Machine Intelligence and Pattern Recognition

Volume 9

Series Editors

L.N. KANAL

and

A. ROSENFELD

University of Maryland
College Park, Maryland, U.S.A.

NORTH-HOLLAND
AMSTERDAM • NEW YORK • OXFORD • TOKYO

Uncertainty in Artificial Intelligence 4

Edited by

Ross D. SHACHTER
Stanford University
Stanford, California, U.S.A.

Tod S. LEVITT
Advanced Decision Systems
Mountain View, California, U.S.A.

Laveen N. KANAL
University of Maryland
College Park, Maryland, U.S.A.

John F. LEMMER
Knowledge Systems Concepts
Rome, New York, U.S.A.

1990

NORTH-HOLLAND
AMSTERDAM • NEW YORK • OXFORD • TOKYO

ELSEVIER SCIENCE PUBLISHERS B.V.
Sara Burgerhartstraat 25
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

Distributors for the United States and Canada:
ELSEVIER SCIENCE PUBLISHING COMPANY, INC.
655 Avenue of the Americas
New York, N.Y. 10010, U.S.A.

ISBN: 0 444 88650 8 (hardbound)
0 444 88737 7 (paperback)

Printed in The Netherlands

PREFACE

The papers in this volume clearly illustrate the current relationship between Uncertainty and Artificial Intelligence. They show that while much has been accomplished, much remains to be done to fully integrate Uncertainty Technology with AI.

It has been said that the research in Artificial Intelligence revolves around five basic questions asked relative to some particular domain: (1) what knowledge is required, (2) how can this knowledge be acquired, (3) how can it be represented in a system, (4) how should the knowledge be manipulated in order to provide intelligent behavior, and (5) how can the behavior be explained. The early volumes in this series concentrated on manipulation of the uncertainties associated with other knowledge in a system and often implied, at least in a general way, associated data structure.

In this volume we can observe all five of the fundamental questions of AI being addressed. Some of the papers address all five questions. Other papers address only one or several of the questions, with intelligent manipulation still being the most popular.

From the perspective of the relationship of uncertainty to the basic questions of Artificial Intelligence, this volume divides rather naturally into four sections which highlight both the strengths and weaknesses of the current state of the relationship between Uncertainty Technology and Artificial Intelligence.

The first section contains papers describing paradigms that seem to be on the same level as the Expert System Paradigm. It is in this sense that these papers address, at least implicitly, all five basic questions. In most cases the papers themselves do not take such a bold stance, but it is difficult to not understand them in this sense. All these papers seem to use the notion of causality as an organizing principle in much the same way that Expert Systems use heuristic knowledge as an organizing principle.

The second, and by far the largest section addresses specific means of representing and intelligently manipulating uncertainty information. How these representations and manipulations are to be integrated into an intelligent system is left largely to the creativity of the reader. This section naturally divides into two parts: one on manipulating uncertainties, and a second which evaluates or compares one or more representation/manipulations technologies. These papers range from ones which are specifically presented as possible solutions to well known AI problems (e.g. the problem of monotonicity) through improved computational techniques to philosophical discussions of the possible meanings of uncertainty.

The third, and regrettably the smallest of the sections (containing only two papers!) addresses the basic questions of knowledge acquisition and explanation. These questions must be addressed if intelligent systems incorporating uncertainty are ever to be widely accepted. These questions seem to provide almost virgin territory for research. It may be significant that both of the papers in this section come from the group which did some of the earliest research in Expert Systems, the group which did the original MYCIN work.

The final section reports on applications of uncertainty technology.

We hope that readers of this volume will be encouraged to work on explicit connections between Uncertainty Technology and ongoing work in mainstream AI. There has been hopeful evidence that some AI researchers are looking to the uncertainty community for help in certain problem areas such as planning. In turn, uncertainty researchers must look for inspiration from the AI community if they are to avoid working on problems whose solution will have little or no value in the larger context of Artificial Intelligence.

Ross D. Shachter
Stanford, CA

Tod S. Levitt
Mt. View CA

Laveen N. Kanal
College Park, MD

John F. Lemmer
Rome, NY

CONTENTS

[14] A. Shimony. Coherence and the axioms of confirmation. *Journal of Symbolic Logic*, 20:8–20, 1955.

[15] B. Skyrms. *Causal Necessity*. Yale Univ. Press, New Haven, 1980.

[16] B. van Fraassen. Belief and the will. *Journal of Philosophy*, (81):235–256, 1984.

[17] B.C. van Fraassen. A temporal framework for conditionals and chance. In W. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs*, pages 323–340, D. Reidel, Dordrecht, 1980.

