*Curriculum Vitae*

# ***MICHAEL C. MOZER***


**Address** Department of Computer Science
University of Colorado
Boulder, CO 80309–0430

**Phone** (303) 492–4103

**E-mail** `mozer@colorado.edu`

**WWW** `http://www.cs.colorado.edu/~mozer`


**Birthdate** November 20, 1958

**Citizenship** USA

## Research Interests

- I am interested in solving *human optimization problems*, which involves the development of software tools to improve how people learn, remember, and make decisions.
- I build computer simulation models of human cognition that allow us to predict and understand behavior. I have worked in the areas of selective attention, awareness, memory, learning, executive control, decision making, and neuropsychological disorders.
- Using these models, we can determine the most effective means of teaching and the manner in which to best present information for human consumption. One project was the Colorado Optimized Language Tutor, which helps students learn facts (e.g., foreign language vocabulary) by scheduling review to promote long-term retention.
- In a complementary line of work, I use insights from cognitive science to make computer systems smarter and easier to use. A past project that got some notoriety was the adaptive house, a control system that learns to manage energy resources (air heat, water heat, lighting, and ventilation) in an actual residence to maximize the satisfaction of the inhabitants and minimize energy consumption.

## Education

Ph.D. University of California, San Diego, 1987 (Psychology and Cognitive Science)

M.A. University of California, San Diego, 1982 (Psychology)

B.A. Brown University, 1980 (Computer Science)

## Academic Awards

Educational Data Mining Society, Best Paper Award, 2016

Educational Data Mining Society. Best Paper Award, 2014

Cognitive Science Society, Computational Modeling Prize, 2013

Faculty Fellowship, University of Colorado, Boulder, 1995–1996, 2009–2010

Distinguished Cognitive Scientist Award, Glushko-Samuelson Foundation, UC Merced, 2010

Presidential Young Investigator Award, National Science Foundation, 1990

Junior Faculty Development Award, University of Colorado, Boulder, 1989

IBM Graduate Fellowship, 1985–1987

Institute for Cognitive Science SDF Graduate Fellowship, UCSD, 1981–1985

Sigma Xi (honorary scientific society), Brown University chapter, 1980

Phi Beta Kappa, Brown University chapter, 1980

B.A. degree *summa cum laude*, Brown University, 1980

## Professional Experience

Professor, Department of Computer Science and Institute of Cognitive Science, University of Colorado, Boulder, 2001–

Associate Professor, Department of Computer Science and Institute of Cognitive Science, University of Colorado, Boulder, 1992–2001

Assistant Professor, Department of Computer Science and Institute of Cognitive Science, University of Colorado, Boulder, 1988–1992

Lecturer, Department of Psychology, University of Toronto, 1987–1988

Postdoctoral Fellow, Departments of Psychology and Computer Science, University of Toronto, Dr. Geoffrey Hinton, Supervisor, 1987–1988

## Professional Activities

Faculty, International Summer School on Deep Learning, Bilbao, Spain. July 2017

Advisory Board, NSF/Cyberlearning project on “Modeling perceptual fluency with visual representations in an intelligent tutoring system for undergraduate chemistry”, University of Wisconsin, 2016-2020.

Advisory Board, NSF/Integrative Strategies project on “Using computational cognitive neuroscience to predict and optimize memory”, New York University, 2016-2019

Editorial Board, Springer series on *Applied Machine Learning*, 2016–present.

Technical Advisory Board, Open Table, 2016–present

Technical Advisory Board, Imagen Technologies, 2015–present

Editorial Board, *Neural Computation*, 2015–present

Technical Advisory Board, Cognilytics, Inc., 2011-present

Editorial Board, *Machine Learning*, 2003–present

Technical Advisory Board, AnswerOn, 2001–present

Editorial Board, *Consciousness and Cognition*, 1998–present

Board Member and Secretary, Neural Information Processing Systems Foundation, 1995–present

Co-Founder, Sensory Inc., Sunnyvale CA, 1992–present. (Sensory produces inexpensive neural network speech-recognition hardware for toys, consumer electronics, and telephony applications.)

Symposium Co-organizer, *Enhancing Education Through Cognitive Psychology*. Psychonomics 2015. Chicago, IL. November 2015.

Workshop Co-organizer, *Machine Learning For Education*. ICML 2015. Lille, France. July 2015

Workshop Co-organizer, *Human Propelled Machine Learning*. NIPS 2014. Montreal, Canada, December 2014

Workshop Co-organizer, *Approaching Twenty Years of Knowledge Tracing: Lessons Learned, Open Challenges, and Promising Developments.* EDM 2014. London, UK, July 2014.

Workshop Co-organizer, *Personalizing Education With Machine Learning*, NIPS 2012. Lake Tahoe, CA, December 2012

Technical Advisory Board, J.D. Powers and Associates, Web Intelligence Division (formerly Umbria Communications), 2003–2010

Chair, Finance Committee, Cognitive Science Society, 2005–2009

Board of Governors, Cognitive Science Society, 1998-2008

Technical Advisory Board, Green Planet Software, 2001–2008

Executive Committee, Cognitive Science Society, 2005–2008

Conference Liaison, Cognitive Science Society, 2008

Chair, Cognitive Science Society, 2006–2007.

Symposium Co-Organizer, Computational Cognitive Neuroscience Conference. *Emergent Cognitive Control*, November 2006

Advisory Board, Series on Natural Computing, Springer-Verlag, 1998-2005

Editorial Board, *Visual Cognition*, 1992–2004

Faculty, International Summer School in Cognitive Science, July 2002

Editorial Board, *Neural Networks*, 1994–2001

Symposium Co-organizer, Neural Information Processing Systems Conference. *Computational Neuropsychology,* December 2001

Chief Scientist, Athene Software, Boulder, CO, 1998–2001. (Machine learning techniques used to predict and manage subscriber churn, credit risk, and profitability in the telecommunications industry.)

