Bibliography

[1] J. Amsterdam. Temporal reasoning and narrative conventions. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR' 91)*, pages 15–21, 1991.

[2] M. Anthony and N.Biggs. *Computational Learning Theory*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1992.

[3] K.R. Apt. *From Logic Programming to Prolog*. Prentice Hall, 1997.

[4] K.R. Apt and M. Bezem. Acyclic programs. *New Generation Computing*, 1991.

[5] F. Bacchus, A.J. Grove, J.Y. Halpern, and D. Koller. From statistical knowledge bases to degrees of belief. *Artificial Intelligence*, pages 75–143, 1997.

[6] L.J. Bain and M. Engelhardt. *Introduction to Probability and Mathematical Statistics*. PWS-Kent, Boston, 1989.

[7] A.B. Baker. A simple solution to the Yale shooting problem. In *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning (KR' 89)*, pages 11–20, 1989.

[8] A.B. Baker. Nonmonotonic reasoning in the framework of situation calculus. *Artificial Intelligence*, 49:5–23, 1991.

[9] V. Balasubramanian. Statistical inference, Occam's Razor, and statistical mechanics on the space of probability distributions. *Neural Computation*, 9:349–368, 1997.

[10] C. Baral and M. Gelfond. Logic programming and knowledge representation. *Journal of Logic Programming*, 19,20:73–148, 1994.

[11] C. Baral and M. Gelfond. Reasoning about actual and hypothetical occurrances of concurrent and non-deterministic actions. In *Proceedings First Workshop on Nonmonotonic Reasoning, Action and Change*, 1995. Available via ftp://ftp.newcastle.edu.au/pub/papers/nrac95/baral.final.ps.

[12] C. Baral, M. Gelfond, and A. Provetti. Representing actions: Laws, observations and hypotheses. *Journal of Logic Programming*, 12:1–44, 1997.

[13] A.R. Barron. Complexity regularization with application to artificial neural networks. In G. Roussas, editor, *Nonparametric Functional Estimation and Related Topics*, pages 561–576, Dordrecht, 1990. Kluwer Academic Publishers.

[14] A.R. Barron and T.M. Cover. Minimum complexity density estimation. *IEEE Transactions on Information Theory*, 37(4):1034–1054, 1991.

[15] R.A. Baxter and J.O. Oliver. MDL and MML: Similarities and differences. Technical Report 207, Department of Computer Science, Monash University, 1994.

[16] A. Berger, S. Della Pietra, and V. Della Pietra. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1), 1996.

[17] J.O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics. Springer-Verlag, New York, revised and expanded second edition, 1985.

[18] J.M. Bernardo and A.F.M Smith. *Bayesian theory*. John Wiley, 1994.

[19] C. Boutilier and M. Goldszmidt. The frame problem and Bayesian network action representations. In *Proceedings of the Canadian Conference on Artificial Intelligence (CCAI-96)*, 1996.

[20] H. Bozdogan. Personal Communication.

[21] G. Brewka. *Nonmonotonic Reasoning: Logical Foundations of Commonsense*, volume 12 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 1991.

[22] M. Browne. Cross-validation methods. *Journal of Mathematical Psychology*, 1998. To appear.

[23] G.J. Chaitin. On the length of programs for computing finite binary sequences: statistical considerations. *Journal of the ACM*, 16:145–159, 1969.

[24] P. Cheeseman. In defense of probability. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence (IJCAI-85)*, pages 1002–1009, 1985.

[25] P. Cheeseman. An inquiry into computer understanding. *Computational Intelligence*, 4(1):58–66,67–129, 1988. Discussion: pages 67–129.

[26] B.S. Clarke and A.R. Barron. Information-theoretic asymptotics of Bayes methods. *IEEE Transactions on Information Theory*, IT-36(3):453–471, 1990.

[27] B.S. Clarke and A.R. Barron. Jeffreys' prior is asymptotically least favorable under entropy risk. *Journal of Statistical Planning and Inference*, 41:37–60, 1996.

[28] G.F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42(2–3):393–405, 1990.

[29] G.F. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:309–347, 1992.

[30] T.M. Cover and J.A. Thomas. *Elements of Information Theory.* Wiley Interscience, New York, 1991.

[31] R.T. Cox. *The Algebra of Probable Inference.* Johns Hopkins University Press, 1961.

[32] J.M. Cozzolino and M.J. Zahner. The maximum-entropy distribution of the future market price of a stock. *Operations Research*, 21:1200–1211, 1973.

[33] A. Darwiche and J. Pearl. Symbolic causal networks. In *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*, 1994.

[34] A. Darwiche and J. Pearl. Symbolic causal networks for reasoning about actions and plans. In *Working notes of the AAAI Spring Symposium on Decision-Theoretic Planning*, 1994.

[35] A.P. Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series A*, 41(1):1–31, 1979.

[36] A.P. Dawid. Present position and potential developments: Some personal views, statistical theory, the prequential approach. *Journal of the Royal Statistical Society, Series A*, 147(2):278–292, 1984.

[37] A.P. Dawid. Prequential analysis, stochastic complexity and Bayesian inference. In J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics*, volume 4, pages 109–125. Oxford University Press, 1992. Proceedings of the Fourth Valencia Meeting.

[38] B. de Finetti. *Theory of Probability. A critical introductory treatment.* John Wiley & Sons, London, 1974.

[39] M.H. DeGroot. *Optimal statistical decisions.* McGraw-Hill, 1970.

[40] W. Derkse. *On Simplicity and Elegance: An Essay in Intellectual History.* PhD thesis, University of Amsterdam, 1993.

[41] D. Dubois and H. Prade. *Possibility theory: an approach to computerized processing of uncertainty.* Plenum, New York, 1988.

[42] R.O. Duda and P.E. Hart. *Pattern classification and scene analysis.* John Wiley, 1973.

[43] C. Elkan. A perfect logic for reasoning about action. manuscript, 1989. University of Toronto.

[44] R.L. Ellis. Remarks on an alleged proof of the method of least squares, contained in a late number of the *edinburgh review.* In W. Walton, editor, *Mathematical and other Writings of R.L. Ellis*, pages 53–61. Cambridge University Press, Cambridge, 1863.

[45] C. Evans. Negation-as-failure as an approach to the hanks and mcdermott problem. In *Proceedings of the Second International Symposium on Artificial Intelligence*, Monterrey, Mexico, 1989.

[46] M. Feder. Maximum entropy as a special case of the minimum description length criterion. *IEEE Transactions on Information Theory*, 32(6):847–849, 1986.

[47] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley, 1968. Third edition.

[48] R.A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A*, 222:309–368, 1925.

[49] K. Friedman and A. Shimony. Jaynes' maximum entropy prescription and probability theory. *Journal of Statistical Physics*, 9:265–269, 1971.

[50] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29:131–163, 1997.

[51] H. Geffner. Causal theories for nonmonotonic reasoning. In *Proceedings of the Eighth National Conference on Artificial Intelligence (AAAI-90)*, pages 524–530, 1990.

[52] H. Geffner. Causality, constraints and the indirect effects of actions. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97)*, pages 555–560, 1997.