Uncertainty in Artificial Intelligence 4
R.D. Shachter, T.S. Levitt, L.N. Kanal, J.F. Lemmer (Editors)

A GENERAL NON-PROBABILISTIC THEORY OF INDUCTIVE REASONING

Wolfgang SPOHN

Institut für Philosophie
Universität Regensburg
8400 Regensburg, West Germany

1. INTRODUCTION

Probability theory, epistemically interpreted, provides an excellent, if not the best available account of inductive reasoning. This is so because there are general and definite rules for the change of subjective probabilities through information or experience; induction and belief change are one and same topic, after all. The most basic of these rules is simply to conditionalize with respect to the information received; and there are similar and more general rules.[1] Hence, a fundamental reason for the epistemological success of probability theory is that there at all exists a well-behaved concept of conditional probability.

Still, people have, and have reasons for, various concerns over probability theory. One of these is my starting point: Intuitively, we have the notion of *plain belief*; we believe propositions[2] to be true (or to be false or neither). Probability theory, however, offers no formal counterpart to this notion. Believing A is not the same as having probability 1 for A, because probability 1 is incorrigible[3]; but plain belief is clearly corrigible. And believing A is not the same as giving A a probability larger than some $1 - \varepsilon$, because believing A and believing B is usually taken to be equivalent to believing A & B.[4] Thus, it seems that the formal representation of plain belief has to take a non-probabilistic route.

Indeed, representing plain belief seems easy enough: simply represent an epistemic state by the set of all propositions believed true in it or, since I make the common assumption that plain belief is deductively closed, by the conjunction of all propositions believed true in it. But this does not yet provide a theory of induction, i.e. an answer to the question how epistemic states so represented are changed through information or experience. There is a convincing partial answer: if the new information is compatible with the old epistemic state, then the new epistemic state is simply represented by the conjunction of the new information and the old beliefs. This answer is partial because it does not cover the quite common case where the new information is incompatible with the old beliefs. It is, however, important to complete the answer and to cover this case, too; otherwise, we would not represent plain belief as corrigible. The crucial problem is that there is no good completion. When epistemic states are represented simply by the conjunction of all propositions believed true in it, the answer cannot be completed; and though there is a lot of fruitful work, no other representation of epistemic states has been proposed, as far as I know, which provides a complete solution to this problem.

In this paper, I want to suggest such a solution. In [4], I have more fully argued that this is the only solution, if certain plausible desiderata are to be satisfied. Here, in section 2, I will be content with formally defining and intuitively explaining my proposal. I will compare my proposal with probability theory in section 3. It will turn out that the theory I am proposing is structurally homomorphic to probability theory in important respects and that it is thus equally easily implementable, but moreover computationally simpler. Section 4 contains a very brief comparison with various kinds of logics, in particular conditional logic, with Shackle's functions of potential surprise and related theories, and with the Dempster - Shafer theory of belief functions.

2. THE THEORY

The algebraic framework has to be settled first. Let W be some non-empty set of possibilities (possible worlds, possible courses of events, or what have you). Propositions, denoted by $A,B,C,...$, are represented simply by subsets of W. Subfields of the field of all propositions will be denoted by $\mathcal{A},\mathcal{B},\mathcal{C},...$[5] Usually, W will have a structure: there will be a family $(W_i)_{i\in I}$ of variables or factors - where I is some index set and each W_i ($i\in I$) is some non-empty set - such that $W = \Pi_{i\in I} W_i$.[6] That is, each $w\in W$ is a function defined on I with $w_i\in W_i$ for all $i\in I$ and thus represents one way how all the variables may get realized. In many physical applications, e.g., each W_i will be identical to the state space and I to the real time axis. For each $J \subseteq I$, $\mathcal{A}_J$ is to be the field $\{A \mid$ for all $w,w'\in W$, if $w_i = w'_i$ for all $i\in J$, then $w\in A$ iff $w'\in A\}$ of all propositions referring at most to the variables in J.

The central concept is now easily defined (and afterwards explained):

Definition 1: Let $\mathcal{A}$ be a field of propositions. Then κ is an *$\mathcal{A}$-measurable natural conditional function ($\mathcal{A}$-NCF)* iff κ is a function from W into the set N of natural numbers such that $\kappa(w) = 0$ for some $w\in W$ and $\kappa(w) = \kappa(w')$ for all atoms[7] A of $\mathcal{A}$ and all $w,w'\in A$.[8] Moreover, we define for each non-empty $A\in \mathcal{A}$: $\kappa(A) = \min\{\kappa(w) \mid w\in A\}$.[9]

The measurability condition is quite obvious; it requires that an $\mathcal{A}$-NCF does not discriminate possibilities which are not discriminated by the propositions in $\mathcal{A}$.

The crucial question, however, is how to interpret an NCF as an epistemic state. The most accurate answer is to say that an NCF κ represents a grading of disbelief: a possibility w with $\kappa(w) = 0$ is not disbelieved at all in κ; if $\kappa(w) = 1$, w is disbelieved to degree 1 in κ; etc. This means that all possibilities w with $\kappa(w) > 0$ are believed in κ not to obtain; i.e., the true possibility is believed in κ to be in $\kappa^{-1}(0) = \{w \mid \kappa(w) = 0\}$; and hence the stipulation of Definition 1 that $\kappa^{-1}(0) \neq \emptyset$. A proposition A is believed true in κ iff the true possibility is believed in κ to be in A, i.e. iff $\kappa^{-1}(0) \subseteq A$, i.e. iff $\kappa(-A) > 0$.[10] Thus, the set of propositions believed true in κ is deductively closed; and it is consistent, because $\kappa^{-1}(0) \neq \emptyset$. Note that $\kappa(A) = 0$ only means that A is not believed false in κ; this is compatible with $\kappa(-A) = 0$, i.e. with A also not being believed true in κ.

One may also talk of integer-valued degrees of firmness of belief, i.e. one may define that A is believed with firmness m in κ iff either $\kappa(A) = 0$ and $\kappa(-A) = m$ or $\kappa(A) = -m > 0$. Thus, A is believed to be true or false iff, respectively, A is believed with positive or negative firmness. This firmness-of-belief function is intuitively easier to grasp because it does not require thinking in negative terms; but it is formally less well-behaved, and the theorems would not look so simple. Therefore I prefer to stick to NCFs.

These explanations well agree with two simple consequences of Definition 1:

Theorem 1: Let κ be an $\mathcal{A}$-NCF. Then we have:

(1) for each contingent[11] $A\in \mathcal{A}$, $\kappa(A) = 0$ or $\kappa(-A) = 0$ or both,

(2) for all non-empty $A,B\in \mathcal{A}$, $k(A\cup B) = \min\{\kappa(A),\kappa(B)\}$.

(1) is the fundamental NCF-law for negation, saying that not both A and $-A$ can be disbelieved. (2) is the fundamental NCF-law for disjunction: It is obvious that $A\cup B$ should be believed at least as firmly as A and B. But $A\cup B$ cannot be believed more firmly than both A and B; otherwise, it might happen that both A and B are disbelieved, though $A\cup B$ is not. In order to discover a fundamental NCF-law for conjunction, we have to look at conditional NCF-values.

This brings up the crucial question how epistemic states represented by NCFs are changed through information or experience. Two plausible assumptions provide a complete answer. The first assumption is that, if the information immediately concerns only the proposition A and nothing else, then neither the grading of disbelief within A, nor that within $-A$ are changed by that information. We define:

Definition 2: Let κ be an $\mathcal{A}$-NCF and A a non-empty proposition in $\mathcal{A}$. Then, *the A-part of* κ is to be that function $\kappa(.|A)$ defined on A for which $\kappa(w|A) = \kappa(w) - \kappa(A)$ for all $w\in A$. If $B\in \mathcal{A}$ and $A\cap B \neq \emptyset$, we also define $\kappa(B|A) = \min\{\kappa(w|A) \mid w\in A\cap B)\} = \kappa(A\cap B) - \kappa(A)$.

The first assumption thus says that an information immediately concerning only A leaves the A-part as well as the $-A$-part of κ unchanged, i.e. its effect can only be that these two parts are shifted in relation to one another. Definition 2, by the way, already contains the fundamental NCF-law for conjunction:

Theorem 1 (cont.):

(3) for all compatible $A,B\in \mathcal{A}$, $\kappa(A\cap B) = \kappa(A) + \kappa(B|A)$.

The second assumption is that information about A may come in various degrees of firmness; seeing A usually informs about A much more firmly than being told about A by some more or less reliable person. Thus, the firmness with which an information is embedded in an epistemic state cannot be fixed once and for all, but has to be conceived as a parameter of the information process itself. In view of the first assumption, this parameter completely determines belief change:

Definition 3: Let κ be an $\mathcal{A}$-NCF, A a contingent proposition in $\mathcal{A}$, and $m \in N$. Then *the A,m-conditionalization* $\kappa_{A,m}$ *of* κ is defined as that $\mathcal{A}$-NCF for which $\kappa_{A,m}(w) = \kappa(w|A)$, if $w \in A$, and $\kappa_{A,m}(w) = m + \kappa(w|\text{-}A)$, if $w \in \text{-}A$.

In the A,m-conditionalization of κ, only the A-part and the -A-part of κ are shifted in relation to one another, and A is believed with firmness m, as specified by the conditionalization parameter.

This account of belief change may be generalized. The information may immediately concern not only a single proposition, but a whole field $\mathcal{B}$ of propositions. The parameter characterizing the information process then consists not in a single number, but in a whole $\mathcal{B}$-NCF λ. And belief change is then defined in the following way:

Definition 4: Let κ be an $\mathcal{A}$-NCF, $\mathcal{B}$ a subfield of $\mathcal{A}$, and λ a $\mathcal{B}$-NCF. Then *the λ-conditionalization* κ_λ *of* κ is defined as that $\mathcal{A}$-NCF for which for all atoms B of $\mathcal{B}$ and all $w \in B$ $\kappa_\lambda(w) = \lambda(B) + \kappa(w|B)$.

In the λ-conditionalization of κ, $\kappa_\lambda(B) = \lambda(B)$ for all $B \in \mathcal{B}$, and only the B-parts of κ, for all atoms B of $\mathcal{B}$, are shifted in relation to each other. Definition 4 corresponds to Jeffrey's much discussed generalized probabilistic conditionalization; cf. [1], ch. 11.

It is to be expected that a workable concept of independence goes hand in hand with this account of conditionalization. This is indeed the case. The following definition is straightforward:

Definition 5: Let κ be an $\mathcal{A}$-NCF and $\mathcal{B}$ and $\mathcal{C}$ two subfields of $\mathcal{A}$. Then $\mathcal{B}$ and $\mathcal{C}$ are *independent with respect to* κ iff for all non-empty $B \in \mathcal{B}$ and $C \in \mathcal{C}$ $\kappa(B \cap C) = \kappa(B) + \kappa(C)$. Furthermore, $\mathcal{B}$ and $\mathcal{C}$ are *independent conditional on* the proposition D *w.r.t.* κ iff for all non-empty $B \in \mathcal{B}$ and $C \in \mathcal{C}$ $\kappa(B \cap C|D) = \kappa(B|D) + \kappa(C|D)$. If $\mathcal{D}$ is a further subfield of $\mathcal{A}$, then $\mathcal{B}$ and $\mathcal{C}$ are *independent conditional on* $\mathcal{D}$ *w.r.t.* κ iff $\mathcal{B}$ and $\mathcal{C}$ are independent conditional on all atoms D of $\mathcal{D}$ w.r.t. κ. Finally, these definitions are specialized to two contingent propositions B and C by taking $\mathcal{B}$ as $\{\emptyset, B, \text{-}B, W\}$ and $\mathcal{C}$ as $\{\emptyset, C, \text{-}C, W\}$.

How do all the concepts so defined behave? This may not be immediately perspicuous, but the next section will provide a surprisingly powerful answer.

3. A COMPARISON WITH PROBABILITY THEORY

The basic definitions and formulae in the previous section look very similar to those in probability theory; we only seem to have replaced the sum, multiplication, and division of probabilities by, respectively, the minimum, addition, and subtraction of NCF-values. In order to see that this is no accident, we have to move for a moment into the context of non-standard arithmetics and non-standard probability theory:

Theorem 2: Let $\mathcal{A}$ be a finite field of propositions. Then, for any non-standard $\mathcal{A}$-NCF[12] κ and for any infinitesimal z there is a non-standard probability measure P such that for all $A,B \in \mathcal{A}$ $\kappa(B|A) = n$ iff $P(B|A)$ is of the same order as z^n (i.e. $P(B|A)/z^n$ is finite, but not infinitesimal). In particular we have: whenever $P(C) = P(A) + P(B)$, then $\kappa(C) = \min\{\kappa(A), \kappa(B)\}$; whenever $P(C) = P(A)\,P(B)$, then $\kappa(C) = \kappa(A) + \kappa(B)$; $\kappa(B|A) = \kappa(A \cap B) - \kappa(A)$, as desired; and whatever is (conditionally) independent w.r.t. P, is so also w.r.t. κ.

Sketch of proof: Define P in the following way: for each atom A of $\mathcal{A}$ with $\kappa(A) = n > 0$ let $P(A) = z^n$, and distribute the rest equally among the other atoms of $\mathcal{A}$ so that the probabilities of all atoms of $\mathcal{A}$ sum up to 1. The claims of Theorem 2 are then easily checked; they in particular turn on the fact that, if x is of the same order as z^m and y of the same order as z^n, then xy is of the same order as z^{m+n} and $x + y$ is of the same order as $z^{\min(m,n)}$.

It is thus not surprising that the laws of the concepts introduced in the previous section are simply translations of the laws of the corresponding probabilistic concepts. For instance, the theorem of total probability translates into this (where $A_1,\ldots,A_s$ partition W):

(4) $\kappa(B) = \min_{r \leq s} [\kappa(A_r) + \kappa(B|A_r)]$.

Bayes' theorem yields this (with $A_1,\ldots,A_s$ as before):

(5) $\kappa(A_q|B) = \kappa(A_q) + \kappa(B|A_q) - \min_{r \leq s} [\kappa(A_r) + \kappa(B|A_r)]$.

Also, the probabilistic laws of independence and conditional independence hold for NCFs - e.g.:

(6) If A and C are independent w.r.t. κ, then B and C are independent w.r.t. κ iff $A \cup B$ and C are independent w.r.t. κ - provided that A and B are disjoint.

Without the proviso, (6) would not necessarily hold. And so on. Let me only mention the most important law concerning conditional independence of subfields. It says in terms of the factorization of W at the beginning of section 2, where J, K, and L are pairwise disjoint subsets of the index set I:

(7) If $\mathcal{A}_J$ is independent of $\mathcal{A}_K$ conditional on $\mathcal{A}_L$ and independent of $\mathcal{A}_L$ or independent of $\mathcal{A}_L$ conditional on $\mathcal{A}_K$ w.r.t. κ, then $\mathcal{A}_J$ is independent of $\mathcal{A}_{K \cup L}$ w.r.t. κ.[13]

These observations have a considerable import. For instance, the theory of probabilistic causation has turned out to be to a large extent a theory of conditional stochastic independence.[14] NCFs would thus allow to extend these ideas to a theory of deterministic causation.[15] In the present context, however, the crucial observation is that conditional independence is an important means for making probability measures computationally manageable. This carries over to the implementation of NCFs. In particular, the results and techniques related to such things as influence diagrams, Markov fields and trees, causal graphs, etc.[16] may be translated into NCF-theory. This is exemplified by [12] and [13]: In [12], Hunter achieves a way of parallel updating of NCFs by adapting methods of parallel probabilistic updating developed by Pearl in [11]; and in [13], Hunter shows that the results reported in [14] and [15] carry over to NCFs, i.e. that for NCFs, too, the conditional independencies implied by a given causal input list according to Definition 5, those derivable from that list by the axioms of semi-graphoids, and those implied by that list via Verma's criterion of d-separation are always the same.[17] Finally, Definition 4 suggests that the concept of a mixture may be meaningfully carried over from probability theory to the theory of NCFs and may there have fruitful applications.

Of course, there also are differences. On the one hand, NCFs are computationally simpler than probabilities; they have the advantage of formally representing the intuitively so important concept of plain belief; and it may be easier to elicit and implement the subjective judgments of experts in the coarser terms of NCFs. On the other hand, I presently do not see how NCFs would allow for a meaningful analogue to the theory of integration and expectation and thus for a useful decision theory (for which something like expected utility is essential). And most importantly, relative frequencies are so intimately tied to probabilities that I do not see how to reasonably deal with statistical data within an NCF-framework.

A final remark: I said in the introduction that plain belief in A cannot be probability 1 for A because of the incorrigibility of probability 1. This seems to be disproved by extensions of standard probability theory which allow for conditionalization by null propositions and thus render probability 1 corrigible. However, Popper measures, the best known extension of this kind, cannot account for iterated epistemic changes, as has already been observed in [16]. According to my diagnosis in [4], sect. 7, which is based on the investigation [17] into the formal structure of Popper measures, this failure can only be overcome by replacing Popper measures by, so to speak, probabilified NCFs. Thus, it seems that the probabilist cannot avoid considering NCFs as long as he takes plain belief seriously.

4. OTHER COMPARISONS

Though many have proposed non-probabilistic representations of epistemic states, I have, to my surprise, nowhere found the very structure described in section 2; aims and intuitions have presumably been different. But often, the importance of stating general and precise rules of belief change, which are tantamount to a theory of induction, has apparently not been clearly recognized; this will in any case be my standard criticism of the further comparisons pursued here.

4.1. *Various Logics*

The following strategy for modelling belief change has attracted many people: Suppose a language with a conditional $\rightarrow$ to be given; represent an epistemic state by a (consistent and deductively closed) set S of sentences of that language; and define the change S_A of S by information A as $S_A = \{B \mid A \rightarrow B \in S\}$.[18] Of course, this strategy crucially depends on the properties of $\rightarrow$. E.g., $\rightarrow$ must not be interpreted as material implication. Strict implication will do neither; all the conditionals in the various many-valued logics that have been proposed

are unsuited, too[19]; and even the conditionals of the variants of relevance logic seem to be unhelpful.[20] However, these remarks are not meant as a criticism, because all the conditionals mentioned were not designed for the present purpose.

Indeed, no monotonic conditional will be adequate for this strategy. The best conditional for this purpose is that of conditional logic (which has always been conceived to be non-monotonic in the sense that $A \rightarrow C$ does not entail $A \& B \rightarrow C$). Most semantics of conditional logic and corresponding models of belief change basically use orderings: orderings of propositions or of certain sets of propositions, well-orderings of possible worlds, and similar or equivalent things.[21] But they don't use numbers and their arithmetical properties. As I argue in [4], this is why these semantics and the corresponding models of belief change get problems with iterated belief change and cannot provide an equally adequate concept of (conditional) independence. Moreover, epistemic changes as defined in Definition 4 seem completely inaccessible to the whole strategy. Again, this is not a criticism of conditional logic, but only of the envisaged strategy of modelling belief change.[22]

4.2. *Plausibility Measures*

One of the first to propose formal alternatives to the beaten tracks of probability theory was Shackle with his functions of potential surprise most extensively presented in [25]. Such a function is a function y from the set of propositions into the closed interval [0,1] such that

(8) $y(\emptyset) = 1$,

(9) either $y(A) = 0$ or $y(-A) = 0$ or both,

(10) $y(A \cup B) = \min\{y(A), y(B)\}$.

(9) and (10) are identical with (1) and (2), and (8) arbitrarily fixes the maximal degree of potential surprise to be 1. Thus, Shackle's and my functions only differ in their ranges. This is not a mere technicality, however. There is reason to accept the generalization of (2) or (10) to countable unions (without weakening min to inf), and this forces the range of these functions to be well-ordered. Moreover, I have avoided a maximal degree of disbelief, because this maximal degree could not be changed according to all rules of belief change and would thus be incorrigible. Therefore, I do not want to allow the possibility accepted by Shackle that non-empty propositions have maximal potential surprise.

The main point, however, is that Shackle didn't get a grip on conditionalization. This is clear from his proposal

(11) $y(A \cap B) = \max\{y(A), y(B|A)\}$,

where he left $y(B|A)$ undefined.[23]

Similar remarks apply to the plausibility indexing which Rescher has proposed since 1964, e.g. in [26], and to Cohen's theory of inductive probability in [27] (which is not mathematical probability, but quite similar to NCFs). The works of these authors show the wide und fruitful applicability of non-probabilistic belief representation in many areas inside and outside philosophy.

4.3. *Dempster - Shafer*

In [28], p.224, Shafer shows that Shackle's theory is a special case of his: the function y is a degree of doubt derived from a consonant belief function in the sense of Shafer iff it satisfies (8) - (10). Since Dempster's rule of combination governs belief change for Shafer's belief functions in general, it may be expected to complete Shackle's theory. It indeed does, but in a different way than I did in section 2:

According to [28], pp.43 + 66f., there are also conditional degrees of doubt given by the formula

(12) $y(B|A) = [y(A \cap B) - y(A)] / [1 - y(A)]$.

Apart from the denominator, this looks like my Definition 2. However, $y(.|A)$ here represents the degree of doubt which results from combining the old belief function with the belief function Bel defined by: $Bel(B) = 1$, if $A \subseteq B$, and $Bel(B) = 0$ otherwise; and this function makes A incorrigibly certain, according to Shafer's theory. Thus, one should rather know how Shafer processes evidence which makes A less than incorrigibly certain, since this is what the above Definition 3 accomplishes. Shafer does this by combining the old belief function with some belief function Bel_s defined by: $Bel_s(B) = 1$, if $B=W$, $Bel_s(B) = s$, if $A \subseteq B \neq W$, and $Bel_s(B) = 0$ otherwise ($0<s<1$). (In a sense, s corresponds to the m of Definition 3.) But now the problem arises that, if the old belief function is consonant, its combination with Bel_s will in general not be consonant; this is easily checked. Thus, my conditionalizations of NCFs move within the set of all NCFs, whereas the set of all consonant belief functions in Shafer's sense is not closed with respect to Dempster's rule of combination. This entails that the NCF-theory presented here cannot be covered by the Dempster - Shafer theory of belief functions.

However, in [5] Shenoy gains a more positive perspective. He proposes a different rule of combination for NCFs which gives an account of belief change equivalent to the conditionalizations given by Definitions 3 and 4. Moreover, he defines marginalization for NCFs and shows that marginalization and combination thus explained obey the axioms presented in [29]. This means that the general scheme of local computation developed in [29] can also be applied to NCFs.[24]

NOTES

[1] Most notably Jeffrey's generalized conditionalization and the principle of maximizing relative entropy; cf. [1], ch. 11, and, e.g., [2].

[2] "Proposition" is the philosophically most common general term for the objects of belief and the one I shall use. The precise nature of these objects is philosophically very problematic, but not my present concern.

[3] Whatever has probability 1 keeps it, according to all rules of belief change within standard probability theory.

[4] I am here alluding to the so-called lottery paradox, which has gained considerable importance in the writings of H.E. Kyburg, jr., I. Levi, and others. Cf., e.g., the various hints in [3].

[5] In the present context W may well be assumed to be finite; so, we need not decide which kinds of fields to consider. In the infinite case, complete fields seem to me to be the most appropriate (cf. [4]), but alternative algebraic frameworks might be adapted, too.

[6] Π denotes the Cartesian product.

[7] A is an atom of $\mathcal{A}$ iff no proper non-empty subset of A ia a member of $\mathcal{A}$.

[8] "Conditional", because these functions can be conditionalized, as we shall see; "natural", because they take natural numbers as values; in [5], Shenoy has proposed the more intuitive label "disbelief function" (which, however, cannot be translated into German). In [4], I have more generally defined "ordinal conditional functions" which take ordinal numbers as values. This generality will not be needed here (all the more as it has some awkward consequences which relate to the fact that addition of ordinal numbers is not commutative).

[9] The latter function for propositions will indeed be the more important one.

[10] $-A$ denotes the complement or the negation of A.

[11] A is contingent iff A and $-A$ both are not empty.

[12] This is to mean that κ takes non-standard natural numbers as values.

[13] For a proof see [4], sect. 6. These are the properties of conditional independence which Pearl calls Contraction and Intersection, e.g. in [6], p.84.

[14] As is manifested by many papers in [7], by [8], and at many other places.

[15] Indeed, I originally invented them for this purpose in [9].

[16] See, e.g., [10], [11], and [6], ch. 3-5. Of course, references could be easily extended.

[17] Perhaps NCFs allow an easier investigation of conditional independence than probability measures, because they are mathematically simpler, because NCFs correspond only to strictly positive probabilities, and because the disturbing property of what Pearl calls weak transitivity (cf. [6], pp.128ff.), which is a special probabilistic law for binary variables, does not hold for NCFs.

[18] This is the so-called Ramsey test, most thoroughly propounded by Gärdenfors, e.g. in [18], who has summarized his work in [19]. See also [20].

[19] As may be easily confirmed with the help of the list in [21].

[20] In order to substantiate this remark, we would have to go more deeply into [22].

[21] Cf., e.g., the pioneering work [23], the overview in [24], and [19], ch. 7 together with ch. 4.

[22] Another serious problem for this strategy is presented by the trivialization result in [18]. I have here avoided this problem by excluding conditional propositions as objects of belief.

[23] In [25], p.205, Shackle mentions that he has considered the law (3) for NCFs instead of (11). But he says almost nothing about why he finally stuck to (11).

[24] I am very grateful to Dan Hunter for having introduced my thoughts and myself to the AI community.

REFERENCES

[1] Jeffrey, R.C., *The Logic of Decision* (University Press, Chicago, 1965, 2nd ed. 1983)

[2] Skyrms, B., "Maximum Entropy Inference as a Special Case of Conditionalization", *Synthese* **63** (1985) 55-74.

[3] Bogdan, R.J. (ed.), *Henry E. Kyburg, jr. & Isaac Levi* (Reidel, Dordrecht, 1982)

[4] Spohn, W., "Ordinal Conditional Functions: A Dynamic Theory of Epistemic States", in: Harper, W.L., Skyrms, B., (eds.), *Causation in Decision, Belief Change, and Statistics* (Kluwer, Dordrecht, 1988) pp.105-134.

[5] Shenoy, P.P., "On Spohn's Rule for Revision of Beliefs", School of Business Working Paper No. 213 (University of Kansas, Lawrence, Kansas, 1989)

[6] Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, San Mateo, Ca., 1988)

[7] Skyrms, B, Harper, W.L., (eds.), *Causation, Chance, and Credence* (Kluwer, Dordrecht, 1988)

[8] Skyrms, B, "Probability and Causation", *Journal of Econometrics* **39** (1988) 53-68.

[9] Spohn, W., *Eine Theorie der Kausalität* (unpublished Habilitationsschrift, Munich, 1984)
[10] Kiiveri, H., Speed, T.P., Carlin, J.B., "Recursive Causal Models", *Journal of the Australian Mathematical Society*, Series A, **36** (1984) 30-52.
[11] Pearl, J., "Fusion, Propagation, and Structuring in Belief Networks", *Artificial Intelligence* **29** (1986) 241-288.
[12] Hunter, D., "Parallel Belief Revision", this volume.
[13] Hunter, D., "Graphoids, Semi-graphoids, and Ordinal Conditional Functions", NRTC Research Paper (Palos Verdes, Ca., 1988)
[14] Geiger, D., Pearl, J., "On the Logic of Causal Models", this volume.
[15] Verma, T., Pearl, J., "Causal Networks: Semantics and Expressiveness", this volume.
[16] Harper, W.L., "Rational Belief Change, Popper Functions and Counterfactuals", in: Harper, W.L., Hooker, C.A., (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, vol. I (Reidel, Dordrecht, 1976) pp. 73-115.
[17] Spohn, W., "The Representation of Popper Measures", *Topoi* **5** (1986) 69-74.
[18] Gärdenfors, P., "Belief Revisions and the Ramsey Test for Conditionals", *Philosophical Review* **95** (1986) 81-93.
[19] Gärdenfors, P., *Knowledge in Flux* (MIT Press, Cambridge, Mass., 1988)
[20] Rott, H., "Conditionals and Theory Change: Revisions, Expansions, and Additions", *Synthese*, in print.
[21] Rescher, N., *Many-valued Logic* (McGraw-Hill, New York, 1969)
[22] Anderson, A.R., Belnap jr., N.D., *Entailment* (University Press, Princeton, 1975)
[23] Lewis, D., *Counterfactuals* (Blackwell, Oxford, 1973)
[24] Nute, D., *Topics in Conditional Logic* (Reidel, Dordrecht, 1980)
[25] Shackle, G.L.S., *Decision, Order, and Time in Human Affairs* (University Press, Cambridge, 1969)
[26] Rescher, N., *Plausible Reasoning* (Van Gorcum, Assen, 1976)
[27] Cohen, L.J., *The Probable and the Provable* (Clarendon Press, Oxford, 1977)
[28] Shafer, G., *A Mathematical Theory of Evidence* (University Press, Princeton, 1976)
[29] Shenoy, P.P., Shafer, G., "An Axiomatic Framework for Bayesian and Belief-function Propagation", this volume.

Uncertainty in Artificial Intelligence 4
R.D. Shachter, T.S. Levitt, L.N. Kanal, J.F. Lemmer (Editors)

EPISTEMOLOGICAL RELEVANCE AND STATISTICAL KNOWLEDGE*

Henry E. KYBURG, Jr.

Computer Science and Philosophy
University of Rochester
Rochester, NY, USA

1. BACKGROUND

For many years, at least since McCarthy and Hayes [10], writers have lamented, and attempted to compensate for, the alleged fact that we often do not have adequate statistical knowledge for governing the uncertainty of belief, for making uncertain inferences, and the like. It is hardly ever spelled out what "adequate statistical knowledge" *would* be, if we had it, and *how* adequate statistical knowledge could be used to control and regulate epistemic uncertainty.

One response to the lack of adequate statistics has been to search for non-statistical measures of uncertainty. The minimal variant has been to propose "subjective probability" as a concept to which we can turn when we lack statistics.

This proposal comes in widely differing flavors, corresponding to the dreadful ambiguity of "subjective". Sometimes this means merely "indexed by a subject". In this sense there is no conflict with statistical representations: the "subjectivity" involved just represents the fact that statistical knowledge is related to (had by) a knower. (This appears to be the sense of "subjective" employed by Cheeseman [3].)

At the other extreme, "subjective" may mean arbitrary, whimsical, subject to no objective control or constraint. Those who think we must turn in this direction are influenced by the feeling that in many cases there may be nothing better to turn to. The philosopher F. P. Ramsey, who did much to make the subjective approach to uncertainty respectable, apparently felt this way; he wrote: "...a man's expectation of drawing a white or a black ball from an urn ... may within the limits of consistency be any he likes..." [12].

Other proposals concern non-probabilistic measures of uncertainty: the certainty factors of Buchan [2], the belief functions of Shafer [14], the fuzzy membership relation of Zadeh [16].

Our purpose here is not to evaluate these alternative treatments of uncertainty, but rather to explore the question of how far you can go on the basis of statistical knowledge that you *do* have, and what considerations must be taken account of in this attempt. Relatively few people have explored the question of how far you can go using statistical knowledge. One writer who has taken this question seriously is Bacchus [1].

*Research underlying the results reported here has been partially supported by the Signals Warfare Center of the United States Army.