not only when these are observational. We agree that the graphs suggested by Glymour and Spirtes could possibly be chosen as another description of our nondecomposable models but we do not regard them as indicating useful potential processes to generate the data, the point of our distinction.

In a recent paper, Stone (1993) elucidates requirements for particular causal interpretations. He also examines critically strongly ignorable treatment allocation. Pearl in his contribution gives an important graphical interpretation exactly of this assumption, this facilitating the judgement of the effects of interventions in a hypothesized causal process.

Several contributors mention the role of latent variables, including as a special case the occurrence of measuring errors. We agree that their use, preferably sparingly, especially in elucidating nondecomposable models, needs further study. For instance, the tetrad conditions studied by Spirtes, Glymour and Scheines (1993) for linear relations become relevant as well for binary variables having a quadratic exponential distribution. This distribution has some of the properties of the multivariate normal distribution and provides exact or approximate answers to Hill's question about graphical theory for binary distributions and to Whittaker's comments on complete independence.

Dempster favours shrinking estimates toward zero as opposed to setting parameters exactly to zero. We agree when empirical prediction is the objective, but not where essentially qualitative understanding via simple representations is involved, and the latter is our main concern.

The issue, raised by Whittaker, of labelling the edges of a graph can be solved in various ways if a single degree of freedom is attached to each edge (by partial correlation coefficients or by standardized regression coefficients, for instance). The introduction of graphs with dashed edges has, however, a different objective, because it leads to structures of independence different from those discussed by Whittaker, thus enriching the class of graphical chain models, as pointed out by Hill. Whittaker's graphs (ai) and (aii) do not represent the multivariate regression of our Figure 1c because the essential association between the two responses is omitted.

Whittaker points out the relation of the Schur complement to partial correlations and inverse covariance matrices. An early treatment of this in the statistical literature is by Cramér (1946, subsections 22.7, 23.4 and 23.5). The connection between partial correlation and canonical parameters in the exponential family has opened the road to defining analogous independence structures for discrete variables and for mixed discrete and continuous variables, known now as block regression (full edge) chain models.

In general distributional assumptions are necessary, in addition to the independence graph, for a full specification of a statistical model. Indeed some research hypotheses may not be possible for a particular joint distribution of specified form. For example, $X \perp\!\!\!\perp Y|A$ cannot hold without additional independences if the joint distribution is given by the linear logistic regression of the binary variable A on the bivariate normal variable (X, Y). Similarly if (X, Y) are conditionally bivariate normal given the discrete variable A, then marginal independence of X and Y is possible only with additional independences. See Cox and Wermuth (1992b) for further details.

We were glad to see that Sobel regards our introduction of multivariate regression (dashed edge) chain graphs as a step toward more traditional analyses in the social sciences. In fact, it was one of our purposes to provide simple examples which help one to recognize similarities and distinctions between different approaches, the latter being explicitly appreciated by both Sobel and Dempster.

Because of the particular focus of our paper, we have put little emphasis on such issues as description of sample selection, checking data quality, testing model adequacy, examining the need of data transformation and comparison of the fits of different kinds of models. All of these are a normal if often difficult part of applied statistical work. From our present perspective, whether the formal aspects to the analysis are in frequentist or Bayesian terms is a secondary issue.

A special topic for further work concerns the role of graphs with both kinds of edge, for example, in representing the regression for multivariate binary data studied by Zhao and Prentice (1990) and by Fitzmaurice and Laird (1993).

Rejoinder

David J. Spiegelhalter, A. Philip Dawid, Steffen L. Lauritzen and Robert G. Cowell

We are grateful to the discussants for their thoughtful comments: since our paper is already quite long enough we shall try to restrict our responses. We shall first deal with representations of causality, followed by some technical points on zero probabilities. Automatic model construction will then be considered, and whether a

network or tree representation is appropriate in the example in our paper and in general. Finally, we shall discuss the practical use of DAGs.

