain types of affine transform)on t-norm, t-conorm pair (prod,probsum), by only (prod,probsum) itself. (As usual, prod indicates arithmetic multiplication and probsum is its DeMorgan transform.)

<u>Lemma</u> 4.2 Again, assume the hypothesis of Theorem <u>4.3</u>, but here lessen the requirement by replacing cop by just a continuous t-norm and cocop by the corresponding (DeMorgan) t-conorm $\check{t}$ only. Then, for all $\underline{f}$, $\underline{x}\epsilon\underline{D}$ $(f_{ij}: D_{ij} \to [0,1])$, and all $\underline{s} \epsilon [0,1]^n$, for appropriately determined n, using notation $s_K \overset{d}{=} (s_j)_{j\epsilon K}$, the following 8 statments are equivalent:

(i) For all $\underline{s}$ $\quad t(\underline{s}) = \sum_{\emptyset \neq K \subseteq \{1,\ldots,n\}} (-1)^{\text{card}(K)+1} \cdot \check{t}(s_K) \qquad (4.17)$

(ii) For all $s_1, s_2, s_3 \epsilon [0,1]$,

$$t(s_1, \check{t}(s_2, s_3)) = t(s_1, s_2) + t(s_1, s_3) - t(s_1, s_2, s_3) \qquad (4.18)$$

(iii) $(t, \check{t})$ is either (min,max)or(prod,probsum). ∎

<u>Lemma</u> 4.3. If the hypothesis of Theorem 4.3 holds, then by combining Lemmas 4.1 and 4.2, eq.(4.12) holding implies that (cop,cocop) must be either (min,max) or (prod,probsum). ∎

<u>Lemma</u> 4.4 If the hypothesis of Theorem 4.3 holds, and eqs.(4.12) and (4.13) also both hold, then necessarily cop must be distributive over cocop. ∎

Finally, the proof of Theorem 4.3 is completed when it is noted that prod in general is not distributive over probsum, and hence the conclusion of Lemma 4.3 must be restricted to (min,max). Also, conversely, as a check, it can be easily verified that (min,max) does satisfy both eqs. (4.12) and (4.13) for all possible allowable variables and that this determined that eq.(4.14b) holds. ∎

Many fuzzy set concepts -such as the extension of ordinary functions to fuzzy sets ("extension principle") and fuzzy relations - can be defined through combinations of conjunctions and disjunctions. Thus, the import of Theorem 4.3 is that: Unless prior information indicates other facts, *the only universal interpretation of fuzzy set theory compatible with probability theory in the one point coverage sense must be through the canonical nested random sets and only Zadeh's original* (min,max,1-( )) *system provides this compatibility.* Hence, in combining linguistic information with stochastic information, it is natural to convert all of the linguistic information first to the nested random set form, and then combine the resulting all-stochastic information as will be seen in section 5.2.

4.2 An Extended Numerical Division of Indicator Functions Applied to the Development of Conditional Events and Fuzzy Conditional Events

Throughout this subsection, T will always refer to an arbitrary fixed nondecreasing continuous function with boundary conditions the same as for t in (4.1). That is, T is a generalization of classical logic table for conjunction, and in fact that of the form of t-norms or copulas described in section 4.1.

Next, define *T-extended division* $(./..)_T:[0,1]^2 \to P[0,1]$, where, for any $r,s \in [0,1]$ (including 0 values traditionally banned from division in ordinary arithmetic division),

$$(r/s)_T \overset{d}{=} \{x : x \in [0,1] \ \& \ T(x,s) = \min(r,s)\} = (\min(r,s)/s)_T \ . \quad (4.19)$$

Extending this further, define $(./..)_T:P[0,1]^2 \to P[0,1]$ where for any $\alpha,\beta \subseteq [0,1]$,

$$(\alpha/\beta)_T = (\min(\alpha,\beta)/\beta)_T \overset{d}{=} \{\gamma : \gamma \subseteq [0,1] \ \& \ T(\gamma,\beta) = \min(\alpha,\beta)\}$$

$$= \text{maximal (classwise inclusion)} \ \gamma \ \text{such that} \ T(\gamma,\beta) = \min(\alpha,\beta) \ , \quad (4.20)$$

where $\min(\alpha,\beta)$ is the usual functional image extension of min applied to α,β, and similarly, for $T(\gamma,\beta)$.