[53] M. Gelfond and V.L. Lifschitz. Classical negation in logic programs and disjunctive databases. *New Generation Computing*, 9:365–385, 1991.

[54] M. Gelfond and V.L. Lifschitz. Representing actions in extended logic programming. In K.R. Apt, editor, *Proceedings of the Tenth Joint International Conference and Symposium on Logic Programming*, pages 559–573, 1992.

[55] S. Geman, E. Bienenstock, and R. Doversat. Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1):1–58, 1992.

[56] M.L. Ginsberg, editor. *Readings in Nonmonotonic Reasoning*. Morgan Kaufmann, Los Altos, California, 1987.

[57] E. Giunchiglia and V. Lifschitz. Dependent fluents. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, pages 1964–1969, 1995.

[58] M. Goldszmidt, P. Morris, and J. Pearl. A maximum entropy approach to nonmonotonic reasoning. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 15(3):220–232, 1993.

[59] M. Goldszmidt and J. Pearl. Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence*, 84(1-2):57–112, 1996.

[60] P.D. Grünwald. Causal networks and nonmonotonic temporal reasoning. In *Proceedings Eighth Dutch Conference on Artificial Intelligence (NAIC-96)*, pages 157–166, Utrecht, 1996.

[61] P.D. Grünwald. A minimum description length approach to grammar inference. In G. Scheler S. Wermter, E. Riloff, editor, *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*, number 1040 in Springer Lecture Notes in Artificial Intelligence, pages 203–216. 1996.

[62] P.D. Grünwald. Causation and nonmonotonic temporal reasoning. In G. Brewka, C. Habel, and B. Nebel, editors, *KI-97: Advances in Artifical Intelligence*, number 1303 in Springer Lecture Notes in Artifical Intelligence, pages 159–170, 1997.

[63] P.D. Grünwald. Nonmonotonic temporal reasoning as a search for explanations. In *Proceedings NRAC '97 (Second IJCAI Workshop on Nonmonotonic Reasoning, Action and Change)*, pages 91–102, Nagoya, Japan, 1997.

[64] P.D. Grünwald. The sufficient cause principle and reasoning about action. Technical Report INS-R9709, CWI Amsterdam, 1997. Available at ftp://ftp.cwi.nl/pub/pdg/R9709.ps.Z.

[65] P.D. Grünwald, P. Kontkanen, P. Myllymäki, T. Silander, and H. Tirri. Minimum encoding approaches for predictive modeling. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI '98)*, 1998.

[66] J. Gustafsson and P. Doherty. Embracing occlusion in specifying the indirect effects of actions. In *Proceedings of the Fifth International Conference on Principles of Knowledge Representation and Reasoning (KR' 96)*, 1996.

[67] S. Hanks and D. McDermott. Default reasoning, non-monotonic logics and the frame problem. In *Proceedings of the Fifth National Conference on Artificial Intelligence (AAAI-86)*, pages 328–333, 1986.

[68] B.A. Haugh. Simple causal mimimizations for temporal persistence and projection. In *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87)*, pages 218–223, 1987.

[69] D. Haussler, M. Kearns, H.S. Seung, and N.Z. Tishby. Rigorous learning curve bounds from statistical mechanics. *Machine Learning*, 25:195–236, 1996.

[70] D. Heckerman. A tutorial on learning with Bayesian networks. Technical Report MSR-TR-95-06, Microsoft Research, Advanced Technology Division, One Microsoft Way, Redmond, WA 98052, 1996.

[71] D. Heckerman, D. Geiger, and D.M. Chickering. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20(3):197–243, 1995.

[72] J. Hertz, A. Krogh, and R.G. Palmer. *Introduction to the theory of neural computation*. Lecture Notes of the Santa Fe Institute. Addison-Wesley, 1991.

[73] E.T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, 1957.

[74] E.T. Jaynes. Information theory and statistical mechanics. In K.W. Ford, editor, *Statistical Physics (1962 Brandeis Lectures)*, pages 181–218, New York, 1963. W.A. Benjamin, Inc.

[75] E.T. Jaynes. Where do we stand on maximum entropy? In R.D. Levine and M. Tribus, editors, *The Maximum Entropy Formalism*, pages 15–118. MIT Press, Cambridge, MA, 1978.

[76] E.T. Jaynes. *Papers on Probability, Statistics and Statistical Physics.* Kluwer Academic Publishers, second edition, 1989.

[77] E.T. Jaynes. Probability theory: the logic of science. Available at ftp://bayes.wustl.edu/Jaynes.book/, 1996. Jaynes' forthcoming monumental treatise on probability theory as extended logic. This preliminary version (1996) is available via the web.

[78] H. Jeffreys. *Theory of Probability.* Oxford University Press, London, third edition, 1961.

[79] F.V. Jensen. *An Introduction to Bayesian Networks.* Springer, New York, 1996.

[80] G. N. Kartha. Soundness and completeness theorems for three formalizations of action. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI-93)*, pages 724–729, 1993.

[81] G.N. Kartha. Two counterexamples related to Baker's approach to the frame problem. *Artificial Intelligence*, 69:379–391, 1994.

[82] R.E. Kass and P.W. Voss. *Geometrical Foundations of Asymptotic Inference.* Wiley Interscience, 1997.

[83] H.A. Kautz. The logic of persistence. In *Proceedings of the Fifth National Conference on Artificial Intelligence (AAAI-86)*, pages 401–405, 1986.

[84] M. Kearns, Y. Mansour, A.Y. Ng, and D. Ron. An experimental and theoretical comparison of model selection methods. *Machine Learning*, 27:7–50, 1997.

[85] A.N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer-Verlag, 1933.

[86] A.N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems Inform. Transmission*, 1(1):1–7, 1965.

[87] P. Kontkanen, P. Myllymäki, T. Silander, H. Tirri, and P.D. Grünwald. Bayesian and information-theoretic priors for Bayesian network parameters. In *Proceedings of the Tenth European Conference on Machine Learning (ECML-98)*, Chemnitz, Germany, 1998.

[88] L.G. Kraft. A device for quantizing, grouping and coding amplitude modulated pulses. Master's thesis, Dept. of Electrical Engineering, M.I.T., Cambridge, Mass., 1949.

[89] A. Krogh and G. Mitchison. Maximum entropy weighting of aligned sequences of proteins or DNA. In *Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology (ISMB '95)*, pages 215–221, 1995.

[90] P. Langley and S. Sage. Induction of selective Bayesian classifiers. In *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence (UAI '94)*, pages 399–406, 1994.

[91] P.S. Laplace. *Essai philosophique sur les Probabilités.* H. Remy, Bruxelles, fifth edition, 1829.

[92] S.L. Lauritzen. *Graphical Models.* Oxford Science Publications, 1996.

[93] M. Li and P.M.B. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications.* Springer-Verlag, New York, revised and expanded second edition, 1997.

[94] V.L. Lifschitz. Computing circumscription. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence (IJCAI-85)*, pages 121–127, 1985.

[95] V.L. Lifschitz. Formal theories of action. In M. L. Ginsberg, editor, *Readings in Nonmonotonic Reasoning.* Morgan Kaufmann Publishers, Los Altos, CA, 1987.