Before starting on particular themes we should reiterate the essential benefits of graphical models. Many of the discussants describe the general value of a representation of structural independence assumptions, and its power for communicating important ideas with subject matter specialists. Normand and Dempster point out that exact inference is only possible in restricted circumstances, but as graphs become complex either through size or distributional forms we may adopt simulation techniques to derive the appropriate posterior or predictive distributions. Fortunately the recent development of Markov Chain Monte Carlo (MCMC) methods [see, e.g., Gelfand and Smith (1990)] means that this is not a great restriction, and hence attention can properly shift away from purely computational issues towards the more important themes of what the model represents, and how it may be constructed and criticised. Incidentally, in answer to a question from Hill, DAGs with both discrete and continuous variables have been considered by Lauritzen (1992).

1. CAUSALITY AND DIRECTED ACYCLIC GRAPHS

Pearl, Glymour and Spirtes all emphasize the potential in DAG models for representing causality and causal thinking.

We have no fundamental disagreement with the idea that DAGs are objects suitable for description of causal structure and that causal structure is of fundamental importance, in particular when discussing interventions. There has been a fascinating development in recent years and a number of authors, including the discussants, have seriously taken up the fundamental challenge of tackling causality issues in a systematic and illuminating way, where just a few years ago this seems to have been considered too difficult a subject.

Our approach has been relatively modest, emphasizing that even if a clear causal understanding is not present, DAG models or the more general chain graph models may be helpful. Of course, if a causal model can be established with reasonable reliability, this would be more useful than the more descriptive models we have discussed in our paper.

It does, however, seem appropriate to point out that phenomena in the world surrounding us may sometimes be described more clearly and directly by association rather than causation, for example in terms of constraints that must be satisfied in equilibrium states of various systems. There may be an underlying causal explanation, but it is conveniently used to derive the association constraints rather than being explicitly represented in the model.

Below we have some more specific comments to issues raised by the discussants concerning causal interpretation of DAG models.

1.1 Specific Issues: Pearl

We find the discussion of directed acyclic graphs with intervention nodes most illuminating and clear. It may be worthwhile pointing out that in this context conditional independence formally needs an asymmetric interpretation; as described in Dawid (1979a), there is an inherent asymmetry between interventions and random variables since the former do not have probabilities attached to them. However, all arguments given in the contribution seem to remain correct under this interpretation.

We believe the condition for equivalence of external intervention to passive observation in (7) is sufficient, but not necessary as it is stated. For example, if we have a network with three variables I, J and K and we study the effect of the intervention $set(J = j)$, we have

$$p_j(k) = \sum_i p(k|j,i)p(i)$$

whereas

$$p(k|j) = \frac{\sum_i p(k|j,i)p(j|i)p(i)}{\sum_i p(j|i)p(i)}.$$

However, the equality of these two expressions does not imply that $K \perp\!\!\!\perp I \mid J$ as (7) would suggest.

Finally some readers may find it simpler to use separation in moral graphs of ancestral sets (Lauritzen et al., 1990) to investigate the conditional independence statements in (9).

1.2 Specific Issues: Glymour and Spirtes

The authors seem to have misinterpreted our comments concerning reciprocal causation and chain graphs, made in connection with the specific link between *Birth Asphyxia?* and *Disease?*. This was not meant as a general suggestion for representing mixtures of DAGs by undirected links and chain graphs. Rather we wanted to point out that the direction on this particular link did not have a causal interpretation and an undirected link would have been a way of representing this.

Generally we find it questionable to represent feedback or reciprocal causation without having a small time delay between effect and feedback. In an idealised world, the time delay becomes infinitesimally small compared to the general delay in "causal time," so that distributions can be assumed to reach an equilibrium. Causal chain graph models represent such equilibrium systems. This is seen most easily in the case where all links are undirected: modern MCMC methods precisely exploit this fact to calculate probabilities in such models by simulating dynamical "DAG"-systems with the correct equilibrium distribution.

Certainly, as also demonstrated clearly by the companion paper of Cox and Wermuth, there is an abun-

dance of other types of independence structures where chain graphs are not helpful and other graphical representations of the structure are illuminating.

2. CONDITIONAL INDEPENDENCE

Hill points out the virtues of the various graphical representations of conditional independence, but contrasts their seductive simplicity with the complexity of detail in interpretation of Markov properties when the probabilities involved are not everywhere positive.