Tutorials Chair, Neural Information Processing Systems Conference, December 2000

Symposium Co-organizer, Cognitive Science Conference. *Bayesian approaches to cognitive modeling*. August 2000

Editorial Board, *Cognitive Science*, 1999-2000

Advisory Board, *Connectionist Surveys*, 1996–2000

Symposium Organizer, Cognitive Neuroscience Conference. *Principles of computation in the brain*. April 1998

Workshop Co-organizer, Neural Information Processing Systems (NIPS), *Interfacing models of language*. December 1997

Co-Editor, Special issue of *Neurocomputing* on recurrent networks, 1997

Consultant, Lifestyle Technologies, Los Angeles, California, 1995–1997

General Chair, Neural Information Processing Systems Conference, December 1996

Program Chair, Neural Information Processing Systems Conference, November 1995

Faculty, James S. McDonnell Foundation Summer Institute in Cognitive Neuroscience, July 1995

Workshop Chair, Neural Information Processing Systems Conference, November 1993

Co-Organizer, Connectionist Models Summer School, June 1993

Local Arrangements Chair, Neural Information Processing Systems Conference, November 1991

Faculty, James S. McDonnell Foundation Summer Institute in Cognitive Neuroscience, July 1991

Participant, James S. McDonnell Foundation Summer Institute in Cognitive Neuroscience, June 1988

Research Assistant, Cognitive Science Laboratory, UCSD, 1981–1987

Programmer/Research Assistant, Electronic Speech Systems, Santa Clara, California, 1975–1987

Participant, Connectionist Models Summer School, June 1986

Visiting Scholar, Department of Computer Science, Carnegie-Mellon University, 1984–1985

Occasional Reviewer for *Neural Information Processing Systems Conference, Cognitive Science Society Conference, Journal of Cognitive Neuroscience, IEEE Transactions on Neural Networks*, *Neural Computation*, *Connection Science*, *Artificial Intelligence*, *Cognitive Science*, *Cognitive Psychology*, *Cognitive Neuropsychology*, *Psychological Science*, *Consciousness and Cognition, Neurocomputing*, *Neuropsychologia, Neural Networks*, *Psychological Review*, *Journal of Experimental Psychology*, *Canadian Journal of Psychology*, *Quarterly Journal of Experimental Psychology*, *Psychological Research*, NSF, AFOSR, NSERC

Program Committee Member for: Cognitive Science Society Conference, 2010; Workshop on Music and Cognition, 2007;International Conference on Unconventional Computation, 2005; Cognitive Science Conference, 2004; International Conference on Cognitive Modeling, 2004; International Conference on Neural Information Processing, 2000; International Joint Conference on Artificial Intelligence, 1997; American Association for Artificial Intelligence, 1997; International Conference on Pattern Recognition, 1994; World Conference on Neural Networks, 1993; American Association for Artificial Intelligence, 1992; Neural Information Processing Systems, 1993, 1995

## Sponsored Research

“Operationalizing students’ textbook annotations to improve comprehension and long-term retention”, NSF IIS (NCS-FO), $1,000,000 (Boulder share $300,000), 2016-2019

“Bayesian optimization for exploratory experimentation in the behavioral sciences”, NSF SES, $400k, 2015-2018

“Context-Award Music Recommendation”, Samsung, $11,400. 2014-2015

“Aphasia rehabilitation: Modulating cues, feedback, and practice conditions” (L. Cherney and S. Van Vuuren, PIs), NIH, $3.1M ($75,000 my share), 2011-2016

“Temporal dynamics of human learning and memory” (Garrison Cottrell, PI), TDLC Science of Learning Center, National Science Foundation, $15M (my share ~$500k), 2006-2016

“Improving memory retention via spacing of practice: Computational and empirical investigations” (Harold Pashler, co-Pi), National Science Foundation, BCS, $450,000 total ($224,977 my share), 2007-2010.

“Understanding the performance of modern systems” (Amer Diwan, PI), National Science Foundation, SMA, $400,000 total ($200,000 my share), 2005-2009.

“Control and adaptation of attentional processing: Empirical and computational investigations” (Shaun Vecera, Co-PI), National Science Foundation, Human and Social Dynamics Program, $430,000 total ($240,000 my share), 2004-2008.

“Enhancing learning through testing: Theoretical and practical issues” (subcontract to University of Colorado, Michael Mozer, PI; Harold Pashler, overall PI), National Institute of Health, $450,000 total (my share $119,510), 2000-2004

"Discrete representations in working memory: Developmental, neuropsychological, and computational investigations" (Randy O’Reilly, Yuko Munakata, Akira Miyake, Co-PIs), National Science Foundation, Knowledge and Distributed Intelligence Program, $800,000 total ($200,000 my share), 1998–2002

"Temporal dynamics of cognition in a modular cortical architecture", McDonnell-Pew Program in Cognitive Neuroscience, $105,000, 1997–2000

"Artificial Intelligence and Home Automation", Lifestyle Technologies, $40,000, 1997

"Rapid Behavioral Tuning to Task Demands: Computational Modeling of Empirical Data" (Clark Fagot, Co-PI), McDonnell-Pew Program in Cognitive Neuroscience, $90,000, 1994–1996

Research Experience for Undergraduates Supplement, National Science Foundation, $23,000, 1993–1995

Connectionist Models Summer School, American Association for Artificial Intelligence, National Science Foundation, and Siemens Research Center, $45,000, 1993

CRCW Grant In Aid, University of Colorado, $3,000, 1992

Digital Equipment Corporation External Research Grant, $45,750, 1991

"Connectionist Modeling and Cognitive Neuroscience", James S. McDonnell Foundation, $187,500, 1990–1995

Presidential Young Investigator Award, National Science Foundation, $312,500, 1990–1995

Junior Faculty Development Award, University of Colorado, $5,000, 1989

"Connectionist Models of Selective Attention and Object Recognition", James S. McDonnell Foundation, $9,000, 1988–1989

## U.S. Patents

"A Speech Recognition Apparatus For Consumer Electronic Applications", Forrest Mozer, Michael Mozer, and Todd Mozer. Submitted September 1994; issued August 4, 1998. US Patent 5,790,754.

"Speech Recognition in Consumer Electronic Products", Todd Mozer, Michael Mozer, and Forrest Mozer. Issued February 1, 2000. US Patent 6,021,387.

## Graduate student supervision

Adam Winchell, Ph.D. 2020 (expected)

Shirly Montero Quesada, Ph.D. 2019 (expected)

Camden Elliott-Williams, M.S. 2018 (expected).

Karl Ridgeway, Ph.D. 2018 (expected)

Brett Roads, Ph.D., 2017.

Shruthi Sukumar, M.S. 2017

Mohammad Khajah, Ph.D. 2017.