[96] V.L. Lifschitz. Pointwise circumscription. In M. L. Ginsberg, editor, *Readings in Nonmonotonic Reasoning.* Morgan Kaufmann Publishers, Los Altos, CA, 1987.

[97] V.L. Lifschitz. Frames in the space of situations. *Artificial Intelligence*, 46:365–376, 1990.

[98] V.L. Lifschitz and A. Rabinov. Miracles in formal theories of action. *Artificial Intelligence*, 38:225–237, 1989.

[99] F. Lin. Embracing causality in specifying the indirect effects of actions. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, 1995.

[100] F. Lin. Embracing causality in specifying the indeterminate effects of actions. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96)*, 1996.

[101] F. Lin and R. Reiter. State constraints revisited. *Journal of Logic and Computation*, 4(5):655–677, 1994.

[102] F. Lin and Y. Shoham. Provably correct theories of action (preliminary report). In *Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91)*, 1991.

[103] N. McCain and H. Turner. A causal theory of ramifications and qualifications. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, 1995.

[104] N. McCain and H. Turner. Causal theories of action and change. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI-97)*, 1997.

[105] N. McCain and H. Turner. On relating causal theories to other formalisms. Unpublished Manuscript, 1997.

[106] J. McCarthy. Circumscription – a form of non-monotonic reasoning. *Artificial Intelligence*, 13(1,2):27–39,171–172, 1980.

[107] J. McCarthy and P.J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Mitchie, editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, 1969.

[108] D. McDermott. A critique of pure reason. *Computational Intelligence*, 3(3):151–237, 1987. Discussion: pages 161–237.

[109] D. McDermott and J. Doyle. Non-monotonic logic I. *Artificial Intelligence*, 13:41–72, 1980.

[110] R.S. Michalski, J.G. Carbonell, and T.M. Mitchell. *Machine Learning, An Artificial Intelligence Approach*. Morgan Kaufmann, 1983.

[111] D. Michie, D.J. Spiegelhalter, and C.C. Taylor, editors. *Machine Learning, Neural and Statistical Classification*. Ellis Horwood, London, 1994.

[112] L. Morgenstern and L.A. Stein. Why things go wrong: A formal theory of causal reasoning. In *Proceedings of the Seventh National Conference on Artificial Intelligence (AAAI-88)*, pages 5–23, 1988.

[113] I.J. Myung. Maximum entropy interpretation of decision bound and context models of categorization. *Journal of Mathematical Psychology*, 38:335–365, 1994.

[114] J.B. Paris. *The Uncertain Reasoner's Companion*. Number 39 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1994.

[115] J.B. Paris and A. Vencovská. A note on the inevitability of maximum entropy. *International Journal of Approximate Reasoning*, 4(3):183–224, 1990.

[116] J. Pearl. *Probabilistic Reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[117] J. Pearl. Graphical models, causality, and intervention. *Statistical Science*, 8(3):266–273, 1993.

[118] J. Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–709, 1995.

[119] J. Pearl. Causation, action, and counterfactuals. Technical Report R-223-U, University of California, Los Angeles, Computer Science Department, 1995. Presented at UNICOM Seminar, London, April 3-5, 1995.

[120] J. Pearl. Causation, action and counterfactuals. In Y. Shoham, editor, *Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge (TARK-VI)*, pages 51–73. Morgan Kaufmann, 1996.

[121] D. Poole. What the lottery paradox tells us about default reasoning. In *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning (KR' 89)*, pages 333–340, 1989.

[122] J. Quinlan and R. Rivest. Inferring decision trees using the minimum description length principle. *Information and Computation*, 80:227–248, 1989.

[123] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings IEEE*, 77(2):257–286, 1989.

[124] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[125] B.D. Ripley. *Pattern Recognition and Neural Networks.* Cambridge University Press, 1996.

[126] J. Rissanen. Modeling by the shortest data description. *Automatica*, 14:465–471, 1978.

[127] J. Rissanen. Stochastic complexity. *Journal of the Royal Statistical Society, series B*, 49:223–239, 1987. Discussion: pages 252–265.

[128] J. Rissanen. *Stochastic Complexity in Statistical Inquiry.* World Scientific Publishing Company, 1989.

[129] J. Rissanen. Fisher information and stochastic complexity. *IEEE Transactions on Information Theory*, 42(1):40–47, 1996.

[130] J.S. Rowlinson. Probability, information and entropy. *Nature*, 225(1196), 1970.

[131] E.J. Sandewall. Filter preferential entailment for the logic of action in almost continuous worlds. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI-89)*, 1989.

[132] E.J. Sandewall. The range of applicability of nonmonotonic logics for the inertia problem. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI-93)*, 1993.

[133] E.J. Sandewall. *Features and Fluents.* Oxford University Press, 1994.

[134] E.J. Sandewall. Assessments of ramification methods that use static domain constraints. In *Proceedings of the Fifth International Conference on Principles of Knowledge Representation and Reasoning (KR' 96)*, 1996.

[135] E.J. Sandewall and Y. Shoham. Nonmonotonic temporal reasoning. In *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 4. Oxford University Press, 1995.

[136] L.K. Schubert. Explanation closure, action closure and the Sandewall test suite for reasoning about change. *Journal of Logic and Computation*, 4(5):679–700, 1994.

[137] R.D. Shachter. Probabilistic inference and influence diagrams. *Operations Research*, 36(4):589–604, 1988.

[138] G. Shafer. *A mathematical theory of evidence.* Princeton University Press, 1976.

[139] M. Shanahan. *Solving the Frame Problem – A Mathematical Investigation of the Common Sense Law of Inertia.* MIT Press, Cambridge, MA, 1997.

[140] J.L. Shapiro and A. Prügel-Bennett. Maximum entropy analysis of genetic algorithm operators. In *Springer Lecture Notes in Computer Science (Proceedings AISB-95)*, volume 993, pages 14–24, 1995.

[141] A. Shimony. The status of the principle of maximum entropy. *Synthese*, 63:35–53, 1985.

[142] Y. Shoham. A semantical approach to nonmonotonic logics. In M.L. Ginsberg, editor, *Readings in Nonmonotonic Reasoning*, pages 227–250. Morgan Kaufmann, Los Altos, California, 1987.

[143] Y. Shoham. Chronological ignorance: experiments in nonmonotonic temporal reasoning. *Artificial Intelligence*, 36:279–331, 1988.

[144] J.E. Shore and R.W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory*, IT-26:26–37, 1980.

[145] R.J. Solomonoff. A formal theory of inductive inference, part 1 and part 2. *Information and Control*, 7:1–22, 224–254, 1964.

[146] M. Steijvers and P.D. Grünwald. A recurrent network that performs a context-sensitive prediction task. In *Proceedings Eighteenth Annual Conference of the Cognitive Science Society*, San Diego, CA, 1996.

[147] L.A. Stein and L. Morgenstern. Motivated action theory: A formal theory of causal reasoning. *Artificial Intelligence*, 71(1):1–42, 1994.