Clearly, for many pairs of α,β, $(\alpha/\beta)_T$ will be vacuous (except for T=min). But for a large class, including intervals from [0,1], nontrivial extended divisions will result. Note also the reduction of the definition to the ordinary arithmetic division for any $r,s \in [0,1]$:

$$(r/s)_{prod} = r/s \quad (s > 0) \ . \quad (4.21)$$

In turn, now using the obvious componentwise definition, T-extended division can be applied to any pair of functions $f,g:D \to [0,1]$ to yield $(f/g)_T:D \to P[0,1]$, where for any $x \in [0,1]$,

$$(f/g)_T(x) \overset{d}{=} (f(x)/g(x))_T . \quad (4.22)$$

By restricting f and g to ordinary set indicator functions $\phi(a),\phi(b):D \to \{0,1\}$, for any $a,b \in P(D)$, and to the three-valued indicator functions of conditional events $\phi(a|b)$, as given in eq.(2.21), where here - and from now on, unless otherwise indicated - one identifies the third value

$$u = [0,1] \ , \quad (4.23)$$

the following result obtains:

Theorem 4.4 Derivation of conditional events through their indicator functions being T-extended divisions of unconditional consequent indicator functions by antecedent ones and $\bar{u}$-closure of higher order conditional events through T-extended division. (new result [40]).

Let D be any (nonempty) set and $a,b,c,d \subseteq D$ arb. Then for T arbitrary as before:

(i)

$$(\phi(a)/\phi(b))_T = (\phi(ab)/\phi(b))_T = \phi(a|b); \quad (4.24)$$

(ii) Interpreting conditional event conjunction · as GNW (eqs.(2.23),(2.24)),

$$(\phi(a|b)/\phi(c|d))_T = (\phi((a|b)\cdot(c|d))/\phi(c|d))_T = \phi(\bar{u}((a|b)|(c|d))), \quad (4.25)$$

where $\bar{u}$ is the class union operator used to reduce higher order conditional events homomorphically (see Theorem 3.4). ∎

Motivated by the last satisfactory tie-ins between the previous algebraic/syntactic approach to conditioning and the more numeric/semantic by T-extended division, we now define *T-conditional fuzzy sets* as simply the componentwise definition given in eq.(4.22), to be read "f given g through T" or "if f then g, wrt T", etc.

The next result provides min-based fuzzy conditionals.

Theorem 4.5 (new result [40])

Let T = min here. Then, for any $f,g,f_j,g_j:D \to [0,1]$ and any $x \in D$:

(i)

$$(f/g)_{min}(x) = \begin{cases} [g(x),1], \text{ if } g(x) \le f(x) \ , \\ f(x) \ , \text{ if } f(x) < g(x) \end{cases}$$

$$= [\min(f(x),g(x)) \ , \ g(x) \to f(x) \ ], \quad (4.26)$$

analogous to the interval form in eq.(2.9), where, here → in (4.26) denotes the relative pseudocomplement (as is ⇒ for boolean algebra) for the lattice [0,1](min,max), where for any $r,s \in [0,1]$

$$r \to s = \begin{cases} 1 \ , \text{ if } r \le s \ , \\ s \ , \text{ if } s < r \ . \end{cases} \quad (4.27)$$

(ii)

$$((f_1/g_1)_{min}/(f_2/g_2)_{min})_{min}(x) = [\min(f_1(x),g_1(x),f_2(x),g_2(x)) \ , \ (g_2(x) \to f_2(x)) \to (g_1(x) \to f_1(x))], \quad (4.28)$$

resulting in a [0,1]-sub interval valued function, readily computable.