Ron Kneusel, Ph.D. 2015. *Improving hybrid human-machine search through soft highlighting.*

Robert Lindsey, Ph.D. 2014. *Probabilistic models of student learning and forgetting.*

Karl Ridgeway, M.S. 2014. *Forgetting of foreign language skills: A corpus based analysis of Rosetta Stone®*

Ahbishek Jainantilil, Ph.D. 2013. *Feature selection via iterative reweighting: An exploration of algorithms for linear models and random forests*

Brett Roads, M.S. 2013. *Using attentional highlighting to train visual expertise.*

Matthew Wilder, Ph.D. 2012. *Probabilistic modeling of sequential effects in human behavior: Theory and practical applications.*

Daniel Knights, Ph.D. 2012. *Predictive modeling of metagenomes* (co-advised with Robin Knights, received College of Engineering Outstanding Dissertation Award, #1 of 390)

Benjamin Link, M.S. 2011. *Modeling the effect of recent experience on judgments.*

Karthik Venkatesh, M.S. 2010 (Electrical and Computer Engineering). *Experience guided search: A theory of attentional control.*

Owen Lewis, M.S. 2010 (Applied Math). *A review of mathematical techniques in machine learning*.

Samuel Reid, Ph.D. 2010. *Model combination in multiclass classification*.

Adrian Fan, M.S. 2008. *A synthesis of theoretical and empirical perspectives on repetition suppression*.

Scott Richardson, M.S. 2007. *Discovering the runtime structure of software with probabilistic generative models*.

Thomas Borchert, M.S. 2007. *Computational correlates of access consciousness*.

Brian Loughery, M.S. 2003. *Learning working memory tasks by reward prediction in the basal ganglia and prefrontal cortex* (co-advisor with Randall O'Reilly)

Michael Colagrosso, Ph.D. 2003. *A rational theory of skilled performance and practice: Modeling long-term repetition priming*.

David Nix, Ph.D. 1998. *Machine learning methods for inferring vocal-tract articulation from speech acoustics*

Torleif Mohling, M.S., 1998. *Predicting human performance on anagram solving: A computational model*

Donald Mathis, Ph.D., 1997. *A computational theory of consciousness in cognition*

Srecko Vidmar, M.S., 1997. *Optimal control of home heating systems*

Kelvin Fedrick, M.S., 1996. *A decompositional approach to time series forecasting*

Debra Miller, M.S., 1995. *Adaptive lighting control*

Kevin Markey, Ph.D., 1994. *The sensorimotor foundations of phonology: A computational model of early childhood articulatory and phonetic development*

Sreerupa Das, Ph.D., 1994. *Connectionist models of language induction incorporating symbolic constraints*

John Allison, M.S., 1994. *Explorations of Bayesian input relevance determination for neural networks*

Jay Alexander, M.S., 1993. *Template-based procedures for neural network interpretation*

Ken Parker, M.S., 1993. *Selecting regression estimators for the generalized ensemble method*

Clayton McMillan, Ph.D., 1992. *Rule induction in a neural network through integrated symbolic and subsymbolic processing*

Stefanie Lindstaedt, M.S., 1992. *Comparison of unsupervised neural network models for redundancy reduction*

## Books and Edited Volumes

Mozer, M. C. (1991). *The perception of multiple objects: A connectionist approach*. Cambridge, MA: MIT Press/Bradford Books.

Mozer, M. C., Smolensky, P., Touretzky, D. S., Elman, J. L., & Weigend, A. S. (Eds.). (1994). *Proceedings of the 1993 Connectionist Models Summer School.* Hillsdale, NJ: Erlbaum.

Smolensky, P., Mozer, M. C., & Rumelhart, D. E. (Eds.). (1996). *Mathematical perspectives on neural networks*. Hillsdale, NJ: Erlbaum.

Touretzky, D. S., Mozer, M. C., & Hasselmo, M. (Eds.). (1996). *Neural Information Processing Systems 8*. Cambridge, MA: MIT Press.

Mozer, M. C., Jordan, M. I., & Petsche, T. (Eds.). (1997). *Neural Information Processing Systems 9*. Cambridge, MA: MIT Press.

### Journal Publications

Mozer, M. C. (1983). Letter migration in word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 531–546.

McClelland, J. L., & Mozer, M. C. (1986). Perceptual interactions in multi-word displays: Familiarity and similarity effects. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 18–35.

Mozer, M. C. (1989). Types and tokens in visual letter perception. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 287–303.

Mozer, M. C. (1989). A focused back-propagation algorithm for temporal sequence recognition. *Complex Systems*, *3*, 349–381.

Mozer, M. C., & Smolensky, P. (1989). Using relevance to reduce network size automatically. *Connection Science*, *1*, 3–16.

Mozer, M. C., & Behrmann, M. (1990). On the interaction of spatial attention and lexical knowledge: A connectionist account of neglect dyslexia. *Cognitive Neuroscience*, *2*, 96–123.

Behrmann, M., Moscovitch, M., Black, S. E., & Mozer, M. C. (1990). Perceptual and conceptual mechanisms in neglect dyslexia: Two contrasting case studies. *Brain*, *113*, 1163–1183.

Mozer, M. C., & Bachrach, J. (1990). Discovering the structure of a reactive environment by exploration. *Neural Computation*, *2*, 447–457.

Behrmann, M., Moscovitch, M., & Mozer, M. C. (1991). Directing attention to words and nonwords in normal subjects and in a computational model: Implications for neglect dyslexia. *Cognitive Neuropsychology*, *8*, 213–248.

Mozer, M. C., & Bachrach, J. (1991). SLUG: A connectionist architecture for inferring the structure of finite-state environments. *Machine Learning*, *7*, 139–160.

Behrmann, M., & Mozer, M. C. (1992). A connectionist account of neglect dyslexia. *Journal of Clinical and Experimental Neuropsychology*, *14*, 48–49.

Mozer, M. C., Zemel, R. S., Behrmann, M., & Williams, C. K. I. (1992). Learning to segment images using dynamic feature binding. *Neural Computation*, *4*, 650–665.

Dodier, R. H., Lukianow, D., Ries, J., & Mozer, M. C. (1994). Comparison of neural net and conventional techniques for lighting control. *Applied Mathematics and Computer Science*, *4*, 447–462.

Mozer, M. C. (1994). Neural network music composition by prediction: Exploring the benefits of psychophysical constraints and multiscale processing. *Connection Science*, *6*, 247-280.