[148] A. Stolcke. *Bayesian Learning of Probabilistic Language Models.* PhD thesis, ICSI, Berkeley, 1994.

[149] J. Takeuchi and A.R. Barron. Asymptotically minimax regret by Bayes mixtures. In *Proceedings of the 1998 International Symposium on Information Theory (ISIT 98)*, 1998.

[150] M. Thielscher. The logic of dynamic systems. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, pages 1956–1962, 1995.

[151] M. Thielscher. Ramification and causality. *Artificial Intelligence*, 89:317–364, 1997.

[152] Y. Tikochinsky, N.Z. Tishby, and R.D. Levine. Alternative approach to maximum entropy inference. *Physical Review A*, 30:2638–2644, 1984.

[153] Y. Tikochinsky, N.Z. Tishby, and R.D. Levine. Consistent inference of probabilities for reproducible experiments. *Physical Review Letters*, 52:1357–1360, 1984.

[154] H. Tirri, P. Kontkanen, and P. Myllymäki. Probabilistic instance-based learning. In *Proceedings of the Thirteenth International Conference on Machine Learning (ICML-96)*, pages 507–515, 1996.

[155] N.Z. Tishby. Statistical physics models of supervised learning. In D. Wolpert, editor, *The Mathematics of Generalization*, volume XX of *SFI Studies in the Sciences of Complexity*, pages 215–242. Addison-Wesley, 1995.

[156] D.S. Touretzky, J.F. Horty, and R.H. Thomason. A clash of intuitions: The current state of nonmonotonic multiple inheritance systems. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI-89)*, pages 476–482, 1989.

[157] J. Uffink. Can the maximum entropy principle be explained as a consistency requirement? *Studies in History and Philosophy of modern Physics*, 26B:223–261, 1995.

[158] J. Uffink. The constraint rule of the maximum entropy principle. *Studies in History and Philosophy of modern Physics*, 27:47–79, 1996.

[159] V.N. Vapnik. *Estimation of Dependences Based on Empirical Data*. Springer-Verlag, Berlin, 1982.

[160] P.M.B. Vitányi and M. Li. Minimum Description Length induction, Bayesianism, and Kolmogorov Complexity, 1997. Submitted to *IEEE Transactions on Information Theory.*

[161] C.S. Wallace and D.M. Boulton. An information measure for classification. *Computing Journal*, 11:185–195, 1968.

[162] C.S. Wallace and P.R. Freeman. Estimation and inference by compact coding. *Journal of the Royal Statistical Society, Series B*, 49:240–251, 1987. Discussion: pages 252–265.

[163] L. Wasserman. Bayesian model selection. *Journal of Mathematical Psychology*, 1998. To appear.

[164] S.S Wilks. *Mathematical statistics.* John Wiley, 1962.

[165] S. Wright. The method of path coefficients. *Annals of Mathematical Statistics*, 5:161–215, 1934.

[166] K. Yamanishi. On-line maximum likelihood prediction with respect to general loss functions. *Journal of Computer and System Sciences*, 55(1):105–118, 1997.

[167] K. Yamanishi. A decision-theoretic extension of stochastic complexity and its applications to learning. *IEEE Transactions on Information Theory*, 44(4):1424–1439, 1998.

List of Symbols

$\mathcal{A}(\mathbf{V})$: intervention variables, 211, 215
$\mathbf{A}, \mathbf{A}^n, \mathbf{A}^+, \mathbf{A}^*$: (sets made up of) data alphabet(s), 7
Ab: set of atoms representing abnormalities, 270
Ab, Ab_1, Ab_2: abnormalities, 184, 225
arg: argument of function, 16

B: $\mathbf{B} = \{0, 1\}$ (parts I and II), 7
B: $\mathbf{B} = \{\text{TRUE}, \text{FALSE}\}$ (part III), 210

C: code, 7
$C_{2\text{-}p}$: two-part code, 29
C_{sc}: stochastic complexity code, 36
C_{uni}: uniform code, 164
cov: covariance, 57
$\mathcal{C}$: set of constraints over a probability distribution, 50
$\mathbf{C}^n(\phi, t)$: set of data for which $\overline{\phi(x)}^n = t$, 53
$\mathcal{C}_e$: set of constraints over data, 53
Circum: circumscription, 226
CONS: set of constraints in causal theories, 210, 225

D: set of dependent fluents, 228
$\mathcal{D}$: decision space, 70
$D(\cdot||\cdot)$: relative entropy, 52
D: data sequence, 7
d: precision, 7
DC: Domain-Closure-Axioms, 225
Di: Dirichlet distribution, 140
Do: syntactic construct or predicate representing an intervention, 211
Do: syntactic construct or predicate representing an intervention, 224

$E_P[\phi(X)]$: expectation of $\phi(X)$ under distribution P, 51
E: sample space (parts I and II), 12
E: set of events (part III), 212

EQ: set of structural equations, 210, 225
ER: error function, 31
ER_{01}: 0/1-error function, 31
ER_{lg}: logarithmic error, 90
ER_{sq}: squared error function, 21
$\widetilde{\text{ER}}$: minimum expected error, 99

F: set of fluents, 212
F: normalizing sum, 37
f: density function, 12

$\mathcal{G}$: Bayesian network structure, 137
$\mathbf{G}^n(\gamma)$: set of sequences with frequencies γ, 58

$\mathcal{H}$: entropy, 52
H: hypothesis (model), 10
H_{mdl}, *see* c_{mdl}
$\widehat{H}$, *see* $\hat{c}$
$\widetilde{H}$, *see* $\tilde{c}$
H^*, *see* c^*
$\ddot{H}$, *see* $\ddot{c}$
Ho: 'holds', 188, 224
$\widehat{H}(D,\ \beta)$: maximum likelihood H for fixed β,91

$I(\theta), I(x^n; \theta)$: Fisher information matrix; observed information matrix, 40, 163
$\mathcal{I}$: indicator function, 51
Ind: set of atoms one is ignorant about, 270
int: interior of a set, 55

k_i: number of elements in sub-sample space $\mathbf{E}_i$, 130
K_{sc}: model cost, 36

$\mathcal{L}$: first-order language, 188
L: code length function, 12
$L_{2\text{-}p}$: two-part code length, 17
$L_{av,w}$: mixture approximation to stochastic complexity, 40

$\widetilde{L}$: minimum expected codelength, 96
L_{sc}: stochastic complexity, 36
ln: logarithm to base e, 12
log: logarithm to base 2, 12
LOSS: loss function, 32, 70
LOSS_{lg}: logarithmic loss, 44

$\mathcal{M}$: model class (parts I and II), 10, 17
$\mathcal{M}$: structure, interpretation or model of a logical theory (part III), 189, 210
$\mathbf{M}$: class of models for a logical theory, 212
$\mathcal{M}_{nb}$: Naive Bayes model class, 146
$\mathcal{M}_{me}$: maximum entropy model class, 55
$\mathbf{M}(\phi, t)$: set of P satisfying $E_P[\phi(X)] = t$, 53