For application in expert systems it is mandatory that cases of zero probabilities are covered by the theory, since the aim is to extend classical logic to cope with uncertainty. We find the statement that "no results exist for distributions with zero probabilities" to be somewhat inaccurate. For DAGs, positivity plays no role at all for the interpretation of the Markov property (Pearl, 1988; Dawid, 1979a; Lauritzen et al., 1990) and if Hammersley and Clifford had considered directed graphs instead of the more complicated–but seemingly easier–case of undirected graphs, much might have looked different today.

But even in the undirected case there is some clarity for nonpositive distributions, partly due to Moussouris (1974). Define the following types of Markov properties for a finite and undirected graph $\mathcal{G}$:

Factorization: $P \in M_F(\mathcal{G})$ if $P(x) = \Pi_a \Psi_a(x_a)$, the product ranging over all complete sets.
Limit Markov: $P \in M_E(\mathcal{G})$ if there exists a sequence $P_n \in M_F(\mathcal{G})$ such that $P_n \to P$ for $n \to \infty$;
Global Markov: $P \in M_G(\mathcal{G})$ if $A \perp\!\!\!\perp B \mid S[P]$ whenever S separates A from B in $\mathcal{G}$;
Local Markov: $P \in M_L(\mathcal{G})$ if $\alpha \perp\!\!\!\perp V \setminus \mathrm{cl}(\alpha) \mid \mathrm{bd}(\alpha)[P]$;
Pairwise Markov: $P \in M_P(\mathcal{G})$ if $\alpha \perp\!\!\!\perp \beta \mid V \setminus \{\alpha, \beta\}[P]$ whenever $\alpha \not\sim \beta$.

If we let $M_+(\mathcal{G})$ denote the strictly positive distributions that obey the pairwise Markov property, it is not difficult to show that

$$M_+(\mathcal{G}) \subset M_F(\mathcal{G}) \subset M_E(\mathcal{G}) \subset M_G(\mathcal{G}) \subset M_L(\mathcal{G}) \subset M_P(\mathcal{G}).$$

All inclusions are strict in general but may turn into equalities for special types of graphs. Hill also mentions (Theorem 1*) that $M_F(\mathcal{G}) = M_G(\mathcal{G})$ for decomposable graphs $\mathcal{G}$ (Dawid and Lauritzen, 1993) and not otherwise. Analogous results hold for other of the properties (Matúš, 1992b).

Some points are unclear to us concerning split graphs. Hill probably assumes that the triplets of conditional independence necessarily involve all variables, else Theorem 3+ must be incorrect as stated. The split graph of the conditional independences $\{A \perp\!\!\!\perp B \mid (C, D), C \perp\!\!\!\perp D\}$ is surely the chordless 4-cycle, but the independence $C \perp\!\!\!\perp D \mid (A, B)$ does not follow from the first two. Also it is not clear whether "implied by" should be understood as implication from semigraphoid axioms (Dawid, 1979a; Pearl, 1988), probabilistic implication, graph separation implication or multivalued dependency implication. In general such implications can be quite different (Studený, 1992; Matúš, 1992a; Matúš and Studený, 1993; Studený, 1993). Similarly we doubt that positive probabilistic independence is graph-generated in full generality; see the same references.

3. MODEL CONSTRUCTION

Glymour and Spirtes raise a number of issues concerning automatic data-based model construction. We agree that there is a need for a systematic and thorough investigation of these issues and believe that much activity is to be expected in this area in the near future.

The problems involved are, however, quite difficult, although some impressive experimental results have appeared. Wedelin (1993) seems to be inaccurately quoted. Although DAGs are discussed in his paper, the part of his algorithm identifying directions is not discussed in depth and he searches in effect for an undirected structure. It is this undirected structure that his method would in principle find asymptotically. The last statement is less interesting than it sounds. It demands an infinite database and an infinite amount of computer time and this asymptotic correctness is shared by many other methods.