Zemel, R. S., Williams, C. K. I., & Mozer, M. C. (1995). Lending direction to neural networks. *Neural Networks*, *8*, 503–512.

Mozer, M. C. (1996). Neural network speech processing for toys and consumer electronics. *IEEE Expert*, *11*, 4–5.

Calder, B., Grunwald, D., Jones, M., Lindsay, D., Martin, J., Mozer, M., & Zorn, B. (1997). Evidence-based static branch prediction using machine learning. *Transactions on Programming Languages and Systems*, *19*, 188-222. [Authorship order is alphabetical.]

Mozer, M. C., Halligan, P. W., Marshall, J. C. (1997). The end of the line for a brain-damaged model of unilateral neglect. *Journal of Cognitive Neuroscience*, *9*, 171–190.

Das, S., & Mozer, M. C. (1998). Dynamic on-line clustering and state extraction: An approach to symbolic learning. *Neural Networks*, *11*, 53–64.

Behrmann, M., Zemel, R. S., and Mozer, M. C. (1998). Object-based attention and occlusion: Evidence from normal subjects and a computational model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1011–1036.

Alexander, J. A., & Mozer, M. C. (1999). Template-based procedures for neural network interpretation. *Neural Networks*, *12*, 479–498.

Mozer, M. C. (1999). An intelligent environment should be adaptive. *IEEE Intelligent Systems and their Applications*, *14(2)*, 11–13.

Behrmann, M., Zemel, R. S., & Mozer, M. C. (2000). Occlusion, symmetry, and object-based attention: Reply to Saiki (1999). *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1497–1505.

Mozer, M. C., Wolniewicz, R., Grimes, D., Johnson, E., & Kaushansky, H. (2000). Maximizing revenue by predicting and addressing customer dissatisfaction. *IEEE Transactions on Neural Networks, 11*, 690–696.

Sitton, M., Mozer, M. C., & Farah, M. J. (2000). Superadditive effects of multiple lesions in a connectionist architecture: Implications for the neuropsychology of optic aphasia. *Psychological Review, 107*, 709–734.

Zemel, R. S., & Mozer, M. C. (2001). Localist attractor networks. *Neural Computation*, *13*, 1045–1064.

Mozer, M. C. (2002). Frames of reference in unilateral neglect and visual perception: A computational perspective. *Psychological Review, 109,* 156–185.

Pashler, H., Mozer, M. C., & Harris, C. R. (2002). Mating strategies in a Darwinian microworld: Simulating the consequences of female reproductive refractoriness. *Adaptive Behavior,* 9*, 5-15.*

Zemel, R. S., Behrmann, M., & Mozer, M. C. (2002). Experience-dependent perceptual grouping and object-based attention. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 202–217.

Kinoshita, S., & Mozer, M. C. (2006). How lexical decision is affected by recent experience: Symmetric versus asymmetric frequency blocking effects. *Memory and Cognition*, *34*, 726–742.

Bohte, S.M., & Mozer, M. C. (2007). Reducing the variability of neural responses: A computational theory of spike-timing dependent plasticity. *Neural Computation, 19,* 371–403.

Kinoshita, S., Forster, K. I., & Mozer, M. C. (2008). Unconscious cognition isn't that smart: Modulation of masked repetition priming effect in the word naming task. *Cognition*, *107*, 623–649.

Mozer, M. C., & Fan, A. (2008). Top-down modulation of neural responses in visual perception: A computational exploration. *Natural Computing, 7*, 45–55.

Mozer, M. C., Pashler, H., & Homaei, H. (2008). Optimal predictions in everyday cognition: The wisdom of individuals or crowds? *Cognitive Science: A Multidisciplinary Journal, 32*, 1133–1147.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology, 56*, 236-246.

Lee, H., Mozer, M. C., & Vecera, S. (2009). Mechanisms of priming of pop out: Stored representations or feature gain modulations? *Attention, Perception, & Psychophysics, 71*, 1059-71.

Kang, S. H. K., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology, 103,* 48–59.

Kinoshita, S., Mozer, M. C., & Forster, K. I. (2011). Dynamic adaptation to history of trial difficulty explains the effect of congruency proportion on masked priming. *Journal of Experimental Psychology: General, 140*, 622-636.

Knights, D., Kuczynski, J., Charlson, E., Zaneveld, J., Collman, R. G., Mozer, M. C., Bushman, F. D., Knight, R., & Kelley, S. T. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nature Methods, 8,* 761–763.

Wilder, M. H., Mozer, M. C., & Wickens, C. D. (2011). An integrative experience-based theory of attentional control. *Journal of Vision, 11,* 1–30.

Doshi, A., Tran, C., Wilder, M., Mozer, M. C., & Trivedi, M. (2012). Sequential effects in driving. *Cognitive Science, 36,* 948–963.

Lee, H., Mozer, M. C., Kramer, A. F., & Vecera, S. P. (2012). Object-based control of attention is sensitive to recent experience. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 314–325.

Chukoskie, L., Snider, J., Mozer, M. C., Krauzlis, R. J., & Sejnowski, T. J. (2013). Learning where to look: An empirical, computational, and theoretical account of hidden target search performance. *Proceedings of the National Acadamy of Sciences, 110,* 10438-10445.

Jones, M., Curran, T., Mozer, M. C., & Wilder, M. H. (2013). Sequential effects in response time reveal learning mechanisms and event representations. *Psychological Review*, *120*, 628-666.

Pashler, H. Kang, S., & Mozer, M. C. (2013). Reviewing erroneous information facilitates memory updating. *Cognition*, *128*(3), 424–430.

Pashler, H., & Mozer, M. C. (2013). When does fading help perceptual category learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1162-73.

Wilder, M. H., Jones, M., Ahmed A. A., Curran, T., & Mozer, M. C. (2013). The persistent impact of incidental experience. *Psychonomic Bulletin and Review*, *20*, 1221–1231.

Khajah, M. M., Lindsey, R. V., & Mozer, M. C. (2014). Maximizing students’ retention via spaced review: Practical guidance from computational models of memory. *Topics in Cognitive Science*, *6*, 157–169.

Lindsey, R., Shroyer, J., Pashler, H., & Mozer, M. C. (2014). Improving long-term knowledge retention through personalized review. *Psychological Science, 25, 639–647. doi: 10.1177/0956797613504302.*

Kang, S. H. K., Lindsey, R. V., Mozer, M. C., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review*, *21*, 1544-50.