$\mathbf{N}_0$: the nonnegative integers, 188, 224

$o(1)$: asymptotics notation, 40

P^*: 'true' distribution, 26, 34
$\mathcal{P}$: predictive distribution, 131
P: probability distribution, probabilistic model, 12, 13
Pa_i: set of parents of node i in a Bayesian network structure, 137
$\mathcal{P}_{av}$: evidence predictive distribution, 132
P_{av}: marginal distribution (Bayesian evidence), 39
$\mathcal{P}_{jef}$: Jeffrey's evidence predictive distribution, 136
P_{jef}: marginal distribution based on Jeffrey's prior, 134
$\mathcal{P}_{map}$: MAP predictive distribution, 132
P_{me}: maximum entropy distribution, 53
$\mathcal{P}_{sc}$: stochastic complexity predictive distribution, 133
P_{sc}: stochastic complexity distribution, 37

R: Rectangle, 114
$\mathbf{R}$: the real numbers, 12
$\mathcal{R}$: regret, 36

s: width between adjacent parameter values, 166
SCORE: score (predictive accuracy), 148

T: transpose, 53
T: logical theory / causal theory, 210
T_{C} : Two-point or first-order causal theory of our approach, 242, 252
T_{DF}: causal theory for the baby-and-the-table scenario, 228
$T_{\text{H-M}}$: naive theory for Yale Shooting Scenario, 190
T_{LIN} : logical theory in Lin's approach, 244
T_{MT} : domain description of McCain and Turner's approach, 242
T_{SW}: causal theory for basic two-switches scenario, 217
$T_{\text{SW},2}$: causal theory for advanced switches scenario, 218
T_{WT}: causal theory for walking turkey scenario, 216
T_{YSP}: causal theory for Yale Shooting Scenario, 227
$\mathbf{T}$: set of names for time points, 249

$\mathbf{U}$: range of a function appearing in a maximum entropy distribution (part I), 53
$\mathbf{U}$: unobserved propositional variables (part III), 210
$\vec{u}, \vec{U}$: vector of clamped (given) (random) variables, 131
UNA: Uniqueness-of-Names-Axioms, 225

$\mathbf{V}$: observable propositional variables, 210
$\mathbf{V_E}$: set of event–time pairs, 212

$\mathbf{V_F}$: set of fluent–time pairs, 212
$\vec{v}, \vec{V}$: vector of free (random) variables, 131

W: prior distribution over countable space., 39
w: prior density function, 40

Z, Z_H: normalizing factor ('partition function') in maximum entropy distribution, 53, 88

β: parameter vector for maximum entropy distribution, 53
$\hat{\beta}$, *see* $\hat{c}$
$\hat{\beta}(D,\ H)$: maximum likelihood β for fixed H,91
$\tilde{\beta}$, *see* $\tilde{c}$
δ: decision, 70
Γ: set of logical formulas (part III), 210
Γ: set of parameter values (parts I and II), 17
Γ_{nat}: natural parameter space, 56
γ: frequency distribution, 58
λ: empty sequence, 7
μ: hyperparameters for Dirichlet distribution, 140
ϕ: function appearing in maximum entropy distribution, 50
π: Jeffrey's prior, 41
$\hat{\sigma}$, *see* $\hat{c}$
$\tilde{\sigma}$, *see* $\tilde{c}$
$(\sigma^*)^2$, *see* c^*
σ^2: variance, 22
θ: parameter vector denoting a probabilistic model, 17
$\hat{\theta}$, *see* $\hat{c}$
$\breve{\theta}$, *see* $\breve{c}$
$\bar{\theta}$, *see* $\breve{c}$
$\ddot{\theta}$, *see* $\ddot{c}$
$\tilde{\theta}$, *see* $\tilde{c}$
θ^*, *see* c^*
θ_{mdl}, *see* c_{mdl}
$\hat{\theta}_{\mathrm{fut}}$: future-optimal model, 67
θ': MML estimator, 163
θ'_{p}: pointwise revised MML estimator, 165, 168
θ'_{v}: volumewise revised MML estimator, 167, 168
θ'_{wf}: MML WF estimator, 163
$\bar{c}$ where $c \in \{\theta, \beta, H\}$: mean of Bayesian posterior, 143
$\breve{c}$ where $c \in \{\theta, \beta, H\}$: Bayesian MAP estimator, 132
$\ddot{c}$ where $c \in \{\theta, \beta, H\}$: (unspecified) estimator, 67, 161
$\hat{c}$ where $c \in \{\theta, \beta, H, \sigma\}$: model that minimizes empirical error/ maximizes likelihood, 18
$\tilde{c}$ where $c \in \{\theta, \beta, H, \sigma\}$: model that minimizes expected error, 78, 97
c^* where $c \in \{\theta, \beta, H, \sigma\}$: 'true' model generating the data, 34
c_{mdl} where $c \in \{\theta, \beta, H, \sigma\}$: model that minimizes two-part code length, 16
$\overline{\phi(x)}^n$: average of ϕ over $x_1, \ldots, x_n$, 53
$\lceil \cdot \rceil$: ceiling, 12
$[K]$: interpretation of K, 189
$\models_{\mathbf{c}}$: causal entailment, 215, 226
$\models$: propositional and first-order entailment, 189, 210
$\mathbf{0}$: identically 0-function, 55
$\propto$: proportionality, 41
$|\cdot|$: determinant/absolute value (parts I and II), 40
$|\cdot|_c$: universe of sort c (part III), 224
$|\cdot|$: number of elements in set $\cdot$, 12
$\wedge$: logical AND, 184
$\vee$: logical OR, 184
$\supset$: material implication, 184
$\neg$: logical NOT, 184
$\langle \mathcal{M} \rangle$: 'entropification' of $\mathcal{M}$, 81

Index

abnormality, 184
action, **188**
alphabet, **7**
Apt, K.R., 192

Baral, C., 245–252
Barron, A.R., 104
Bayes, T., xi
Bayesian
  evidence, **39**, 45
  marginal distribution, *see* Bayesian evidence
  mixture, *see* Bayesian evidence
  statistics, 44–46
Bayesian networks, *see* model class, Bayesian networks
Bernoulli, *see* model class
Bezem, M., 192

causal chains example, 231
circumscription, 225, **234**
code, **7**
  efficient, **29**
  prefix, **13**
  redundant, **29**
  Shannon-Fano, **15**
  uniform, **164**
coin tossing, 222
common-sense reasoning, 187
completion mapping, 200
concentration phenomenon, 58–59
concept learning, *see* model class, of concepts
conditional independence, 137
conditional maximum likelihood estimator, *see* estimation, conditional
conjugate, *see* probability distribution, prior, conjugate
constraint
  empirical, **53**
Cox, R.T., xi
cross-validation, **148**, 148–149

data alphabet, *see* alphabet
Dawid, A.P., 44, 46
decision theory, 49
dependent fluent example, 228
description method, **7**
Dirichlet, *see* probability distribution, prior, Dirichlet
disjunctive effect example, 221

effect axiom, 188
Ellis, L., 61
entropification, **88**, 81–117
entropy, 52, **52**
  empirical, 58
  relative, 52
error function, **31**
  0/1-error, **31**, 89, 106
  logarithmic, 101–102
  logarithmic error, **44**
  simple, **93**, 98–101
  squared error, 11, 20, **31**, 82–85, 102–104
estimation
  conditional, **91**
  MAP, **45**
  maximum likelihood, 17, 132
  MDL, **18**
  MML, 161–168
event, **188**
evidence, *see* Bayesian evidence
expectation, **51**
exponential family, **56**
  dimension of, 56
  full, 56