(iii) In general, though all higher order min-conditional fuzzy sets can be obtained as straightforward generalizations of (4.28), they do not reduce to the same forms, as is the case in (4.25) for ordinary conditional events. ∎

Theorem 4.6 (new result [40])

Suppose T is now arbitary strictly increasing in its arguments. Then, for any $f,g,f_j,g_j:D \to [0,1]$, and $x \in [0,1]$:

(i)

$$(f/g)_T(x) = \begin{cases} [0,1] \ , \text{ if } g(x) = 0, \\ (f(x)/g(x))_T \in [0,1] \ , \text{ if } g(x) > 0 \ . \end{cases} \quad (4.29)$$

(ii)

$$((f_1/g_1)_T/(f_2/g_2)_T)_T(x) \begin{cases} = [0,1] \ , \text{ if } g_1(x) = 0 \ , \\ = [0,(f_1(x)/g_1(x))_T] \ , \text{ if } g_1(x) > 0, g_2(x) = 0 \ , \\ \in [0,1], \text{ if } g_1(x),g_2(x) > 0 \end{cases} \quad (4.30)$$

(iii) All third and higher order T-conditional fuzzy sets reduce to the second order form as given in (ii) ∎

The next result shows compatibility between the ideas of T-conditional fuzzy sets, the one point representation of fuzzy sets by nested random sets, and the conditional events formed from these random sets.

First, consider $f,g:D\to[0,1]$ and recall (see (4.8),(4.9)) the one point coverage equivalent representations of them through the nested random sets $S_f(U), S_g(U)$, so that one has for all $x\in D$,

$$p(x\in S_f(U)) = E_U(\phi(S_f(U))(x)) = f(x) \ . \qquad (4.31)$$
$$p(x\in S_g(U)) = E_U(\phi(S_g(U))(x)) = g(x) \ .$$

Recall also the fixed point weighted aver. $\lambda_0[s,t]$ for any interval of numbers $[s,t]\subseteq[0,1]$ given in eq.(3.21) and denote, as usual, ϕ to mean either the ordinary set indicator function or the generalization to the indicator function of conditional vents, with the identification for u given in (4.23). Also note that $(S_f(U)|S_g(U)):\Omega\to(P(D)|P(D))$ is a *random conditional event*, where for any $\omega\in\Omega$

$$(S_f(U)|S_g(U))(\omega) \overset{d}{=} (S_f(U(\omega))|S_g(U(\omega))). \quad (4.32)$$

In addition, for any $x\in D$, analogous to (4.31), it is natural to define the *one point coverage function of a random conditional set* typically as

$$p(x\in(S_f(U)|S_g(U))) \overset{d}{=} E_U(\phi(S_f(U)\ |\ S_g(U))(x))$$
$$= 1\cdot p(x\in(S_f(U)\cdot S_g(U))) + 0\cdot p(x\in(S_f(U)'\cdot S_g(U)))$$
$$+ [0,1]\cdot p(x\in S_g(U)')$$
$$= [\min(f(x),g(x))\ ,\ \min(f(x),g(x))+1-g(x)]$$
$$\subseteq [0,1]\ , \qquad (4.33)$$

using (2.21).

All of the above, finally leads to:

Theorem 4.7 (new result [40])

Using the above notation, let $f,g:D\to[0,1]$ be any two given fuzzy set membership functions. Then, for all $x\in D$,

$$\lambda_0(p(x\in(S_f(U)|S_g(U))))$$
$$= (f/g)_{prod}(x)$$
$$= E_U(\phi(S_{(f/g)_{prod}}(U)\ )(x))\ , \qquad (4.34)$$

the bottom equation holding except when $g(x) = 0$. ■

Remarks. Previously in [17], eq.(9.21) (see also [13], Chapter 7) fuzzy conditional events were defined in a, more or less, ad hoc manner, which in the notation here led to the form $(f/g)_{prod}$, for any given $f,g:D\to[0,1]$, fortunately the same as in Theorem 4.7 !.

Also, fuzzy set operations, such as Zadeh's (min,max,1-( )), or any t-norm, t-conorm system, can all be applied to fuzzy conditional sets as defined above by using the functional image approach. For example, for any $f_j,g_j:D\to[0,1]$, $x\in D$, for conjunction represented by Zadeh's min,

$$(\min((f_1/g_1)_{min},(f_2/g_2)_{min}))(x) = [a(x),b(x)]\ , \qquad (4.35)$$

where
$$a(x) \overset{d}{=} \min(f_1(x),g_1(x),f_2(x),g_2(x)) \quad (4.36)$$
$$b(x) \overset{d}{=} \min(g_1(x)+f_1(x),g_2(x)+f_2(x)). \quad (4.37)$$