BIFROST would simply not be able to deal with the Alarm network which is so impressively reconstructed by Wedelin's and other methods. BIFROST uses a backward search strategy, searching from the fully connected network and trying to simplify, and is not designed to deal with large networks. However, BIFROST is only a master program for CoCo (Badsberg, 1991). Using CoCo directly, with an initially forward strategy that exploits sparseness of the network, reconstructs the (moral graph of) Alarm network at a level similar to but not quite as well as Wedelin. The results are, however, not fully comparable. The Minimum Description Length criterion needs calibration (this is easier for simulated data where the true network is known), and the results reported by Wedelin were based on a calibrated run. The CoCo reconstruction which we refer to was performed on Wedelin's simulated data using an uncalibrated BIC criterion which is asymptotically equivalent to MDL. But is 10,000 close to infinity?

Anyway, it is clear that many of these recent simulation experiments seem to indicate the possibility of reconstructing network structures of a size and complexity that hitherto would have been unthinkable. A point of caution may be appropriate: the simulation experiments tell little about the behaviour of automatic methods in real situations, trying to find structures that may only be there in an approximate sense.

4. TREES OR NETWORKS?

Glymour and Spirtes, Dempster and Hill all contrast the diagnostic performance of the network models with that of the hand-crafted algorithm reported in Franklin et al. (1991). First we should say that the domain of paediatric cardiology appears well suited to algorithmic solution – particular configurations of well-defined features pointing strongly to specific diagnoses. We are investigating data-based algorithms using the tree command in S (Chambers and Hastie, 1992), and finding prospective performance equivalent to the expert-based classification trees. However, in defence of the network models, we should emphasize that Table 6 reports performance while the network is being sequentially trained. When the penalized-EM procedure is used for batch learning (see subsection 7.1) and then reapplied to the training set of 168 cases, the (admittedly slightly optimistic) accuracy is 137 compared to the expert-algorithm's 132. As mentioned in subsection 5.4, fully prospective evaluation of trained networks gives similar performance.

Nevertheless, we should not be surprised if a tree-based approach gives better classification, since it is solely designed for that task. In contrast, a graphical model is just that: a full model for the relationship between all variables, without *Disease?* being singled out as particularly important, and thus a network can be used for prediction of any quantity. A classification tree can be thought of as knowledge *compiled* for a particular purpose: indeed a classification tree can be derived from a network by introducing, for example, an entropy-based mechanism for selecting successive questions to ask. If users are solely interested in classification then we might recommend tree-based procedures particularly since, as Dempster emphasizes, they often fit within the recognized pattern of teaching and discourse. The network models appear more appropriate when an underlying mechanism is postulated, and interest is in the interacting components of a whole system.

5. GENERAL ISSUES IN USING DAGs

Dempster makes some important comments on the explicit consideration of context in the structuring of the model, particularly with regard to selection of cases. Dawid (1976) contrasts direct modelling in the *diagnostic* direction, from disease to symptoms, to the *sampling* models we use in the paper. The former should be more robust to variations in patient selection on the basis of their clinical findings, providing further support for tree-based structures if interested only in classification. Within the network paradigm we could explicitly model the selection process by including a node indicating *Included in sample?* which has as parents those clinical features that influence referral. A classic occurrence of selective reporting is in the context of adverse drug reactions: Cowell et al. (1993) provide an example of explicitly representing the reporting process within a graphical model. Unfortunately the construction of conditional independence graphs is complicated by selective reporting, since in general conditional independence properties may be somewhat different within selected and unselected populations.

We fully agree with Dempster that subject matter expertise should be exploited wherever it is available, but that it should be subject to as much critical evaluation as given to data. The ability to create parsimonious and yet realistic models using evidence from a variety of sources is a skill that is difficult to formalise, and clearly different strategies exist. Dempster feels our approach has been too "discrete," and more smoothing may be possible through hierarchical modelling, Within our context this could mean, for example, not assuming "local independence" when learning and instead regarding the underlying frequencies exchangeable over different parent configurations. We have admittedly somewhat constrained ourselves to seeing how far we can push analytic methods, but in practice we would support the use of hierarchical models where appropriate, using MCMC techniques for learning and evidence propagation. We do, however, diverge from Dempster with regard to belief functions, and do not agree that the (now essentially solved) computational issues have been the main hindrance to their use. We still have strong reservations regarding the essential interpretation of these quantities, their elicitation and use in initialising *directed* structures, their criticism in the light of data, and in particular their interface with external environments involving calibration or decision making.