Fisher information matrix, *see* information matrix
fluent, **188**
frame axiom, 188
frame problem, **188**
Fred, **190**
frequency distribution, 58

gambling, 42

Gaussian, *see* model class of normal distributions
Gelfond, M., 245–252

Hanks, S., 189
Hayes, P., 188
hypothesis, 38

i.i.d., 13
ignorance, 51
independence, 51
indicator function, 51
inertia, **188**
information matrix, 40, 135, 144–145, 163, 167–172
interior, 55
internal preference structure, 200

Jaynes, E.T., xi, 50, 76
Jeffrey's prior , *see* probability distribution, prior, Jeffrey's

Kolmogorov, A.N., 8
Kraft Inequality, **14**
Kullback-Leibler Divergence, *see* entropy, relative

Laplace, P.S., xi, 50
Li, M., 46
Lifschitz, V.L., 193
Lin, F., 242–245
loss function, **70**
  logarithmic loss, **44**, 66, 69

MAP, *see* estimation, MAP
maximum a posteriori, *see* estimation, MAP
maximum entropy distribution, 51
maximum likelihood, *see* estimation, maximum likelihood
McCain, N., 240–242
McCarthy, J., 188
McDermott, D., 189
MDL
  estimator, *see* estimation, MDL
  Model Selection Criterion, 41
  philosophy of, 5, 24–28
  predictive MDL, **44**
  refined two-part code MDL, **168**
  two-part code MDL, **16**
Minimum Description Length Principle, *see* MDL
  stochastic complexity version, *see* stochastic complexity
Minimum Message Length Principle, *see* MML
misspecified, 27
mixture, **40**
MML, 159–161
  estimators, *see* estimation, MML
model
  complete, 30
  complexity, **36**
  cost, **36**
  future optimal, 67
  intended, **190**
  non-probabilistic, 20, 30–34
  preferred, **184**
  probabilistic, **12**, 30–34
  selection, *see* model class selection
model class
  Bayesian networks, **139**, 137–145
  Bernoulli, **18**, 55, 106
  Markov chains, 25
  maximum entropy, **55**
  Naive Bayes, 172
  non-probabilistic, 23
  normal distribution, 22
  of concepts, **31**, 106
  probabilistic, 23
  selection, 30, 38, 41, 104
multinomial distribution, 140

naive persistence formalization, 190
normal distribution, *see* model class of normal distributions

odds, 42, 273
overfitting, 24–28

parameterization, 56
  mean-value, **56**
  natural, **56**

parent variable, 137
Pearl, J., 205–238
persistence, **188**
prediction, 30, 38, *see also* estimation
predictive distribution, *see* probability distribution, predictive
predictive MDL, *see* MDL, predictive
Principle of Insufficient Reason, 60
prior, *see* probability distribution, prior
probabilistic model, *see* model
probability distribution, *see also* model class
    generating, 23, 26, **34**, 39
    maximum entropy, 51
    modeling, 23, 26, **34**
    normal, 51
    predictive, **131**, 130–136, 142–145
    prior, **40**
        conjugate, 140
        Dirichlet, 140
        ESS, 172–174
        Jeffrey's, 41, 134–136, 144–145, 171
        subjective, 161, 172–174
        uniform, 132, 140, 172–174
    true, *see* probability distribution, generating
    uniform, 51
proportional betting, 42
Provetti, A., 245–252

ramification problem, 188, 193–195
regression, **21**
Regret, **36**
relative entropy, KL divergence, 52
reliable
    predictions, decisions, 67, **71**, 70–75
    under P, **96**
risky statistics, 75–79
Rissanen, J., 7–177

safe statistics, **78**, 75–79
sample space, **12**
    continuous, **12**
    discrete, **12**
Sandewall, E.J., 192
score, 147
Shanahan, M., 187
Shannon-Fano code, *see* code, Shannon-Fano
Solomonoff, R.J., 8
stochastic complexity, 30, **35**
    approximations of, **41**
    betting interpretation of, **42**
    mixture approximation, 39–40
stolen car, **191**, **232**
suitcase example, 217
supervised learning, **21**
support, 51
switches example, 217

Turner, H., 240–242
two-part code MDL, *see* MDL, two-part code

underfitting, 24–28
utility, 45

Vitányi, P.M.B., 46

walking turkey, 193, **216**
Wallace, C.S., 46, 159–177

Yale Shooting Problem, **190**, 227
Yamanishi, K., 105, 167

Nederlandse Samenvatting

Centraal in dit proefschrift staat het *Beginsel van de Minimale Beschrijvingslengte* (in het Engels 'Minimum Description Length Principle'; vanaf nu 'MDL Principe' genoemd). Het MDL Principe stelt 'leren' gelijk aan 'comprimeren'. In zijn eenvoudigste vorm ziet het er als volgt uit:

MDL Principe De beste hypothese om een verzameling gegevens te verklaren is de hypothese H die de som van ...

- de beschrijvingslengte van de hypothese H en
- de beschrijvingslengte van de gegevens, wanneer de gegevens beschreven worden met behulp van hypothese H,

... minimaliseert.

In deze vorm zorgt het MDL Principe voor een afweging tussen *complexiteit* van de hypothese en de *fout* die de hypothese maakt op de gegevens. Het MDL Principe kan worden toegepast op alle vormen van *inductieve inferentie*. Met 'inductieve inferentie' wordt bedoeld het postuleren van algemene wetmatigheden op grond van een beperkte hoeveelheid gegevens. In dit proefschrift richten we ons met name op het gebruik van MDL in de statistiek en in het vakgebied genaamd 'machinaal leren' (machine learning). Dit is het deelgebied van de Kunstmatige Intelligentie dat zich bezighoudt met het leren door computers.