Finally, it can be observed that the fixed point weighted average function can also be used to evaluate probabilistically T-conditional fuzzy sets, since they are actually set-valued and, in particular, for T=min or T stictly increasing, interval- or point-valued relative to [0,1]. For example, let V be any r.v. playing the role of an identity function so that $p(V \text{ is in } a) = p(a)$, for all $a\in P(D)$. Then for T being strictly increasing, using Theorem 4.6(i),

$$p((f/g)_T) \overset{d}{=} p(V \text{ is in } (f/g)_T) \overset{d}{=} \lambda_0(E_V(((f/g)_T(V)))$$
$$= \lambda_0(\alpha\cdot[0,1] + (1-\alpha)\cdot\beta)$$
$$= \lambda_0[(1-\alpha)\cdot\beta\ ,\ (1-\alpha)\cdot\beta+\alpha\ ]$$
$$= \beta\ , \qquad (4.38)$$

where

$$\alpha \overset{d}{=} p(g(V)=0);\ \beta \overset{d}{=} E((f(V)/g(V))_T|g(V)>0). \quad (4.39)$$

See again [13], section 7.5 for previous related results. ■

5 Combining Information

With all of the major results set in place, some applications to combining evidence will be considered next. First, the previously introduced diagonalization procedure will be reviewed and a new loss function further justifying its use will be presented. This will be followed by an illustrative example and related concepts.

5.1 Diagonalization

A key element in combining evidence, from this author's viewpoint, is diagonalization of information, mentioned briefly in item (3) of the list of new mathematical techniques developed for analyzing and treating this problem given in the Introduction. This procedure, originally introduced in [3] with additional properties presented in [12], section 2.2, is related to (but distinct from) the logarithmic pooling procedure [52]. In brief, let $f_j:D\to\mathbb{R}^+$ represent the j^{th} source's description of a common but unknown parameter of interest $\theta\in D\subseteq\mathbb{R}^m$, where f_j is a pdf or pf (probability density or probability function), known, and $f:D^n\to\mathbb{R}^+$ is a constructed known joint pdf or pf, whose j^{th} marginal is f_j, $j=1,\ldots,n$. Often, with no other information present, appealing e.g. to the maximal entropy principle, f_j can be considered statistically independent and, hence, f as a product of the f_j in different arguments. Then, $diag(f):D\to\mathbb{R}^+$ is given, for any $x\in D$, using the pdf form,

$$diag(f)(x) \overset{d}{=} f(x,\ldots,x)/c_f\ , \qquad (5.1)$$

where

$$c_f \overset{d}{=} \int_{x\in D} f(x,\ldots,x)dx. \qquad (5.2)$$

It was shown in [3] (see also [12]) that diag has quite a number of desirable properties, including: extending optimal estimation and regression, such as gaussian linear regression; for the independent source case, being a symmetric, associative, bayesian updating-invariant, related to surface integral representation for probaility measures conditioned on particular regions, namely here

$$\text{diag}(D) = \{(x,\ldots,x): x \in D\} \subseteq D^n . \quad (5.3)$$

However, despite all of the above justification for use, no basic loss function emerged for deriving use of the procedure until recently [49]. This is summarized below:

Let n be any positive integer and choose any weights w_i, λ_j such that

$$0 \le w_i, \lambda_j \le 1 \;;\; 1 = w_1 + w_2 = \lambda_1 + \cdots + \lambda_n . \quad (5.4)$$

Let $f: D^n \to \mathbb{R}^+$ be the joint pdf describing the joint density of the n sources describing $\theta \in D$ and let $h: D^n \to D$ be any procedure which reduces the joint description space to the space of the parameter itself. Then, define the following loss function $L(h)$ representing an expected weighted combination of squared distances between any given value of $h(\underline{x})$ and the possible components of $\underline{x}$, a direct measure of fit between $h(\underline{x})$ and f, plus a weighted amount of the dispersion of f interacting multiplicatively with the square of $h(\underline{x})$:

$$L(h) \overset{d}{=} \int_{\underline{x} \in D^n} \Big(\big(w_1 \cdot \sum_{j=1}^{n} \lambda_j \cdot (x_j - h(\underline{x}))^2 + w_2 \cdot \sum_{(1 \le i < j \le n)} (x_i - x_j)^2 h(\underline{x})^2 \big) f(\underline{x}) d\underline{x} . \quad (5.5)$$