Roads, B. D., Mozer, M. C., & Busey, T. A. (2016). Using highlighting to train attentional expertise. *PLoS ONE* 11(1): e0146266. doi:10.1371/journal.pone.0146266

Kneusel, R. T., & Mozer, M. C. (2017). Improving human-machine cooperative visual search with soft highlighting. *ACM Transactions on Applied Perception*. Accepted for publication. Also arXiv:1612.08117 [cs.HC]

Ridgeway, K., Mozer, M. C., & Bowles, A. (2017). Forgetting of foreign language skills: A corpus-based analysis of online tutoring software. *Cognitive Science: A Multidisciplinary Journal, 41,* 924-949. *doi:10.1111/cogs.12385.*

Roads, B. D., & Mozer, M. C. (2017). Improving human-computer cooperative classification via cognitive theories of similarity. *Cognitive Science: A Multidisciplinary Journal*. doi:10.1111/cogs.12400

Mozer, M. C., Pashler, H., Lindsey, R. V., & Jones, J. (Submitted). Efficient training of visual search via attentional highlighting. Submitted for publication.

Mozer, M. C., Sukumar, S., Elliott-Williams, C., Hakimi, S., & Ward, A. F. (Submitted). Overcoming temptation: Incentive design for intertemporal choice. Submitted for publication.

Vatterott, D. B., Mozer, M. C., & Vecera, S. P. (Submitted) Rejecting salient distractors: Generalization from experience. Submitted for publication.

## Refereed Conference Proceedings

Mozer, M. C. (1987). Early parallel processes in reading: A connectionist approach. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 83–104). Hillsdale, NJ: Erlbaum.

Mozer, M. C. (1987). RAMBOT: A connectionist expert system that learns by example. In M. Caudillo & C. Butler (Eds.), *Proceedings of the IEEE First Annual International Conference on Neural Networks* (pp. 693–700). San Diego: IEEE Publishing Services.

Mozer, M. C. (1988). A connectionist model of selective attention in visual perception. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 195–201). Hillsdale, NJ: Erlbaum.

Mozer, M. C., & Smolensky, P. (1989). Skeletonization: A technique for trimming the fat from a network via relevance assessment. In D. Touretzky (Ed.), *Advances in Neural Information Processing Systems I* (pp. 107–115). San Mateo, CA: Morgan Kaufmann.

Mozer, M. C. (1990). Discovering faithful “Wickelfeature” representations in a connectionist network. *Proceedings of the Twelfth Annual Conference of the Cognitive Society* (pp. 356–363). Hillsdale, NJ: Erlbaum.

Mozer, M. C., & Bachrach, J. (1990). Discovering the structure of a reactive environment by exploration. In D. Touretzky (Ed.), *Advances in neural information processing systems II* (pp. 439–446). San Mateo, CA: Morgan Kaufmann.

Zemel, R. S., Mozer, M. C., & Hinton G. E. (1990). TRAFFIC: Object recognition using hierarchical reference frame transformations. In D. Touretzky (Ed.), *Advances in neural information processing systems II* (pp. 266–273). San Mateo, CA: Morgan Kaufmann.

Mozer, M. C., & Soukup, T. (1991). Algorithmic music composition with melodic and stylistic constraints. In R. P. Lippmann, J. Moody, and D. S. Touretzky (Eds.), *Advances in neural information processing systems III* (pp. 789–796). San Mateo, CA: Morgan Kaufmann.

Mozer, M. C. (1991). Discovering discrete distributed representations with iterative competitive learning. In R. P. Lippmann, J. Moody, and D. S. Touretzky (Eds.), *Advances in neural information processing systems III* (pp. 627–634). San Mateo, CA: Morgan Kaufmann.

McMillan, C., Mozer, M. C., & Smolensky, P. (1991). Learning explicit rules in a neural network. In *Proceedings of the International Joint Conference on Neural Networks, Volume II* (pp. 83–88). Piscataway, NJ: IEEE Publishing Services.

McMillan, C., Mozer, M. C., & Smolensky, P. (1991). The connectionist scientist game: Rule extraction and refinement in a neural network. *Proceedings of the Thirteenth Annual Conference of the Cognitive Society* (pp. 424–430). Hillsdale, NJ: Erlbaum.

Mozer, M. C. (1992). The induction of multiscale temporal structure. In J. E. Moody, S. J. Hanson, & R. P. Lippmann (Eds.), *Advances in neural information processing systems IV* (pp. 275–282). San Mateo, CA: Morgan Kaufmann.

Mozer, M. C., Zemel, R. S., & Behrmann, M. (1992). Learning to segment images using dynamic feature binding. In J. E. Moody, S. J. Hanson, & R. P. Lippmann (Eds.), *Advances in neural information processing systems IV* (pp. 436–443). San Mateo, CA: Morgan Kaufmann.

McMillan, C., Mozer, M. C., & Smolensky, P. (1992). Rule induction through integrated symbolic and subsymbolic processing. In J. E. Moody, S. J. Hanson, & R. P. Lippmann (Eds.), *Advances in neural information processing systems IV* (pp. 969–976). San Mateo, CA: Morgan Kaufmann.

Markey, K., & Mozer, M. C. (1992). Comparison of reinforcement algorithms on learning discrete functions: Learnability, time complexity, and scaling. *Proceedings of the International Joint Conference on Neural Networks* (Volume I, pp. 853-859). San Diego, CA: IEEE Publishing Services.

Mozer, M. C., Zemel, R. S., & Behrmann, M. (1992). Discovering and using perceptual grouping principles in visual information processing. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 283–288). Hillsdale, NJ: Erlbaum.

Wagner, K., Mozer, M., Smolensky, P., Miyata, Y., & Fellows, M. (1992). Optical neural networks using a new radial nonlinear neural layer. *Proceedings of the SPIE*, *1773A-10*.

Bonnlander, B. V., & Mozer, M. C. (1993). Metamorphosis networks: An alternative to constructive methods. S. J. Hanson, J. D. Cowan, & C. L. Giles (Eds.), *Advances in Neural Information Processing Systems V* (pp. 131–138). San Mateo, CA: Morgan Kaufmann Publishers.