Het proefschrift bestaat uit drie delen. Deel I bevat een introductie tot het MDL Principe (hoofdstuk 1–3) en een bijdrage aan de theorievorming rond MDL (hoofdstuk 4–5). Hoofdstuk 1 geeft een algemene introductie en kan gelezen worden zonder kennis van statistiek of informatietheorie. Hoofdstuk 2 en 3 geven een voortgezette introductie, met de nadruk op drie zaken: ten eerste, de tamelijk ongebruikelijke interpretatie die MDL aan *kansverdelingen* toekent: volgens MDL dient men een empirisch bepaalde kansverdeling in eerste instantie te beschouwen als een code (preciezer, code-lengte functie). Ten tweede, het begrip 'stochastische complexiteit'. Dit is de centrale notie in de theorievorming rond MDL. Ten derde, het verband tussen MDL en het *Beginsel van de Maximale Entropie*, een principe dat oorspronkelijk bedoeld was als methode voor het 'redeneren met onzekerheid'. De theorie die wordt ontwikkeld in hoofdstuk 4 en 5 geeft een eerste aanzet tot het beantwoorden van de volgende vraag: hoe kan het dat simplistische modellen voor ingewikkelde processen vaak toch bruikbaar zijn? Het volgende geldt voor vrijwel alle praktische toepassingen van de statistiek: de uitkomst van de statistische analyse van de gegevens is een model dat in feite onjuist is, vaak zelfs een grove simplificatie. Toch worden 'simplistische' modellen die op deze manier verkregen zijn met succes gebruikt voor het voorspellen en classificeren van toekomstige gegevens. De centrale vraag in deel I van dit proefschrift is: wanneer kan men een 'simplistisch' model zonder problemen gebruiken? De hoofdconclusie luidt dat een simplistisch model op twee manieren gebruikt kan worden: een 'riskante' en een 'veilige'. Als het op de veilige manier gebruikt wordt, dan zal het simplistische model in het algemeen 'betrouwbaar' zijn. Dat wil zeggen dat het model zelf een correcte indruk geeft van de voorspellingsfout die men zal maken als men het gebruikt om toekomstige data mee te voorspellen – zelfs als het model een grove simplificatie is van het proces dat daadwerkelijk aan de gegevens ten grondslag ligt. Deze 'betrouwbaarheid' van een incorrect model kan in veel gevallen zelfs formeel bewezen worden (Hoofdstuk 5, Sectie 5.3, Stellingen 5.16-5.19).

Deel II (hoofdstuk 6 en 7) gaat over praktische toepassingen van het MDL Principe. Centrale vraag is hier: werkt het MDL Principe in de praktijk beter, even goed of minder goed dan andere statistische principes? De gevonden empirische verschillen kunnen voor een groot deel uit de bestaande theorie verklaard worden. Hoofdstuk 6 vergelijkt MDL met methoden uit de Bayesiaanse en klassieke statistiek. Hoofdstuk 7 vergelijkt MDL met het nauw verwante MML (Minimum Message Length): in tegenstelling tot wat vaak gedacht wordt, zijn er kleine theoretische verschillen tussen deze twee aanpakken. Deze leiden tot verschillend gedrag in praktische leerproblemen. De hoofdconclusie van Deel II is dat geavanceerde vormen van zowel MDL als Bayesiaanse inferentie vaak verrassend goed presteren wanneer slechts zeer weinig data gegeven is. MDL lijkt beter te presteren dan MML als weinig data gegeven is, zij het dat het verschil vrijwel verwaarloosbaar is.

Deel III gaat over een onderwerp dat slechts indirect aan MDL gerelateerd is: het ontwikkelen van een theorie over *gezond-verstand redeneren* ('common-sense reasoning') over *gebeurtenissen en veranderingen*. Dit soort theorieën wordt bestudeerd in het 'logicistische' paradigma van de Kunstmatige Intelligentie. In dit paradigma probeert men op wiskundige logica gebaseerde automatische redeneersystemen te ontwikkelen. Zulke systemen kunnen vervolgens worden toegepast in, bijvoorbeeld, robots. Hoofdstuk 8 geeft een inleidend overzicht van dit soort redeneersystemen. Centrale vragen in deel III zijn: hoe kunnen de sterke eigenschappen van bestaande redeneersystemen gecombineerd worden? Wat is de rol van *causaliteit* in dit soort redeneersystemen? En, hoe is de manier waarop deze systemen met onzekerheid omgaan gerelateerd aan kansrekening en de MDL-interpretatie van kansverdelingen? Ter beantwoording van deze vragen stellen wij (in hoofdstuk 9) een nieuw redeneersysteem voor, dat gebaseerd is op het *Beginsel van de Voldoende Oorzaak* ('sufficient cause principle'). Dit is onderdeel van een theorie over causaliteit die is ontwikkeld voor statistische toepassingen, dus buiten het vakgebied van de kunstmatige intelligentie. We bewijzen formeel (in hoofdstuk 10) dat ons redeneersysteem gezien kan worden als een generalisatie van enkele bestaande redeneersystemen (met name de systemen voorgesteld door (1) McCain en Turner, (2) Lin en (3) Baral, Gelfond en Provetti). We laten zien dat deze bestaande redeneersystemen vaak impliciet gebruik maken van het sufficient cause principe. We laten ook zien dat zij problematisch gedrag kunnen vertonen zodra zij ervan afwijken. Hoofdconclusie van deel III is dat het sufficient cause principe ons toestaat om een groot deel van zowel de successen als de mislukkingen van bestaande redeneersystemen te verklaren.

Het voorgestelde redeneersysteem maakt gebruik van *niet-monotone logica*. Deel III eindigt met een Epiloog waarin een formeel verband gelegd wordt tussen aan de ene kant deze niet-monotone logica en aan de andere kant probabilistische redeneermethoden. Er wordt een probabilistische semantiek voor eenvoudige vormen van niet-monotoon redeneren gegeven. Er wordt betoogd dat de 'kansen' die in deze semantiek optreden volgens het MDL Principe geïnterpreteerd dienen te worden.

Curriculum Vitae

Peter Daniel Grünwald was born May 13th, 1970 in Geldrop, the Netherlands. In 1988 he did his final exam Gymnasium β (Grammar School) at the Lorentz Lyceum in Eindhoven. He then moved to Amsterdam, where from 1988 until 1994 he studied Computer Science at the Free University of Amsterdam. During this time, he also obtained his 'propaedeutic exam' in Mathematics (1991), he worked as a teaching assistant at the Department of Artificial Intelligence (1991–1993) and he spent six months at the IRIT (Institut de Recherche en Informatique de Toulouse) and the Université Paul Sabatier in Toulouse, France (1993-1994). In 1994 he graduated 'cum laude' with specializations in Artificial Intelligence and Theoretical Computer Science and a minor in Psychology. 1994 he started working as a Ph.D. student ('Onderzoeker in Opleiding') at the CWI in Amsterdam. During this time he spent two months with the CoSCo group at the University of Helsinki, Finland (1997). The research he did as a Ph.D. student has resulted in the present thesis. He has been awarded an NWO TALENT–grant for a post-doc position at Stanford University, Palo Alto, California, starting October 1998.

Publications

[1] P.D. Grünwald, P. Kontkanen, P. Myllymäki, T. Silander, and H. Tirri. Minimum encoding approaches for predictive modeling. In *Proceedings of the Fourteenth International Conference on Uncertainty in Artificial Intelligence (UAI '98)*, Madison, WI, 1998.

[2] P.D. Grünwald. Causation and nonmonotonic temporal reasoning. In G. Brewka, C. Habel, and B. Nebel, editors, *KI-97: Advances in Artifical Intelligence*, number 1303 in Springer Lecture Notes in Artifical Intelligence, pages 159–170, 1997.

[3] P.D. Grünwald. A minimum description length approach to grammar inference. In G. Scheler S. Wermter, E. Riloff, editor, *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*, number 1040 in in Springer Lecture Notes in Artificial Intelligence, pages 203–216, 1996.

[4] P.D. Grünwald. Model selection based on minimum description length. Accepted for publication in *Journal of Mathematical Psychology*, special issue on Model Selection. To appear in 1999.