Theorem 5.1 (new result [49])

(i) $\inf_{\text{over all } h: D^n \to D} L(h)$ occurs for $h = h_0$,

where for any $\underline{x} \in \mathbb{R}^n$,

$$h_0(\underline{x}) \overset{d}{=} \begin{cases} x, & \text{if } x_1 = \cdots = x_n \in D , \\ w_1 \cdot \sum_{j=1}^{n} \lambda_j x_j / (w_2 \cdot \sum_{(1 \le i < j \le n)} (x_i - x_j)^2) , & \\ & \text{if } x_1,\ldots,x_n \text{ are otherwise } \in D. \end{cases} \quad (5.6)$$

$$\text{(ii)}\; h_1(\underline{x}) \overset{d}{=} \lim_{\substack{w_1 \to 0 \\ w_2 \to 1}} h_0(\underline{x}) = \begin{cases} x , & \text{if } x_1 = \cdots = x_n = x \in D \\ 0 , & \text{if otherwise.} \end{cases} \quad (5.7)$$

(iii) The pdf of $h_1(\underline{x})$ for $\underline{x}$ having pdf f is diag(f), as given in eqs.(5.1),(5.2). ■

5.2 A Generic Example Illustrating Some Application of the Previous Theory

Consider, for simplicity, two different sources of information to be present, both attempting to estimate common unknown parameter $\theta \in D \subseteq \mathbb{R}$, with D a known set. One or both of the sources may be linguistic-based, with the other, if any, stochastic, and similary for conditional information vs unconditional. For simplicity here, modal or non-indicative forms will be omitted. See [19] for a conditional event reduction of such forms.

Case 1. Both sources provide stochastic descriptions in the form of unconditional r.v.'s.

Let $V_1, V_2: \Omega \to D$, (Ω, A, p) a fixed probability space, so that for any $a_1, a_2 \in B$, one can compute the joint probability

$$p(V_1 \text{ in } a_1 \;\&\; V_2 \text{ in } a_2) = p(V_1^{-1}(a_1) \cdot V_2^{-1}(a_2)). \quad (5.8)$$

Then, apply diagonalization via the restriction

$$a_1 = a_2 = \{x\},\; x \in D \quad (5.9)$$

to obtain the desired result. This is the direct application case for diagonalization.

Case 2. Both sources provide stochastic descriptions in conditional r.v. form.

Let $(V_j | W_j)$ be a conditional r.v., where $V_j: \Omega \to D$ and $W_j: \Omega \to B_j \subseteq \mathbb{R}^{m_j}$, $j = 1,2$, for (Ω, A, p) a fixed probability space, so that for any $a_j \in D$, $b_j \in B_j$, $j = 1,2$, one can compute the joint probability $p(([V_1|W_1] \text{ in } [a_1|b_1]) \& ([V_2|W_2] \text{ in } [a_2|b_2]))$ (see Theorem 3.10). Then, keeping in mind Theorem 3.11 and the following Remark, one can apply diag to obtain $\text{diag}(f_{\underline{W}^{-1}(\underline{b})})$, $\underline{b} = (b_1, b_2)$, etc., constraining the a_i as in (5.9). In turn, one can select some reasonable distribution of $\underline{b}$ through $p \circ \underline{W}^{-1}$ - such as singleton restriction - and obtain finally $E_{\underline{W}}(\text{diag}(f_{\underline{W}^{-1}(\cdot)}))$.

In practise, spaces B_1, B_2 can represent auxiliary attributes serving as antecedents in the descriptions of θ, as given in the example in the Introduction (Sources 3 and 4).

Case 3. Source 1 is linguistic-based and provides fuzzy set membership function $g: D \to [0,1]$, while source 2 is stochastic, corresponding to unconditional r.v. $V: \Omega \to D$.

Apply the principle of identifying g with nested random set $S_g(U): \Omega \to P(D)$, by invoking Theorem 4.3 and the ensuing Remark. It is reasonable here to assume r.v. U over $[0,1]$ is statistically independent of V, and hence $S_g(U)$ and V are independent. Furthermore, in order to apply diag in a non-trivial way, either V as a singleton point-valued function should be "brought up" to the typical range level of $S_g(U)$, or $S_g(U)$ restricted down to the point values of V (all with suitable normalization of values).

In the case of the former, one most naturally can replace V by $\hat{V}: \Omega \to P(D)$, where $\hat{V}$ is a random subset of D with -assuming D finite - pf given as

$$p(\hat{V} = a) = p(V \text{ in } a) / \sum_{b \in P(D)} p(V \text{ in } b) , \text{ all } a \in P(D). \quad (5.10)$$

In the case of the latter, one could consider normalized one point coverage probabilities, i.e. normalized fuzzy set membership values, thus diag here yields $\text{diag}(g / \sum_{x \in D} g(x) , p \circ V^{-1})$.

Other cases which can be treated in a similar way include:

Case 4. Both sources are linguistic-based: Use again the principle based on Theorem 4.3.

Case 5. One or both sources are conditional linguistic: Use the principle of identification with ordinary conditional random sets, via Theorem 4.7 and the ensuing comments, and then apply a procedure similar to Case 2.

More generally, consider several information sources s_j, $j=1,\ldots,n$, being present, some possibly providing linguistic-based descriptions, others, stochastic ones, some (either linguistic or stochastic) conditional, others unconditional (either linguistic or stochastic). One then converts each linguistic source, say s_j, $j=1,\ldots,m$, to stochastic form as a nested random set $S(s_j)$, in either unconditional form (when s_j is in unconditional form) based on the ideas of Section 4, especially Theorem 4.3, or in conditional form (when s_j is in conditional form) based on Theorem 4.7. Since the conditional case includes the unconditional, one can assume, without loss of generality, all $S(s_j)$ are represented as conditional random sets.

Following this, utilizing either Theorem 3.11 or a variation, based on the logical combination desired (such as combinations of conjunctions, disjunctions, negations), a single joint conditional probability distribution $F(S(\underline{s}))$ is obtainable with corresponding pdf or pf, say, $f(S(\underline{s}))$, describing the situation of interest - e.g., the unknown parameter θ. Next, applying the diagonalization transform yields $\mathrm{diag}(f(S(s)))$, a single pdf or pf, representing the combined description. Thus, it is reasonable to replace any decision problem based on the initial collection of sources $\underline{s}$, by one based on $\mathrm{diag}(f(S(\underline{s})))$. Hence, standard probability models apply to the choice of optimal decisions, decision errors, and information bounds.

However, it is still of interest to determine what loss of information occurs between $f(S(s))$ and $\mathrm{diag}(f(S(\underline{s})))$, and to compare the gain $f(S(\underline{s}))$ produces compared to prior knowledge of the situation. Analysis of these issues will be forthcoming in a later publication.

Acknowledgments

The author wishes to express his appreciation to the following individuals and offices for their continued support for various aspects of this work: Office of Naval Research (Dr. Ralph Wachter, Code 1133), Independent Research Program at NRaD (Dr. Alan Gordon, Code 0141), Basic Research Group, Technical Panel for C^3, Joint Directors of Laboratories (Michael C. Mudurian, Code 422, NRaD); and the line management at NRaD - Code 42 (John A. Salzmann, Jr.) and Code 421 (Sunny Conwell).

References

1. Goodman, I.R., "A general theory for the fusion of data", *Proc. DFS-87* (1987 Tri-Service Data Fusion Symp., at Applied Phys. Lab., Laurel, MD, June 9-11, 1987), pp. 254-270.

2. Goodman, I.R., "Applications of a conditional event algebra to data fusion", *Proc. DFS-88* (1988 Tri-Service Data Fusion Symp., at Applied Phys. Lab., Laurel, MD, May 17-19, 1988), pp. 179-187.

3. Goodman, I.R., "Applications of a conditional event algebra to data fusion: Part 2", *Proc. DFS-89* (1989 Tri-Service Data Fusion Sympos. at Applied Physics Lab., Laurel, MD, May 16-18, 1989), pp. 181-193.

4. Goodman, I.R. "Extension of the PACT algorithm to a conditional event algebra format", *Proc. DFS-90* (1990 Joint Service Data Fusion Symp. held at Applied Phys. Lab., Laurel, MD, May 15-18, 1990), pp. 275-288.

5. Goodman, I.R., "Modeling and fusing of linguistic, probabilistic, and/or conditional and unconditional forms", hardcopy of invited tutorial, *Proc. DFS-91, Vol. III* (1991 Joint Service Data Fusion Symposium, at Applied Phys. Lab., Laurel, MD, Oct 7-11, 1991).

6. Goodman, I.R., "Combination of evidence in C^3 systems", *Proc. 8th MIT/ONR Wrkshp C^3 Sys* (at MIT, Cambridge, MA, June 24-28, 1985), pp. 161-166.

7. Goodman, I.R., "A probabilistic/possibilistic approach to modeling of C^3 Systems", *Proc. 9th MIT/ONR Wrkshp C^3 Syeems* (at Naval Postgrad. Schl, Monterey, CA, June 2-5, 1986), pp. 53-58.

8. Goodman, I.R., "A probabilistic/possibilistic approach to modeling C^3 systems: Part 2", *Proc. 1987 C^2 Research Symp* (at Nat. Def. Univ., Wash., D.C., June 16-18, 1987), pp. 41-48.

9. Goodman, I.R., "Toward a general theory of C^3 processes", *Proc. 1988 Symp. C^2 Research* (at Naval Postgrad. Schl., Monterey, CA, June 7-9, 1988), pp. 92-105.

10. Goodman, I.R., "Toward a general theory of C^3 processes: Part 2", *Proc. 1989 Symp C^2 Research* (at Nat. Def. Univ., Wash., D.C., June 27-29, 1989), pp. 57-67.

11. Goodman, I.R., "Applications of an exact linearization-gaussian sum technique to the modeling of C^3 nodes", *Proc. 1990 Symp C^2 Res.* (at Naval Postgrad. Schl., Monterey, CA, June 12-14, 1990), pp. 63-72.

12. Goodman, I.R., "Development and application of some new procedures to the modeling and combining of linguistic and probabilistic information for C^3 systems", *Proc. 1991 Symp. C^2 Research* (at Nat.Def.Univ., Wash., D.C., June 24-25, 1991), pp. 179-190.

13. Goodman, I.R., Nguyen, H.T. & Walker, E.A., *Conditional Inference and Logic for Intelligent Systems: A Theory of Measure-Free Conditioning*, North-Holland, Amsterdam, 1991.

15. Goodman, I.R., "Evaluation of combinations of conditioned information: a history", *Inform. Sci., 57-58*, Sept-Dec, 1991, pp. 79-110.

14. Goodman, I.R., Gupta, M.M., Nguyen, H.T., & Rogers, G.S. (eds.), *Conditional Logic in Expert Systems*, North-Holland, Amsterdam, 1991.

16. Goodman, I.R. & Nguyen, H.T., *Uncertainty Models for Knowledge-Based Systems*, North-Holland, Amsterdam, 1985.

17. Goodman, I.R., "Algebraic and probabilistic bases for fuzzy sets and the development of

fuzzy conditioning", in [15],1991, pp. 1-69.

18. Goodman, I.R. & Stein,D.W.,"Extension of the measure-free approach to conditioning of fuzzy sets and other logics",*Proc.3rd Int.Fuzzy Sys.Analysis World Congress*, Univ. of Wash., Seattle, Aug., 1989, pp. 361-364.

19. Goodman, I.R., "Identification of compound modal forms with conditional events, with applications to the combination of evidence problem", to be submitted.

20. Copeland, A.H., "Implicative boolean algebras", *Math.Zeitschr.,53(3)*, 1950, pp. 285-290.

21. Copeland, A.H. & Harary,F., "The extension of an arbitrary boolean algebra to an implicative boolean algebra", *Proc.Amer.Math.Soc.,4*, 1953, pp. 751-758.

22. Calabrese, P.G., "The probability that p implies q", Preliminary report, *Amer.Math.Soc. Notices,22*, 1975, pp. A430 A431.

23. Calabrese, P.G., "An algebraic synthesis of the foundations of logic and probability", *Inform.Sci.,42*, 1987, pp. 187-237.

24. Popper, K., *Conjectures and Refutations*, Routledge and Kegan Paul, London, 1963, p.390, formula 22.

25. Lewis, D., "Probabilities of conditionals and conditional probabilities", *Philos.Rev.,85*, 1976, pp. 297-315.

26. Schay, G., "An algebra of conditional events", *Journ.Math.Analysis & Applic., 24*, 1968,pp. 334-344.

27. Adams, E.W., *The Logic of Conditionals*, D. Reidel Co., Dordrecht, Holland, 1975,pp.46-47.

28. Hailperin,T., *Boole's Logic and Probability*, North-Holland, Amsterdam, 1976. (Also, expanded Second Edition, 1986.)

29. Boole, G., *The Laws of Thought*, Walton & Maberly, London, 1854 (reprinted by Dover Co., New York, 1958), especially Chapters 5,6 et passim.

30. Goodman, I.R., "A measure-free approach to conditioning", *Proc.Third AAAI Wrkshp.Uncert. AI*, Univ. of Wash., Seattle, July, 1987, pp. 270-277.

31. Igoshin, V.V., "An algebraic characterization of interval lattices", Translation of *Uspekhi Mat.Nauk.,40(3)*, May-June, 1985,pp.233-234.

32. Igoshin, V.V., "Identities in interval lattices of lattices",*Colloq.Math.Soc.J.Bolyai,33,Contributions to Lattice Theory*, Szeged, Hungary, 1980, pp. 491-501.

33. Goodman, I.R., "Three-valued logics and conditional event algebra", *Proc.First Inter. Symp.on Uncert. Model & Applic.*, Univ. of MD, Dec, 1990, pp. 31-37.

34. Goodman, I.R., "New axiom systems characterizing some proposed conditional event algebras and corresponding three-valued logic", to be submitted.

35. Goodman, I.R., "Extension of the canonical isomorphism theorem between three-valued logics and conditional event algebras to probability evaluations and the development of a sample space / frequency theory for conditional events", to be submitted.

36. Goodman, I.R., "Non-boolean functional image extensions to conditional event algebra and the development of joint conditional random variables with differing antecedents", to be submitted.

37. Goodman, I.R., "Preservation of Stone algebra property in forming all higher order interval algebras, upper and lower approximations of conditional events, and the closure problem", to be submitted.

38. Goodman,I.R., "GNW conditional event algebra as a non-equational logic and a theory of partial deducts", to be submitted.

39. Goodman, I.R., "The canonical normal disjunctive form for boolean operators has a natural extension to conditional event algebra", to be submitted.

40. Goodman, I.R., "A numerically-based approach to the development of conditional events and fuzzy conditional events, utilizing extended numerical division", to be submitted.

41. Dubois,D. & Prade,H., *Fuzzy Sets & Systems*, Academic Press, New York, 1980.

42. Sklar,A.,"Random variables,joint distributions and copulas,*Kybernetika,9*(Czech),1973, 449-460.

43. Dall'Aglio,G.,Kotz,S. & Salinetti,G.(eds), *Advances in Probability Distributions with Given Marginals*,Kluwer,Dordrecht,Holl.,1991.

44. Hohle,U.,"Editorial",special vol.on math.of fuzzy sets,*Fuzzy Sets & Sys.,40*,1991,253-256.

45. Goodman,I.R. & Bier,V.M., "A re-examination of the relationship between fuzzy set theory and probability theory",lecture,*29th Annual Bayesian Research Conf.*,U.S.C.,Feb.15,1991.

46. Goodman,I.R.,"Zadeh's (min,max,1-()) fuzzy logic as the unique probability-compatible one", to be submitted.

47. Goodman,I.R.,"A new result supporting the nested random set model as the unique probability-compatible interpretation of linguistic based information in combination of evidence", lecture,*30th Annual Bayesian Research Conf.*, Univ.S.Cal.,Los Angeles, Feb. 13, 1992.

48. Frank,M.J., "On the simultaneous associativity of F(x,y) and x+y-F(x,y)",*Aeq.Math,19*,1979,194-226.

49 Goodman,I.R., "A new formulation of the diagonalization technique for combining probabilities", to be submitted.

50. Goodman,I.R.,Nguyen,H.T. & Rogers, G.S.,"On the scoring approach to admissibility of uncertainty measures in expert systems", *Journ. Math. Analysis & Applic.,159(2)*,1991,550-594.

51. Rescher,N., *Many-valued Logic*, McGraw-Hill, New York, 1969.

52. Genest,C. & Zidek,J.V., "Combining probability distributions: a critique and an annotated bibliography", *Statis. Sci.,1(1)*,1988,114-148.