Das, S., & Mozer, M. C. (1993). A connectionist symbol manipulator that induces rewrite rules in context-free grammars. *Proceedings of the first colloquium on grammatical inference: Theory, applications, and alternatives*. London: Institute of Electrical Engineers Digest 1993/092.

Gupta, P., & Mozer, M. C. (1993). Exploring the nature and development of phonological representations. *Proceedings of the Fifteenth Annual Conference of the Cognitive Society* (pp. 516–521). Hillsdale, NJ: Erlbaum.

McMillan, C., Mozer, M. C., & Smolensky, P. (1993). Implicit conflict resolution in a connectionist rule-based system. *Proceedings of the 13th Annual International Joint Conference on Artificial Intelligence* (pp. 1366–1371). San Mateo, CA: Morgan Kaufmann.

Mozer, M. C., & Das, S. (1993). A connectionist symbol manipulator that discovers the structure of context-free languages. In S. J. Hanson, J. D. Cowan, & C. L. Giles (Eds.), *Advances in Neural Information Processing Systems V* (pp. 863–870). San Mateo, CA: Morgan Kaufmann Publishers.

Schmidhuber, J. H., Mozer, M. C., & Prelinger, D. (1993). Continuous history compression. In H. Huening, S. Neuhauser, M. Raus, & W. Ritschel (Eds.), *Proceedings of the International Workshop on Neural Networks, RWTH Aachen* (pp. 87–95). Augustinus.

Williams, C. K. I., Zemel, R. S., & Mozer, M. C. (1993). Unsupervised learning of object models. *Proceedings of the AAAI Fall 1993 Symposium on Machine Learning in Computer Vision* (pp. 20–24). Proceedings available as AAAI Tech Report FSS-93-04.

Zemel, R. S., Williams, C. K. I., & Mozer, M. C. (1993). Directional-unit Boltzmann machines. In S. J. Hanson, J. D. Cowan, & C. L. Giles (Eds.), *Advances in Neural Information Processing Systems V* (pp. 172–179). San Mateo, CA: Morgan Kaufmann Publishers.

Das, S., & Mozer, M. C. (1994). A unified gradient-descent/clustering architecture for finite-state machine induction. In J. D. Cowan, G. Tesauro, & J. Alspector (Eds.), *Advances in Neural Information Processing Systems VI* (pp. 19–26). San Mateo, CA: Morgan Kaufmann Publishers.

Mozer, M. C. (1994). Computational approaches to functional feature learning [Extended abstract]. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 975-976). Hillsdale, NJ: Erlbaum.

Alexander, J. A., & Mozer, M. C. (1995). Template-based algorithms for connectionist rule extraction. In G. Tesauro, D. S. Touretzky, & T. K. Leen (Eds.), *Advances in Neural Information Processing Systems 7* (pp. 609–616). Cambridge, MA: MIT Press.

Calder, B., Grunwald, D., Lindsay, D., Martin, J., Mozer, M., Zorn, B. (1995). Corpus-based static branch prediction. *ACM SIGPLAN '95 Conference on Programming Language Design and Implementation* (pp. 79–92). La Jolla, CA. June, 1995. [alphabetical authorship order]

Mathis, D. A., & Mozer, M. C. (1995). On the computational utility of consciousness. In G. Tesauro, D. S. Touretzky, & T. K. Leen (Eds.), *Advances in Neural Information Processing Systems 7* (pp. 10–18). Cambridge, MA: MIT Press.

Markey, K. L., Menn, L., and Mozer, M. C. (1995). A developmental model of the sensorimotor foundations of child phonology. In D. MacLaughlin & S. McEwen (Eds.), *Proceedings of the Nineteenth Boston University Conference on Language Development*, *19*, 367-378. Sommerville, MA: Cascadilla Press.

Mathis, D., & Mozer, M. C. (1996). Conscious and unconscious perception: A computational theory. In G. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 324–328). Hillsdale, NJ: Erlbaum.

Mozer, M. C., Vidmar, L., & Dodier, R. H. (1997). The Neurothermostat: Adaptive control of residential heating systems. In M. C. Mozer, M. I. Jordan, & T. Petsche (Eds.), *Advances in Neural Information Processing Systems 9* (pp. 953–959). Cambridge, MA: MIT Press.

Uno, Y., & Mozer, M. C. (1997). Neural net architectures in modeling compositional syntax: Prediction and perception of continuity in minimalist works by Phillip Glass and Louis Andriessen. Proceedings of the International Computer Music Conference, Greece.

Mozer, M. C., Sitton, M., & Farah, M. J. (1998). A superadditive-impairment theory of optic aphasia. In M. I. Jordan, M. Kearns, & S. A. Solla (Eds.), *Advances in Neural Information Processing Systems 10* (pp. 66–72). Cambridge, MA: MIT Press.

Mozer, M. C. (1998). The neural network house: An environment that adapts to its inhabitants. In M. Coen (Ed.), *Proceedings of the American Association for Artificial Intelligence Spring Symposium on Intelligent Environments* (pp. 110–114; Technical Report SS-98-02). Menlo Park, CA: AAAI Press.

Mozer, M. C. (1999). A principle for the unsupervised hierarchical decomposition of visual scenes. In M. S. Kearns, S. A. Solla, & D. Cohn (Eds.), *Advances in Neural Information Processing Systems 11* (pp. 52-58). Cambridge, MA: MIT Press.

Mozer, M. C. (1999). Do attention and perception require multiple reference frames? Evidence from a computational model of unilateral neglect. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society* (pp. 456-461). Hillsdale, NJ: Lawrence Erlbaum Associates.

O'Reilly, R. C., Mozer, M. C., Munakata, Y., & Miyake, A. (1999). Discrete representations in working memory: A hypothesis and computational investigations. In *Proceedings of the Second International Conference on Cognitive Science* (pp. 183-188). Tokyo, Japan: Japanese Cognitive Science Society.

Lee, S.-Y., & Mozer, M. C. (2000). Robust recognition of noisy and superimposed patterns via selective attention. In S. A. Solla, T. K. Leen & K.-R. Mueller (Eds.), *Advances in Neural Information Processing Systems 12* (pp. 31–37). Cambridge, MA: MIT Press.

Mozer, M. C., Wolniewicz, R., Grimes, D. B., Johnson, E., & Kaushansky, H. (2000). Churn reduction in the wireless industry. In S. A. Solla, T. K. Leen & K.-R. Mueller (Eds.), *Advances in Neural Information Processing Systems 12* (pp. 935–941). Cambridge, MA: MIT Press.