[5] P.D. Grünwald. Ramifications and sufficient causes. In *Working notes COMMON SENSE '98 (fourth symposium on logical formalizations of commonsense reasoning)*, 1998. Available at www.dcs.qmw.ac.uk/conferences/CS98/CS98Papers.html.

[6] P. Kontkanen, P. Myllymäki, T. Silander, H. Tirri, and P.D. Grünwald. Predictive distributions and Bayesian networks. Accepted for publication in *Journal of Statistics and Computing*. To appear in 1999.

[7] P. Kontkanen, P. Myllymäki, T. Silander, H. Tirri, and P.D. Grünwald. Comparing predictive inference methods for discrete domains. In *Proceedings AISTATS-97*, pages 311–318, Ft.Lauderdale, USA, 1997.

[8] P. Kontkanen, P. Myllymäki, T. Silander, H. Tirri, and P.D. Grünwald. Bayesian and information-theoretic priors for Bayesian network parameters. In *Proceedings of the Tenth European Conference on Machine Learning (ECML-98)*, Chemnitz, Germany, 1998.

[9] M. Steijvers and P.D. Grünwald. A recurrent network that performs a context-sensitive prediction task. In *Proceedings Eighteenth Annual Conference of the Cognitive Science Society*, San Diego, CA, 1996.

Only publications that have appeared or are to appear in journals, books or refereed international conferences have been listed. The material in this thesis is partly based on some of these publications. Specifically, parts of chapters 1 and 2 are new and parts are based on [4]. The material in chapters 3, 4 and 5 is new and as yet unpublished. Chapter 6 is a combination of [6], [7] and [8]. Chapter 7 is based on [1]. Part III of the thesis is partially based on [2] and [5]. The material in publications [3] and [9] does not appear in this thesis.

Titles in the ILLC Dissertation Series:

ILLC DS-1993-01: **Paul Dekker**
Transsentential Meditations; Ups and downs in dynamic semantics

ILLC DS-1993-02: **Harry Buhrman**
Resource Bounded Reductions

ILLC DS-1993-03: **Rineke Verbrugge**
Efficient Metamathematics

ILLC DS-1993-04: **Maarten de Rijke**
Extending Modal Logic

ILLC DS-1993-05: **Herman Hendriks**
Studied Flexibility

ILLC DS-1993-06: **John Tromp**
Aspects of Algorithms and Complexity

ILLC DS-1994-01: **Harold Schellinx**
The Noble Art of Linear Decorating

ILLC DS-1994-02: **Jan Willem Cornelis Koorn**
Generating Uniform User-Interfaces for Interactive Programming Environments

ILLC DS-1994-03: **Nicoline Johanna Drost**
Process Theory and Equation Solving

ILLC DS-1994-04: **Jan Jaspars**
Calculi for Constructive Communication, a Study of the Dynamics of Partial States

ILLC DS-1994-05: **Arie van Deursen**
Executable Language Definitions, Case Studies and Origin Tracking Techniques

ILLC DS-1994-06: **Domenico Zambella**
Chapters on Bounded Arithmetic & on Provability Logic

ILLC DS-1994-07: **V. Yu. Shavrukov**
Adventures in Diagonalizable Algebras

ILLC DS-1994-08: **Makoto Kanazawa**
Learnable Classes of Categorial Grammars

ILLC DS-1994-09: **Wan Fokkink**
Clocks, Trees and Stars in Process Theory

ILLC DS-1994-10: **Zhisheng Huang**
Logics for Agents with Bounded Rationality

ILLC DS-1995-01: **Jacob Brunekreef**
On Modular Algebraic Protocol Specification

ILLC DS-1995-02: **Andreja Prijatelj**
Investigating Bounded Contraction

ILLC DS-1995-03: **Maarten Marx**
Algebraic Relativization and Arrow Logic

ILLC DS-1995-04: **Dejuan Wang**
Study on the Formal Semantics of Pictures

ILLC DS-1995-05: **Frank Tip**
Generation of Program Analysis Tools

ILLC DS-1995-06: **Jos van Wamel**
Verification Techniques for Elementary Data Types and Retransmission Protocols

ILLC DS-1995-07: **Sandro Etalle**
Transformation and Analysis of (Constraint) Logic Programs

ILLC DS-1995-08: **Natasha Kurtonina**
Frames and Labels. A Modal Analysis of Categorial Inference

ILLC DS-1995-09: **G.J. Veltink**
Tools for PSF

ILLC DS-1995-10: **Giovanna Cepparello**
Studies in Dynamic Logic

ILLC DS-1995-11: **W.P.M. Meyer Viol**
Instantial Logic. An Investigation into Reasoning with Instances

ILLC DS-1995-12: **Szabolcs Mikulás**
Taming Logics

ILLC DS-1995-13: **Marianne Kalsbeek**
Meta-Logics for Logic Programming

ILLC DS-1995-14: **Rens Bod**
Enriching Linguistics with Statistics: Performance Models of Natural Language

ILLC DS-1995-15: **Marten Trautwein**
Computational Pitfalls in Tractable Grammatical Formalisms

ILLC DS-1995-16: **Sophie Fischer**
The Solution Sets of Local Search Problems

ILLC DS-1995-17: **Michiel Leezenberg**
Contexts of Metaphor

ILLC DS-1995-18: **Willem Groeneveld**
Logical Investigations into Dynamic Semantics

ILLC DS-1995-19: **Erik Aarts**
Investigations in Logic, Language and Computation

ILLC DS-1995-20: **Natasha Alechina**
Modal Quantifiers

ILLC DS-1996-01: **Lex Hendriks**
Computations in Propositional Logic

ILLC DS-1996-02: **Angelo Montanari**
Metric and Layered Temporal Logic for Time Granularity

ILLC DS-1996-03: **Martin H. van den Berg**
Some Aspects of the Internal Structure of Discourse: the Dynamics of Nominal Anaphora

ILLC DS-1996-04: **Jeroen Bruggeman**
Formalizing Organizational Ecology

ILLC DS-1997-01: **Ronald Cramer**
Modular Design of Secure yet Practical Cryptographic Protocols

ILLC DS-1997-02: **Nataša Rakić**
Common Sense Time and Special Relativity

ILLC DS-1997-03: **Arthur Nieuwendijk**
On Logic. Inquiries into the Justification of Deduction

ILLC DS-1997-04: **Atocha Aliseda-LLera**
Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence

ILLC DS-1997-05: **Harry Stein**
The Fiber and the Fabric: An Inquiry into Wittgenstein's Views on Rule-Following and Linguistic Normativity

ILLC DS-1997-06: **Leonie Bosveld - de Smet**
On Mass and Plural Quantification. The Case of French 'des'/'du'-NP's.

ILLC DS-1998-01: **Sebastiaan A. Terwijn**
Computability and Measure

ILLC DS-1998-02: **Sjoerd D. Zwart**
Approach to the Truth: Verisimilitude and Truthlikeness

ILLC DS-1998-03: **Peter Grünwald**
The Minimum Description Length Principle and Reasoning under Uncertainty