Madigan discusses the practical issues in using networks in AI applications. We agree that only a limited type of knowledge can be directly represented by a graphical model, and integration into hybrid systems is often appropriate. The specification of a joint distribution, even in a transparent and computationally efficient manner, can only be one component in a solution to a complex problem. We look forward to the development of general toolkits for graphical models, which should incorporate tools for model criticism as well as construction. As Normand points out, such criticism is a complex matter and difficult to automate, although standard statistical ideas of residuals, influence and fit all carry over.

Madigan also suggests averaging predictions over models when there is uncertainty concerning structure. As we show in subsection 5.2, our global monitors based on scoring rules may be transformed into Bayes factors for model comparison, which can in turn be transformed to posterior probabilities on alternative

structures. These posterior probabilities could then weight predictions on common nodes. As Madigan and others have shown, such weighted predictions can often perform well, and perhaps should have been included in our evaluation.

ADDITIONAL REFERENCES

ALMOND, R. G., BRADSHAW, J. M. and MADIGAN, D. (1993). Reuse and sharing of graphical belief network components. In *Proceedings of the Fourth International Workshop on Artificial Intelligence and Statistics*. To appear.

BARTHOLOMEW, D. (1987). *Latent Variable Models and Factor Analysis*. Oxford Univ. Press.

BEERI, C., FAGIN, R., MAIER, D. and YANNAKAKIS, M. (1983). On the desirability of acyclic database schemes. *J. Assoc. Comput. Mach.* **30** 479-513.

BEERI, C. and KIFER, M. (1986a). An integrated approach to logical database design of relational database schemes. *ACM Trans. Database Systems* **11** 134-158.

BEERI, C. and KIFER, M. (1986b). Elimination of intersection anomalies from database schemes. *J. Assoc. Comput. Mach.* **33** 423-450.

BEERI, C. and KIFER, M. (1987). A theory of intersection anomalies in relational database schemes. *J. Assoc. Comput. Mach.* **34** 544-577.

BEINLICH, I., SUERMONDT, H., CHAVEZ, R. and COOPER, G. (1989). The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks. In *Proceedings of the Second European Conference on Artificial Intelligence in Medicine* 247-256.

BERNARDINELLI, L. and MONTOMOLI, C. (1992). Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statistics in Medicine* **11** 983-1007.

BLALOCK, H. (1961). *Causal Inferences in Nonexperimental Research*. Univ. North Carolina Press.

BRADSHAW, J. M., CHAPMAN, C. R., SULLIVAN, K. M., ALMOND, R. G., MADIGAN, D., ZARLEY, D., GAVRIN, J., NIMS, J. and BUSH, N. (1993). KS-3000: An application of DDUCKS to bone-marrow transplant patient support. In *Proceedings of the Seventh Annual European Knowledge Acquisition Workshop*, Toulouse, France. To appear.

BRADSHAW, J. M., COVINGTON, S. P., RUSSO, P. J. and BOOSE, J. H. (1991). Knowledge acquisition techniques for decision analysis using AXOTL and AQUINAS. *Knowledge Acquisition* **3** 49-77.

BREESE, J. S. (1989). Construction of belief and decision networks. Technical Memorandum 30, Rockwell Palo Alto Laboratory, 444 High Street, Palo Alto, CA 94301.

CASDAGLI, M. (1992). Chaos and deterministic versus stochastic non-linear modelling. *J. Roy. Statist. Soc. Ser. B* **54** 303-328.

CHAMBERS, J. M. and HASTIE, T. J. (1992). *Statistical Models in S*. Wadsworth, Belmont, CA.

CHENG, B. and TITTERINGTON, D. M. (1993). Neural networks and statistical perspectives. Unpublished manuscript.

COWELL, R. G., DAWID, A. P., HUTCHINSON, T. A., RODEN, S. and SPIEGELHALTER, D. J. (1993). Bayesian networks for the analysis of drug safety. *The Statistician* To appear.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

DAWID, A. P. (1976). Properties of diagnostic data distributions. *Biometrics* **32** 647-658.

DAWID, A. P. (1979b). Some misleading arguments involving conditional independence. *J. Roy. Statist. Soc. Ser. B* **41** 249-252.

DEMPSTER, A. P. (1990). Causality and statistics. *J. Statist. Plann. Inference* **25** 261-278.

DEMPSTER, A. P. and KONG, A. (1988). Uncertain evidence and artificial analysis. *J. Statist. Plann. Inference* **20** 355-368. [Reprinted in G. SHAFER and J. PEARL, eds. (1990). *Readings in Uncertain Reasoning* 522-528. Morgan Kaufmann, San Mateo, CA.]

DEMPSTER, A. P., SCHATZOFF, M. and WERMUTH, N. (1977). A simulation study of alternatives to least squares. *J. Amer. Statist. Assoc.* **72** 77-91.

DONOHO, D. L., JOHNSTONE, I. M., HOCH, J. C. and STERN, A. S. (1992). Maximum entropy and the nearly black object. *J. Roy. Statist. Soc. Ser. B* **54** 41-81.

DRAPER, D. (1993) Assessment and propagation of model uncertainty. Technical report, Interdivisional Program in Statistics, Univ. California, Los Angeles.

EDWARDS, D. and KREINER, S. (1983). The analysis of contingency tables by graphical models. *Biometrika* **70** 553-565.

FAGIN, R., MENDELZON, A. O. and ULLMAN, J. F. (1982). A simplified universal relation assumption and its properties. *ACM Trans. Database Systems* **7** 343-360.

FISHER, F. M. (1970). A correspondence principle for simultaneous equation models. *Econometrica* **38** 73-92.

FISHER, R. A. (1951). *The Design of Experiments*. Oliver and Boyd, Edinburgh.

FITZMAURICE, G. M. and LAIRD, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika* **80** 141-151.

GEIGER, D. (1990). Graphoids: a qualitative framework for probabilistic inference. Ph.D. dissertation, Univ. California, Los Angeles.

GILKS, W. R., CLAYTON, D. G., SPIEGELHALTER, D. J., BEST, N. G., NCNEIL, A. J., SHARPLES, L. D. and KIRBY, A. J. (1993). Modelling complexity: applications of Gibbs sampling in medicine. *J. Roy. Statist. Soc. Ser. B* **55** 39-52.

GOLDMAN, R. P. and BREESE, J. S. (1992). Integrating model construction and evaluation. In *Uncertainty in Artificial Intelligence* (D. Dubois, M. P. Wellman, B. D'Ambrosio and P. Smets, eds.) **8** 104-111. Morgan Kaufman, San Mateo, CA.

GOLDMAN, R. P. and CHARNIAK, E. (1992). Probabilistic text understanding. *Statistics and Computing* **2** 105-114.

GOLDSZMIDT, M. and PEARL, J. (1992). Default ranking: A practical framework for evidential reasoning, belief revision and update. In *Proceedings of the Third International Conference on Knowledge Representation and Reasoning* 661-672. Morgan Kaufmann, San Mateo, CA.

HAMMERSLEY, J. M. (1974). Comment on "Spatial interaction and statistical analysis of lattice systems" by J. Besag. *J. Roy. Statist. Soc. Ser. B* **36** 230-231.

HARRISON, P. J. and STEVENS, C. F. (1976). Bayesian forecasting. *J. Roy. Statist. Soc. Ser. B* **38** 205-247.

HILL, J. R. (1991). Relational databases: A tutorial for statisticians. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface* (E. Kerimidas, ed.) 86-93. Interface Foundation, Fairfax Station, VA.

HODGES, J. S. (1987). Uncertainty, policy analysis and statistics (with discussion). *Statist. Sci.* **2** 259-291.

HOLLAND, P. W. (1986). Statistics and causal inference (with discussion). *J. Amer. Statist. Assoc.* **81** 945-970.

HOLTZMAN, S. (1989). *Intelligent Decision Systems*. Addison-Wesley, Reading, MA.

KADANE, J. and SEIDENFELD, T. (1992). Statistical issues in the analysis of data gathered in the new designs. In *Toward a More Ethical Clinical Trial* (J. Kadane, ed.). Wiley, New York. To appear.

KASS, R. E. and RAFTERY, A. E. (1993). Bayes factors and model

uncertainty. Technical Report 571, Dept. Statistics, Carnegie Mellon Univ.

KATZ, B. P. and HUI, S. L. (1989). Variance estimation for medical decision analysis. *Statistics in Medicine* 8 229–241.

KIIVERI, H. and SPEED, T. P. (1982). Structural analysis of multivariate data: A review. In *Sociological Methodology, 1982* (S. Leinhardt, ed.) 209–289. Jossey Bass, San Francisco.

LAIRD, N. M. and WARE, J. H. (1982). Random effects models for longitudinal data. *Biometrics* 38 963–974.

LIANG, K. Y., ZEGER, S. L. and QAQISH, B. (1992). Multivariate regression analyses for categorical data (with discussion). *J. Roy. Statist. Soc. Ser. B* 54 3–40.

LINDLEY, D. V. and SMITH, A. F. M. (1972). Bayes estimates for the linear model. *J. Roy. Statist. Soc. Ser. B* 34 1–18.

MADIGAN, D. and RAFTERY, A. E. (1991). Model selection and accounting for model uncertainty in graphical models using Occam's window. Technical Report 213, Dept. Statistics, Univ. Washington.

MADIGAN, D. and YORK, J. (1993). Bayesian graphical models. Technical Report, Dept. Statistics, Univ. Washington.

MASON, S. (1956). Feedback theory: Further properties of signal flow graphs. *Proc. Inst. Radio Eng.* 44 920–926.

MATÚŠ, F. (1992a). Ascending and descending conditional independence relations. In *Transactions of the 11th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, 189–200. Kluwer, Dordrecht.

MATÚŠ, F. (1992b). On equivalence of Markov properties over undirected graphs. *J. Appl. Probab.* 29 745–749.

MATÚŠ, F. and STUDENÝ, M. (1993). Conditional independences among four random variables I. Unpublished manuscript.

MORRIS, C. N. (1987). Comment on "The calculation of posterior distributions by data augmentation" by M. A. Tanner and W. H. Wong. *J. Amer. Statist. Assoc.* 82 542–543.

MOUSSOURIS, J. (1974). Gibbs and Markov random systems with constraints. *J. Statist. Phys.* 10 11–33.

NORMAND, S. L. and TRITCHLER, D. (1992). Parameter updating in a Bayes network. *J. Amer. Statist. Assoc.* 87 1109–1115.

PEARL, J. (1993a). Belief networks revisited. *Artificial Intelligence* 59 49–56.

PEARL, J. (1993b). From conditional oughts to qualitative decision theory. In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence* (D. Heckerman and Mamdani, eds.) 12–20. Morgan Kaufmann, San Mateo, CA.

PEARL, J. (1993c). Aspects of graphical models connected with causality. In *Proceedings of 49th Session, International Statistical Institute: Invited Papers*. To appear.

PRATT, J. and SCHLAIFER, R. (1988). On the interpretation and observation of laws. *J. Econometrics* 39 23–52.

RAFTERY, A. E. (1988). Approximate Bayes factors for generalized linear models. Technical Report 121, Dept. Statistics, Univ. Washington.

RAFTERY, A. E. (1993). Bayesian model selection in structural equation models. In *Testing Structural Equation Models* (K. A. Bollen and J. S. Long, eds.). Sage, Beverly Hills, CA.

REICHENBACH, H. (1956). *The Direction of Time*. Univ. California Press, Berkeley, CA.

RIPLEY, B. O. (1993). Statistical aspects of neural networks. *Proc. Semstat.*

RODGERS, R. and MARANTO, C. (1989). Causal models of publishing productivity in psychology. *Journal of Applied Psychology* 74 636–649.

ROSENBAUM, P. and RUBIN, D. (1983). The central role of propensity score in observational studies for causal effects. *Biometrika* 70 41–55.

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology* 66 688–701.

RUBIN, D. B. (1977). Assignment to treatment groups on the basis of a covariate. *Journal of Educational Statistics* 2 1–26.

RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* 6 34–58.

RUBIN, D. B. (1980). Comment on "Randomization analysis of experimental data: The Fisher randomization test" by D. Basu. *J. Amer. Statist. Assoc.* 75 591–593.

SHAFER, G. (1976). *A Mathematical Theory of Evidence*. Princeton Univ. Press.

SHAFFER, C. (1993). Selecting a classification method by cross-validation. *Fourth International Workshop on Artificial Intelligence and Statistics*. To appear.

SIMON, H. (1954). Spurious correlation: a causal interpretation. *J. Amer. Statist. Assoc.* 49 467–479.

SIMON, H. A. (1977). *Models of Discovery: and Other Topics in the Methods of Science*. Reidel, Dordrecht, Holland.

SKUCE, D. (1991). A frame-like knowledge acquisition tool integrating abstract data types and logic. In *Principles of Semantic Networks* (J. Sowa, ed.) 543–563. Morgan Kaufmann, San Mateo, CA.

SMITH, J. Q. (1988). Models, optimal decisions and influence diagrams. In *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 765–776. Oxford Univ. Press.

SOBEL, M. E. (1990). Effect analysis and causation in linear structural equation models. *Psychometrika* 55 495–515.

SOBEL, M. E. (1992). Causation and spurious correlation: A reexamination. Unpublished manuscript.

SOBEL, M. E. (1994). Causal inference in the social and behavioral sciences. In *A Handbook for Statistical Modeling in the Social and Behavioral Sciences* (G. Arminger, C. C. Clogg and M. E. Sobel, eds.) Plenum Press, New York. To appear.

SPIRTES, P. (1992). Building causal graphs from statistical data in the presence of latent variables. In *Proceedings of the Ninth International Congress on Logic, Methodology, and the Philosophy of Science* (B. Skyrms, ed.) Uppsala, Sweden.

SPIRTES, P. and GLYMOUR, C. (1993). Inference, intervention, and prediction. In *Fourth International Workshop on Artificial Intelligence and Statistics*. To appear.

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (1991). An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review* 9 62–72.

SPIRTES, P., SCHEINES, R., MEEK, C. and GLYMOUR, C. (1993). *Tetrad II: Tools for Discovery*. Erlbaum, Hillsdale, NJ. To appear.

SPIRTES, P. and VERMA, T. (1992). Equivalence of causal models with latent variables. Philosophy, Methodology and Logic Technical Report 33, Carnegie Mellon Univ.

STONE, R. (1993). The assumptions on which causal inferences rest. *J. Roy. Statist. Soc. Ser. B* 55 455–466.

STROTZ, R. H. and WOLD, H. O. A. (1960). Recursive vs. nonrecursive systems: An attempt at synthesis. *Econometrica* 28 417–427.

STUDENÝ, M. (1992). Conditional independence relations have no finite complete characterization. In *Transactions of the 11th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes* 377–396. Kluwer, Dordrecht, Holland.

STUDENÝ, M. (1993). Structural semigraphoids. *Internat. J. Gen. Systems*. To appear.

SUPPES, P. (1970). *A Probabilistic Theory of Causality*. North-Holland, Amsterdam.

SZOLOVITS, P. and PAUKER, S. G. (1978). Categorical and probabilistic reasoning in medical diagnosis. *Artificial Intelligence* 11 115–144.

U.S. NUCLEAR REGULATORY COMMISSION. (1990). *Severe Accident*

Risks: An Assessment for Five U. S. Nuclear Power Plants. Final Summary Report (NUREG-1150). Washington, D.C.

VERMA, T. and PEARL, J. (1990). On equivalence of causal models. Technical Report R-150, Dept. Computer Science, Univ. California, Los Angeles.

VOROB'EV, N. N. (1962). Consistent families of measures and their extensions. *Theory Probab. Appl.* 7 147–163.

WEDELIN, D. (1993). Discovering causal structure from data. Technical Report, Dept. Computer Science, Chalmers Univ. of Technology, Sweden.

WINER, B. J. (1971). *Statistical Principles in Experimental Design*, 2nd ed. McGraw-Hill, New York.

ZHAO, L. P. and PRENTICE, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika* 77 642–648.