Hochreiter, S., & Mozer, M. C. (2000). An electric field approach to independent component analysis. In P. Pajunen & J. Karhunen (Eds.), *Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation* (pp. 145–150). Espoo, Finland: Otamedia.

Wolniewicz, R., Mozer, M. C., Dodier, R. H., Yan, L., and Guerra, C. (2000). Enhancing customer relationship management solutions through choice of data representation. In *Proceedings of CRM Infrastructure*. IBC Global Conferences, September 20-21, 2000. London, England. Available at www.ibctelecoms.com/crminfra.s

Zemel, R. S., & Mozer, M. C. (2000). A generative model for attractor dynamics. In S. A. Solla, T. K. Leen & K.-R. Mueller (Eds.), *Advances in Neural Information Processing Systems 12* (pp. 80–86). Cambridge, MA: MIT Press.

Grimes, D., & Mozer, M. C. (2001). The interplay of symbolic and subsymbolic processes in anagram problem solving. In T. K. Leen, T. Dietterich, & V. Tresp (Eds.), *Advances in Neural Information Processing Systems 13* (pp. 17–23). Cambridge, MA: MIT Press.

Hochreiter, S., & Mozer, M. C. (2001). Beyond maximum likelihood and density estimation: A sample-based criterion for unsupervised learning of complex models. In T. K. Leen, T. Dietterich, & V. Tresp, (Eds.), *Advances in Neural Information Processing Systems 13* (pp. 535–541). Cambridge, MA: MIT Press.

Hochreiter, S., & Mozer, M. C. (2001). A discrete probabilistic memory model for discovering dependencies in time. In G. Dorffner, H. Bischof, & K. Hornig (Eds.), *Proceedings of the International Conference on Artificial Neural Networks (ICANN) 2001* (pp. 661-668). Berlin: Springer-Verlag.

Yan, L., Miller, D., Mozer, M. C., & Wolniewicz, R. (2001). Improving prediction of customer behavior in nonstationary environments. *Proceedings of the International Joint Conference on Neural Networks* (pp. 2258–2263). Piscataway, NJ: IEEE Press.

Hochreiter, S., & Mozer, M. C. (2001). Monaural separation and classification of mixed signals: A support-vector regression perspective. *Proceedings of the Third International Conference on Independent Component Analysis and Blind Signal Separation*, San Diego, CA.

Mozer, M. C., Colagrosso, M. D., & Huber, D. H. (2002). A rational analysis of cognitive control in a speeded discrimination task. In T. Dietterich, S. Becker, & Ghahramani, Z. (Eds.) *Advances in Neural Information Processing Systems 14* (pp. 51–57). Cambridge, MA: MIT Press.

Mozer, M. C., Dodier, R., Colagrosso, M. D., Guerra-Salcedo, C., & Wolniewicz, R. (2002). Prodding the ROC curve: Constrained optimization of classifier performance. In T. Dietterich, S. Becker, & Ghahramani, Z. (Eds.) *Advances in Neural Information Processing Systems 14* (pp. 1409–1415). Cambridge, MA: MIT Press.

Hochreiter, S., Mozer, M. C., & Obermayer, K. (2003). Coulomb classifiers: Generalizing support vector machines via an analogy to electrostatic systems. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in Neural Information Processing Systems 15* (pp. 545-552). Cambridge, MA: MIT Press.

Mozer, M. C., Colagrosso, M. D., & Huber, D. E. (2003). Mechanisms of long-term repetition priming and skill refinement: A probabilistic pathway model. In *Proceedings of the Twenty Fifth Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum Associates.

Yan, L., Dodier, R., Mozer, M. C., & Wolniewicz, R. (2003). Optimizing classifier performance via the Wilcoxon-Mann-Whitney statistic. In T. Fawcett & N. Mishra (Eds.), *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)* (pp. 848-855). Menlo Park, CA: AAAI Press.

Mozer, M. C., Howe, M., & Pashler, H. (2004). Using testing to enhance learning: A comparison of two hypotheses. In K. Forbus, D. Gentner, T. Regier (Eds.), *Proceedings of the Twenty Sixth Conference of the Cognitive Science Society* (pp. 975-980). Hillsdale, NJ: Erlbaum Associates.

Mozer, M. C., Kinoshita, S., & Davis, C. (2004). Control of response initiation: Mechanisms of adaptation to recent experience. In K. Forbus, D. Gentner, T. Regier (Eds.), *Proceedings of the Twenty Sixth Conference of the Cognitive Science Society* (pp. 981–986). Hillsdale, NJ: Erlbaum Associates.

Bohte, S., & Mozer, M. C. (2005). Reducing spike train variability: A computational theory of spike-timing dependent plasticity. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems 17* (pp. 201-208). Cambridge, MA: MIT Press.

Colagrosso, M. D., & Mozer, M. C. (2005). Theories of access consciousness. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems 17* (pp. 289-296). Cambridge, MA: MIT Press.

Hauswirth, M., Diwan, A., Sweeney, P. F., & Mozer, M. C. (2005). Automated vertical profiling. In *Proceedings of Object Oriented Programming, Languages, and Applications 2005 (OOPSLA'05)*.

Mozer, M. C., Shettel, M., & Vecera, S. P. (2006). Control of visual attention: A rational account. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Neural Information Processing Systems 18* (pp. 923-930). Cambridge, MA: MIT Press.

Baldwin, D., & Mozer, M. C. (2006). Controlling attention with noise: The cue-combination model of visual search. In R. Sun & N. Miyake (Eds.), *Proceedings of the Twenty Eighth Conference of the Cognitive Science Society* (pp. 42–47). Hillsdale, NJ: Erlbaum Associates.

Mozer, M. C., & Fan, A. (2007). Top-down modulation of neural responses in visual perception: A computational exploration. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 491-496). Austin, TX: Cognitive Science Society.

Mozer, M. C., Jones, M., & Shettel, M. (2007). Context effects in category learning: An investigation of four probabilistic models. In B. Schölkopf, J. Platt, & T. Hoffmann (Eds.), *Neural Information Processing Systems 19* (pp. 993-1000). Cambridge, MA: MIT Press.

Mozer, M. C., & Baldwin, D. (2008). Experience-guided search: A theory of attentional control. In J. Platt, D. Koller, Y. Singer, & Roweis, S. (Eds.), *Advances in Neural Information Processing Systems 20* (pp. 1033-1040). Cambridge, MA: MIT Press.

Mozer, M. C., Pashler, H., & Homaei, H. (2008). Optimal predictions in everyday cognition: The wisdom of individuals or crowds? In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1051-1056). Austin, TX: Cognitive Science Society.

Jones, M., Mozer, M. C., & Kinoshita, S. (2009). Optimal response initiation: Why recent experience matters. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems 21* (pp. 785-792). La Jolla, CA: NIPS Foundation.

Reynolds, J., & Mozer, M. C. (2009). Temporal dynamics of cognitive control. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems 21* (pp. 1353-1360). La Jolla, CA: NIPS Foundation.

Lindsey, R., Mozer, M. C., Cepeda, N. J., & Pashler, H. (2009). Optimizing memory retention with cognitive models. In A. Howes, D. Peebles, & R. Cooper (Eds.), *Proceedings of the Ninth International Conference on Cognitive Modeling (ICCM)*. Manchester, UK.

Knights, D., Mytkowicz, T., Sweeney, P. F., Mozer, M. C., & Diwan, A. (2009). Blind optimization for exploiting hardware features. In O. de Moor & M. I. Schwartzbach (Eds.), *Lecture Notes in Computer Science, v. 5501: Compiler Construction 2009* (pp. 251-265). New York: Springer

Knights, D., Mozer, M. C., Nicolov, N. (2009). Detecting topic drift with compound topic models. In Proceedings of the Third International AAAI Conference on Weblogs and Social Media. AAAI Press.

Mozer, M. C., Pashler, H., Cepeda, N., Lindsey, R., & Vul, E. (2009). Predicting the optimal spacing of study: A multiscale context model of memory. In Y. Bengio, D. Schuurmans, J. Lafferty, C.K.I. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22* (pp. 1321–1329). La Jolla, CA: NIPS Foundation.

Wilder, M., Jones, M., & Mozer, M. C. (2009). Sequential effects reflect parallel learning of multiple environmental regularities. In Y. Bengio, D. Schuurmans, J. Lafferty, C.K.I. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22* (pp. 2053-2061). La Jolla, CA: NIPS Foundation.

Lindsey, R., Lewis, O., Pashler, H., & Mozer, M. C. (2010). Predicting students' retention of facts from feedback during training. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2332-2337). Austin, TX: Cognitive Science Society.

Mozer, M. C., Pashler, H., Wilder, M., Lindsey, R., Jones, M., & Jones, M. (2010). Decontaminating human judgments by removing sequential dependencies. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, & A. Culota (Eds.), *Advances in Neural Information Processing Systems 23* (pp. 1705-1713)*.* La Jolla, CA: NIPS Foundation.

Link, B. V., Kos, B., Wager, T. D., & Mozer M. C. (2011). Past experience influences judgment of pain: Prediction of sequential dependencies. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33d Annual Conference of the Cognitive Science Society* (pp. 1248–1253). Austin, TX: Cognitive Science Society.

Mozer, M. C., Link, B. V., & Pashler, H. (2011) An unsupervised decontamination procedure for improving the reliability of human judgments. In Shawe-Taylor, J., Zemel, R. S., Bartlett, P., Pereira, & Weinberger, K. Q. (Eds.), *Advances in Neural Information Processing Systems 24* (pp. 1791-1799)*.* La Jolla, CA: NIPS Foundation.

Khajah, M., Lindsey, R., & Mozer, M. C. (2013). Maximizing students' retention via spaced review: Practical guidance from computational models of memory. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 758-763). Austin, TX: Cognitive Science Society. [Awarded the Cognitive Science Society Computational Modeling Prize]

Lindsey, R. V., Mozer, M. C., Huggins, W. J., & Pashler, H. (2013). Optimizing instructional policies.In C. J. C. Burges et al. (Eds.), *Advances in Neural Information Processing Systems 26* (pp. 2778-2786). Curran Associates, Inc. [Among the 20 of 1420 submissions accepted for oral presentation]

Khajah, M., Huang, Y., Gonzales-Brenes, J. P., Mozer, M. C., & Brusilovsky, P. (2014). Integrating knowledge tracing and item response theory: A tale of two frameworks. In M. Kravcik, O. C. Santos, J. G. Boticario (Eds.), *Proceedings of the 4th International Workshop on Personalization Approaches in Learning Environments* (pp. 7-15). CEUR Workshop Proceedings, ISSN 1613-0073.

Khajah, M., Wing, R. M., Lindsey, R. V., & Mozer, M. C. (2014) Incorporating latent factors into knowledge tracing to predict individual differences in learning. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the Seventh International Conference on Educational Data Mining* (pp. 99-106). Educational Data Mining Society Press. [Awarded Best Paper of Conference]

Lindsey, R. V., Khajah, M., & Mozer, M. C. (2014). Automatic discovery of cognitive skills to improve the prediction of student learning. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberge (Eds.), *Advances in Neural Information Processing Systems 27* (pp. 1386-1394)*.* La Jolla, CA: Curran Associates Inc. [Acceptance rate 24.7%]

Beckage, N., Mozer, M. C., & Colunga, E. (2015). Predicting a child's trajectory of lexical acquisition. In D. Noelle et al. (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 196-201). Austin, TX: Cognitive Science Society.

Khajah, M., Lindsey, R. V., & Mozer, M. C. (2016). How deep is knowledge tracing? In T. Barnes, M. Chi, & M. Feng (Eds.), *Proceedings of the Ninth International Conference on Educational Data Mining* (pp. 94-101). Educational Data Mining Society. [Awarded Best Paper of Conference]

Khajah, M., Roads, B. D., Lindsey, R. V., Liu, Y.-E., & Mozer, M. C. (2016). Designing engaging games using Bayesian optimization. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5571-5582). New York: ACM. [Acceptance rate 23.4%]

Snell, J., Ridgeway, K. Liao, R., Roads, B. D., Mozer, M. C., & Zemel, R. S. (2017). Learning to generate images with perceptual similarity metrics. *IEEE International Conference on Image Processing*. Beijing, China: IEEE Press.

Mozer, M. C., Lindsey, R. V., & Kazakov, D. (2017). Continuous-time, discrete-event memory nets. Submitted for publication.

## Refereed Book Chapters

Zemel, R. S., