Machine Intelligence and Pattern Recognition 5

UNCERTAINTY IN ARTIFICIAL INTELLIGENCE 2

Edited by J.F. Lemmer and L.N. Kanal

North-Holland

UNCERTAINTY IN ARTIFICIAL INTELLIGENCE 2

Machine Intelligence and Pattern Recognition

Volume 5

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 2

Edited by

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

and

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



NORTH-HOLLAND AMSTERDAM · NEW YORK · OXFORD · TOKYO

© ELSEVIER SCIENCE PUBLISHERS B.V., 1988

All rights reserved. No part of this publication may be reproduced stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN: 0 444 70396 9

Publishers: ELSEVIER SCIENCE PUBLISHERS B.V. P.O. Box 1991 1000 BZ Amsterdam The Netherlands

Sole distributors for the U.S.A. and Canada: ELSEVIER SCIENCE PUBLISHING COMPANY, INC. 52 Vanderbilt Avenue New York, N.Y. 10017 U.S.A.

LIBRARY OF CONGRESS Library of Congress Cataloging-in-Publication Data

Uncertainty in artificial intelligence 2 / edited by John F. Lemmer and Laveen N. Kanal. p. cm. -- (Machine intelligence and pattern recognition ; v. 5) Bibliography: p. ISBN 0-444-70396-9 1. Artificial intelligence. 2. Uncertainty (Information theory) I. Lemmer, John F. II. Kanal, Laveen N. III. Title: Uncertainty in artificial intelligence two. IV. Series. Q335.U532 1988 006.3--dc19 88-415 CIP

PRINTED IN THE NETHERLANDS

PREFACE

The favorable reception accorded to our first volume on Uncertainty in Artificial Intelligence (Kanal &Lemmer, 1986; see e.g. review on IEEE Trans. on SMC, March—April 1987) has encouraged us to put together this second volume. The major portion of this second volume consists of expanded versions of papers originally presented at the Workshop on Uncertainty in Artificial Intelligence which was held August 8—10, 1986 at the University of Pennsylvania in Philadelphia. In addition, this volume contains a number of new papers.

At the time of final preparation the authors knew which other papers would appear, so there is considerable cross referencing among the papers. Because there is by no means a consensus of opinions within this volume, the cross referencing is particularly interesting.

We have placed the papers in four major sections: analysis, tools, theory, and applications. The analysis section contains papers which substantiate, dispute, or compare the attributes of various approaches to uncertainty. The papers in this first section raise many issues important for critical reading of the remaining sections. The tools section contains papers which provide sufficient information for one to implement uncertainty calculations, or aspects of uncertainty, without requiring deep understanding of the underlying theory. Papers in the theory section explicate various approaches to uncertainty without necessarily addressing how the approaches can be implemented. Finally, the applications section describes the difficulties involved in, and results produced by, incorporating uncertainty into implemented systems.

A number of interesting ideas emerge from the papers in this volume, and some are briefly mentioned here. Heckerman and Horvitz describe some newly discovered work showing that first order predicate calculus combined with some simple (and innocuous?) requirements for continuity imply the axioms of probability. They argue that this frees probability theory from its ties to repeatable events. Hummel and Landy show how Dempster—Shafer theory can be fully interpreted in the context of traditional probability theory. Three papers attempt to forge a link between probabilistic and non-numeric uncertainty approaches and thereby develop a basis for providing explanations for inferred probabilities: Norton attempts to directly explain inference results in terms of correlations; Su-Shing Chen attempts to bring probabilistic logic into the framework of probability theory; Hawthorne describes a theory that unifies into a single coherent system a symbolic approach to non-monotonic entailment and a quantitative (Bayesian) approach. Dalkey disproves by counter-example, an often cited property of distributions formed by product extension.

It is clear from the papers that there continues to be no consensus concerning the "best" approach to uncertainty for AI systems. Moreover, there is no agreement on how to measure "best". Further it is likely that best will always depend on the particular use to be made of uncertainty. Because in many experiments all approaches seem to imply essentially the same actions, the idea that the choice of any particular approach is not important was discussed at some length during the workshop without reaching a consensus.

We would stress that any choice of approach to uncertainty must be based on factors in addition to the merits of the uncertainty representation and calculus. These additional factors include the difficulty of acquiring the original uncertainty estimates (knowledge engineering), the computational complexity of inference, the semantics which guide the acquisition of the original estimates and guide the interpretation of computed results, and how the chosen representation is used in decision making.

We hope, that like volume 1, this volume will also stimulate further communication among researchers interested in handling uncertainty in AI systems. We thank all our colleagues who helped in organizing the workshop, and all the authors, reviewers, and the folks at North-Holland for their excellent collaboration in helping to produce the earlier volume and this volume on Uncertainty in Artificial Intelligence.

Laveen N. Kanal College Park, MD John F. Lemmer Rome, NY

CONTRIBUTORS

- Shoshana Abel Mountain View, CA 94043
- Lee A. Appelbaum Advanced Decision Systems, Mountain View, CA 94040
- Jonathan Baron University of Pennsylvania, Philadelphia, PA 19104 and Medical College of Pennsylvania, Philadelphia, PA 19129
- Steven W. Barth PAR Government Systems Corporation, New Hartford, NY 13413
- Lashon B. Booker AI Center, Naval Research Laboratory, Washington, DC 20375
- Roger A. Browse Queens University at Kingston, Ontario, Canada K7L 3N7
- Peter Cheeseman NASA Ames Research Center, Moffett Field, CA 94035
- Kaihu Chen University of Illinois at Urbana-Champaign, 1L 61801 Su-Shing Chen University of North Carolina, NC 28223
- John R. Clarke University of Pennsylvania, Philadelphia, PA 19104 and Medical College of Pennsylvania, Philadelphia, PA 19129
- A. Julian Craddock Queens University at Kingston, Ontario, Canada K7L 3N7
- N.C. Dalkey University of California, Los Angeles, CA 90024
- Hugh F. Durrant-Whyte University of Pennsylvania, Philadelphia, PA 19104
- Brian Falkenhainer University of Illinois, Urbana, IL 61801
- B.R. Fox McDonnell Douglas Research Laboratories, St. Louis, MO 63166
- Sally A. Goldman Massachusetts Institute of Technology, Cambridge, MA 02139
- Benjamin N. Grosof Stanford University, Stanford, CA 94305
- Greg Hager University of Pennsylvania, Philadelphia, PA 19104
- James Hawthorne Honeywell Systems and Research Center, Minneapolis, MN 55418

David Heckerman Stanford University, Stanford, CA 94305

- Steven J. Henkind Courant Institute of Mathematical Sciences, NYU,
- NY 10012 and Mount Sinai School of Medicine, New York, NY 10029 Max Henrion Carnegie-Mellon University, Pittsburgh, PA 15213
- Peter D. Holden McDonnell Douglas Corporation, St. Louis, MO 63166 Naveen Hota JAYCOR, Vienna, VA 22180
- Eric Horvitz Stanford University, Stanford, CA 94305
- Robert Hummel Courant Institute of Mathematical Sciences, NYU, NY 10012
- Karl G. Kempf FMC Corporation, Santa Clara, CA 95052
- Henry E. Kyburg University of Rochester, Rochester, NY 14627
- Michael Landy Courant Institute of Mathematical Sciences, NYU, NY 10012
- Todd S. Levitt Advanced Decisions Systems. Mountain View, CA 94040
- Ze-Nian Li University of Wisconsin, Milwaukee, WI 53211
- Ronald P. Loui University of Rochester, Rochester, NY 14627
- Mary McLeish University of Guelph, Guelph, Ontario, Canada N1G 2W1
- Khaled Mellouli University of Kansas, Lawrence, KS 66045
- Steven W. Norton PAR Government Systems Corporation, New Hartford, NY 13413
- Judea Pearl University of California, Los Angeles, CA 90024
- Bruce M. Perrin McDonnell Douglas Corporation, St. Louis, MO 63166
- Ronald L. Rivest Massachusetts Institute of Technology, Cambridge, MA 02139
- Ross D. Shachter Stanford University, Stanford, CA 94305
- Glenn Shafer University of Kansas, Lawrence, KS 66045
- Prakash P. Shenoy University of Kansas, Lawrence, KS 66045
- Shimon Shocken New York University, NY 10003
- Stanley M. Schwartz University of Pennsylvania, Philadelphia, PA 19104 and Medical College of Pennsylvania, Philadelphia, PA 19129
- Matthew Self University of California, Berkeley, CA 94720
- Randall Smith General Motors Research Laboratories, Warren, MI 48090
- Richard M. Tong Advanced Decision Systems, Mountain View, CA 94040
- Leonard Uhr University of Wisconsin, Madison, WI 53706
- Silvio Ursic Ursic Computing, Madison, WI 53714
- David S. Vaughan McDonnell Douglas Corporation, St. Louis, MO 63166
- Michael P. Wellman MIT Laboratory for Computer Science, Cambridge, MA 02139
- Ben P. Wise Thayer School, Dartmouth, NH 03755
- Robert M. Yadrick McDonnell Douglas Corporation, St. Louis, MO 63166
- Ronald R. Yager Iona College, New Rochelle, NY 10801

CONTENTS

Pre	face	v
Co	ntributors	vii
I.	ANALYSIS	
	Models vs. Inductive Inference for Dealing with Probabilistic Knowledge	
	N.C. Dalkey	3
	An Axiomatic Framework for Belief Updates D.E. Heckerman	11
	The Myth of Modularity in Rule-Based Systems for Reasoning with Uncertainty	
	D.E. Heckerman and E.J. Horvitz	23
	Imprecise Meanings as a Cause of Uncertainty in Medical Knowledge-Based Systems	
	S.J. Henkind	35
	Evidence as Opinions of Experts R. Hummel and M. Landy	43
	Probabilistic Logic: Some Comments and Possible Use for Nonmonotonic Reasoning	
	M. McLeish	55
	Experiments with Interval-Valued Uncertainty R.M. Tong and L.A. Appelbaum	63

Con	tents	S
-----	-------	---

Evaluation of Uncertain Inference Models I: PROSPECTOR R.M. Yadrick, B.M. Perrin, D.S. Vaughan, P.D. Holden, and K.G. Kempf	77
Experimentally Comparing Uncertain Inference Systems to Probability B.P. Wise	89
TOOLS	
Knowledge Engineering within a Generalized Bayesian	
S.W. Barth and S.W. Norton	103
Learning to Predict: An Inductive Approach K. Chen	115
Towards a General-Purpose Belief Maintenance System B. Falkenhainer	125
A Non-Iterative Maximum Entropy Algorithm S.A. Goldman and R.L. Rivest	133
Propagating Uncertainty in Bayesian Networks by Probabilistic Logic Sampling M. Henrion	149
An Explanation Mechanism for Bayesian Inferencing Systems	
S.W. Norton	165
On the Rational Scope of Probabilistic Rule-Based Inference Systems S. Schocken	175
DAVID: Influence Diagram Processing System for the Macintosh R.D. Shachter	191
Qualitative Probabilistic Networks for Planning under Uncertainty	
M.P. Wellman	197

II.

	Contents	
	On Implementing Usual Values R.R. Yager	209
III.	THEORY	
	Some Extensions of Probabilistic Logic SS. Chen	221
	Belief as Summarization and Meta-Support A.J. Craddock and R.A. Browse	229
	Non-Monotonicity in Probabilistic Reasoning B.N. Grosof	237
	A Semantic Approach to Non-Monotonic Entailments J. Hawthorne	251
	Knowledge H.E. Kyburg, Jr.	263
	Computing Reference Classes R.P. Loui	273
	Distributed Revision of Belief Commitment in Composite Explanations J. Pearl	291
	A Backwards View for Asssessment R.D. Shachter and D. Heckerman	317
	Propagation of Belief Functions: A Distributed Approach P.P. Shenoy, G. Shafer, and K. Mellouli	325
	Generalizing Fuzzy Logic Probabilistic Inferences S. Ursic	337
IV.	APPLICATIONS	
	The Sum-and-Lattice-Points Method Based on an Evidential-Reasoning System Applied to the Real-Time Vehicle Guidance Problem	
	S. Abel	365

Probabilistic Reasoning about Ship Images L.B. Booker and N. Hota	371
Information and Multi-Sensor Coordination G. Hager and H.F. Durrant-Whyte	381
Planning, Scheduling, and Uncertainty in the Sequence of Future Events B.R. Fox and K.G. Kempf	395
Evidential Reasoning in a Computer Vision System ZN. Li and L. Uhr	403
Bayesian Inference for Radar Imagery Based Surveillance T.S. Levitt	413
A Causal Bayesian Model for the Diagnosis of Appendicitis S.M. Schwartz, J. Baron, and J.R. Clarke	423
Estimating Uncertain Spatial Relationships in Robotics R. Smith, M. Self, and P. Cheeseman	435

Contents

Ι

ANALYSIS

This page intentionally left blank

MODELS VS. INDUCTIVE INFERENCE FOR DEALING WITH PROBABILISTIC KNOWLEDGE

N. C. DALKEY*

1. Introduction

Two different approaches to dealing with probabilistic knowledge are examined -models and inductive inference. Examples of the first are: influence diagrams [1], Bayesian networks [2], log-linear models [3, 4]. Examples of the second are: games-against nature [5, 6], varieties of maximum-entropy methods [7, 8, 9], and the author's min-score induction [10].

In the modeling approach, the basic issue is manageability, with respect to data elicitation and computation. Thus, it is assumed that the pertinent set of users in some sense knows the relevant probabilities, and the problem is to format that knowledge in a way that is convenient to input and store and that allows computation of the answers to current questions in an expeditious fashion.

The basic issue for the inductive approach appears at first sight to be very different. In this approach it is presumed that the relevant probabilities are only partially known, and the problem is to extend that incomplete information in a reasonable way to answer current questions. Clearly, this approach requires that some form of induction be invoked. Of course, manageability is an important additional concern.

Despite their seeming differences, the two approaches have a fair amount in common, especially with respect to the structural framework they employ. Roughly speaking, this framework involves identifying clusters of variables which strongly interact, establishing marginal probability distributions on the clusters, and extending the subdistributions to a more complete distribution, usually via a product formalism. The product extension is justified on the modeling approach in terms of assumed conditional independence; in the inductive approach the product form arises from an inductive rule.

2. Structures on Event Spaces

An event space is a set $X = X_1, \dots, X_n$ of descriptors which is presumed to cover the subject matter of interest. For example, in a medical context, the X_i 's could be disease states, symptoms, test results, outcomes of treatment, and the like. Each descriptor X_i involves a set of states x_{ij} , which is a partition (exclusive and exhaustive division) of the "universe" of potential cases. The vector $x = (x_1, \dots, x_n)$ is a joint state for a specific case. It is presumed that there is a joint probability distribution P(X) on the set of joint states, so that

^{*} Department of Computer Science, University of California, Los Angeles, CA. 90024. This work was supported in part by National Science Foundation Grant IST 84-05161.

$$\sum_{X} P(x) = 1$$
$$\sum_{j} P(x_{ij}) = 1$$

We define two types of components. An absolute component Y is a subset of descriptors. A conditional component (Z | W) is an ordered pair of absolute components. A probability P(Y) on an absolute component is a joint distribution on the states y for descriptors in Y; P(Y) is a subdistribution (or marginal) of P(X). Let -Y denote the complement of Y (all descriptors in X not in Y). Thus

$$P(Y) = \sum_{-Y} P(X) \tag{1}$$

A probability P(Z | W) is a conditional probability distribution on the states z given the states w. Thus

$$P(Z | W) = \frac{\sum_{-Z.W} P(X)}{\sum_{-W} P(X)}$$
(1a)

(The period in -Z.W denotes the logical product "and".)

A set of components $C = Y_1, \dots, Y_k$ is called a structure. The corresponding probability distributions on members of $C, PC = P(Y_1), \dots, P(Y_k)$, is called a probability system (or system for short.) In this notation, Y may be either absolute or conditional.

A system PC is called consistent if there is a probability distribution P(X) that fulfills (1) or (1a) for all components Y in PC. In general, if PC is consistent, there will be a set K(PC) of distributions compatible with PC.

In the model approach, it is assumed that a system *PC* represents the clustering of descriptors with respect to dependence; i.e., within a component *Y*, the descriptors have "strong" probabilistic interactions, whereas if X_i and X_j do not occur together in any component, then they are conditionally independent. Specifically,

$$P(X_{i}, X_{i} | -X_{i}, X_{i}) = P(X_{i} | -X_{i}, X_{i}) P(X_{i} | -X_{i}, X_{i})$$
⁽²⁾

In the inductive approach, the system PC represents what is known concerning P(X). If X_i and X_j do not occur in a common component, then nothing is known about their probabilistic relationship.

A structure C by itself exhibits many of the general properties of the available knowledge. Thus, it is possible to determine that one structure is uniformly more informative than another, or to specify which structures have a product extension, without reference to the probabilities PC [11]. As the developers of influence diagrams and Bayesian networks have noted, this feature allows a significant amount of preliminary analysis to be carried out before numbers are introduced.

3. Webs

Structures have an internal organization; e.g., components may overlap. There are a number of different ways to represent this organization. A common representation is as a graph, where the components are the nodes and one component is connected by an arc to another if they overlap. A more convenient representation for the purposes of this paper is what I call a web. Let [C] designate the set of all descriptors which belong to some member of C. An absolute component Y is terminal if Y consists of two subcomponents Ζ and $W, Z[C - (Y)] = 0, W \subset [C - (Y)]$. Thus, W "connects" Z to the remainder of C. (W may be vacuous, in which case Z is unconnected to the remainder of C.) A conditional component Y = (Z | W) is terminal if the preceeding conditions hold for Z and W. (W cannot be vacuous for a conditional component.)

A web is a structure which fulfills the recursion:

- 1. Any absolute component is a web.
- 2. If Y is a terminal component, and $C \{Y\}$ is a web, then C is a web.

From the definition, a web contains at least one absolute component. A web can be "unpacked" to generate a linear order on the components by starting with any terminal component Y, labeling it 1, choosing any terminal component in $C - \{Y\}$, labeling it 2, and so on. A web is called conditional if all absolute components are distinct, i.e., do not overlap.

A web is somewhat more general than influence diagrams or Bayesian networks. These can be characterized as conditional webs where for any conditional component Y = (Z | W), Z consists of a single descriptor.

Conditional webs are significant for modeling probabilistic knowledge as a result of two basic properties:

- a. The product $P^{\bullet}(X) = \prod_{i=1}^{n} |P(Y)|$ is a joint probability distribution on X.
- b. $P^{\bullet}(X)$ is an extension of PC; i.e., it fulfills (1) or (1a) for all Y in C.

Proofs for these assertions are readily constructed by induction on the number of components in a web.

What these two properties entail, in effect, is that if you can represent your knowledge concerning a distribution P(X) by the sub-distributions PC for a web C plus assuming conditional independence for descriptors not in common components, then the product $P^{\bullet}(X)$ "automatically" expresses that knowledge.

From the modeling point of view, then, a web is a relatively manageable representation of probabilistic knowledge. All that need be input are the subdistributions PC. The product is quite convenient for computations; e.g., the manipulations feasible for influence diagrams are directly extendable to webs.

4. Induction and maximum-entropy

Turning to the inductive approach, in an earlier publication I demonstrated that for a subspecies of web, namely a forest, the product extension is the maximum entropy extension of *PC* [11]. A forest is a web in which all terminal components Y = (Z:W) fulfill the additional restriction that *W* is contained some component Y' in $C - \{Y\}$. (In a general web, *W* need only be contained in the set of all descriptors "covered" by $C - \{Y\}$.) In the graphical representation mentioned earlier where arcs are defined by overlap of components, a forest is a graph with no loops. A forest corresponds to Goodman's decomposable model [3].

Maximum entropy is an instance of min-score inference which has the dual properties: (a) guaranteed expectation -- in the case of maximum entropy, the conclusion is always at least as informative as it claims to be -- and (b) positive value of information -- a conclusion based on additional knowledge will be at least as informative as a conclusion without that knowledge [10]. Thus, if all you know is a set of subdistributions PC, and PC is a forest, then the product extension is a supportable estimate of the total distribution.

One of the motivations for studying webs was the expectation that the product extension would also turn out to be the maximum entropy extension for a general web. The expectation was based on a purported result of P.M. Lewis frequently cited in the literature to the effect that for a structure with a product extension, the product is the maximum entropy extension [12]. Unfortunately, the Lewis "result" happens to be incorrect.

An elementary counter-example is furnished by the simplest of all possible webs that is not a forest, namely the structure $C = \{X_1, X_2, (X_3 | X_1, X_2)\}$. Set $P(X_1) = P(X_2) = .5$ and define $P(X_3 | X_1, X_2)$ by Table I, where "1" means occurrence in the list of cases.

Table I

X_1	X_2	$P(X_3 X_1, X_2)$
1	1	1
1	0	.5
0	1	.5
0	0	0

The product distribution $P^{\bullet}(X)$ is displayed in Table II, along with another distribution $P^{\bullet}(X)$. $P^{\circ}(X)$ is an extension of PC -- which can easily be verified by summation -- and is also clearly a higher entropy distribution.

The entropy of $P^{\bullet}(X) = -\sum_{X} P(x) \log P(x) = 1.7329$, whereas the entropy of $P^{\circ}(X) = 1.7918$. The numerical difference in entropy is small, but the difference between .25 and 1/6 for P(1, 1, 1), e.g., may not seem trivial.

The elementary structure C of the example is actually a substructure of any web that is not a forest. Hence a similar counter-example can be constructed for any such web. The example is also a counter to the Lewis "result".

Table II

<i>X</i> ₁	X 2	X_3	$P^{\bullet}(X)$	$P^{\circ}(X)$
1	1	1	.25	1/6
1	1	0	0	0
1	0	1	.125	1/6
1	0	0	.125	1/6
0	1	1	.125	1/6
0	1	0	.125	1/6
0	0	1	0	0
0	0	0	.25	1/6

The upshot of this inquiry, then, is that a forest is the most general structure for which the product extension is always the maximum entropy extension.

5. Discussion

At first glance, the fact that the product extension of a web is not in general maximum entropy may appear benign. From the standpoint of the model approach, the basic properties of a web -- the product is a probability and an extension of PC -- make webs a highly convenient representation of probabilistic knowledge. All that is lost is a desirable, but by no means essential, fallback. In the case of a forest, for example, if the assumption of conditional independence for separated descriptors is shaky, then it can still be contended that the product is a reasonable estimate of the joint distribution, given PC. It would be a valuable safety feature if the same could be claimed for a web.

From the standpoint of the inductive approach, it is perhaps unfortunate that the product extension of a web is not maximum entropy. However, the maximum entropy extension can be sought by other means [13]. What is lost is the convenience of the product form. For the complex systems of many descriptors common in expert systems, maximum entropy formalisms are likely to be cumbersome.

On a somewhat deeper level, however, the result is thought-provoking. Independence is a common "simplifying" assumption in expert systems [14]. The maximum entropy property, where germaine, is a good justification of the "assumption" even when there is no evidence either for or against independence. However, as the example shows, maximum entropy does not imply independence, not even conditional independence, if the structure is not a forest. In the example, $P^{\circ}(x_1|x_3) = P^{\circ}(x_2|x_3) = \frac{1}{3}$; but $P^{\circ}(x_1.x_2|x_3) = \frac{1}{3}$, rather than 4/9 as required by conditional independence.

One route that can be taken is to "prune" the structure to a forest. Lemmer [15] has adopted this suggestion, following a program proposed by Lewis [12], Chow and Liu [16], and others. The advantages of this approach are clear: substantive inputs can be restricted to the subdistributions in PC for the forest, the product extension is automatically consistent with the inputs, and, as I mentioned above, the fact that the product extension is maximum entropy carries strong guarantees.

A basic element missing from this program is a measure of the information that is lost by the pruning process. Concommitantly, there is no systematic procedure for determining the most informative forest contained in the knowledge available to the analyst. Given a general probability system PC, if PC is consistent, the amount of information in PC can be defined as

max Entropy (P). At present, there is no way to determine this quantity directly from $PC - P \in \mathcal{K}(PC)$ or, for that matter, determining whether PC is consistent. These issues appear to be one area of potentially fruitful research.

References

- 1. Howard, R. A. and J. E. Matheson, "Influence Diagrams," in R. A. Howard and J. E. Matheson, *The Principles and Applications of Decision Analysis*, Vol. II, Strategic Decisions Group, Menlo Park, 1981.
- Pearl, J., "Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning," UCLA Computer Science Department *Technical Report* 850021 (R-43); in *Proceedings, Cognitive Science Society*, UC Irvine, August 15-17, 1985, pp. 329-334.
- Goodman, L. A., "The Multivariate Analysis of Qualitative Data: Interaction among Multiple Classifications," J. Amer. Statist. Assoc., 65, 1970, pp. 226-256.
- 4. Lauritzen, S. L., *Lectures on Contingency Tables*, Aalberg University Press, 1982.
- 5. Wald, A., Statistical Decision Functions, John Wiley and Sons, New York, 1950.
- 6. Blackwell, D. and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954.
- 7. Kullback, S., Information Theory and Statistics, Wiley, New York, 1959.
- 8. Jaynes, E. T., "Prior Probabilities," *IEEE Transactions on Systems Science and Cybernetics*, SSC-4, 1968, pp. 227-241.
- 9. Shore, J. E. and R. W. Johnson, "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," *IEEE Transactions on Information Theory*, Vol. IT-26, January, 1980, pp. 26-37.
- 10. Dalkey, N. C., "Inductive Inference and the Maximum Entropy Principle," in C. Ray Smith and W. T. Grandy Jr. (eds.) Maximum-Entropy and Bayesian Methods in Inverse Problems, D. Reidel, 1985.
- Dalkey, N. C., "Min-Score Inference on Probability Systems," UCLA-ENG-CSL-8112, June, 1981.
- 12. Lewis, P. M., "Approximating Probability Distributions to Reduce Storage Requirements," *Information and Control*, 2, 1959, pp. 214-225.
- 13. Cheeseman, P., "A Method of Computing Generalized Bayesian Probability Values for Expert Systems," *Proceedings of the 1983 International Joint Conference on Artificial Intelligence.*

- Duda, R., P. Hart, K. Konolige, and R. Reboh, "A Computer Based Consultant for Mineral Exploration," SRI Project 6415 Final Report, SRI International, Sept., 1979.
- 15. Lemmer, J., "Generalized Bayesian Updating of Incompletely Specified Distributions," in *Large Scale Systems*, 5, Elsevier, North-Holland, 1983, pp. 51-68.
- 16. Chow, C. K. and C. N. Liu, "Approximating Discrete Probabilities with Dependence Trees," *IEEE Transactions on Information Theory*, May 1968, pp. 462-467.

This page intentionally left blank

AN AXIOMATIC FRAMEWORK FOR BELIEF UPDATES

David E. Heckerman

Medical Computer Science Group, Knowledge Systems Laboratory Medical School Office Building, Room 215, Stanford, California 94305

In the 1940's, a physicist named Cox provided the first formal justification for the axioms of probability based on the subjective or Bayesian interpretation. He showed that if a measure of belief satisfies several fundamental properties, then the measure must be some monotonic transformation of a probability. In this paper, measures of *change* in belief or *belief updates* are examined. In the spirit of Cox, properties for a measure of change in belief are enumerated. It is shown that if a measure satisfies these properties, it must satisfy other restrictive conditions. For example, it is shown that belief updates in a probabilistic context must be equal to some monotonic transformation of a likelihood ratio. It is hoped that this formal explication of the belief update paradigm will facilitate critical discussion and useful extensions of the approach.

1. INTRODUCTION

As researchers in artificial intelligence have begun to tackle real-world domains such as medical diagnosis, mineral exploration, and financial planning, there has been increasing interest in the development and refinement of methods for reasoning with uncertainty. Much of the work in this area has been focused on methods for the representation and manipulation of measures of *absolute* belief, quantities which reflect the absolute degree to which propositions are believed. There has also been much interest in methodologies which focus on measures of *change* in belief or *belief updates*¹, quantities which reflect the degree to which beliefs in propositions change when evidence about them becomes known. Such methodologies include the MYCIN certainty factor model [1], the PROSPECTOR scoring scheme [2], and the application of Dempster's Rule to the combination of "weights of evidence" [3].

In this paper, a formal explication of the belief update paradigm is given. The presentation is modeled after the work of a physicist named R.T. Cox. In 1946, Cox [4] enumerated a small set of intuitive properties for a measure of *absolute* belief and proved that *any* measure that satisfies these properties must be some monotonic transformation of a probability. In the same spirit, a set of properties or axioms that are intended to capture the notion of a belief update are enumerated. It is then shown that these properties place strong restrictions on measures of change in belief. For example, it is shown that the only measures which satisfy the properties in a probabilistic context are monotonic transformations of the likelihood ratio $\lambda(H,E,e) = p(E|H,e)/p(E|-H,e)$, where H is a hypothesis, E is a piece of evidence relevant to the hypothesis, and e is background information.

It should be emphasized that the goal of this axiomization is not to prove that belief updates can only take the form described above. Rather, it is hoped that a formal explication of the update paradigm will stimulate constructive research in this area. For example, the axioms presented here can serve as a tool for the identification and communication of dissatisfaction with the update approach. Given the properties for a belief update, a researcher may be able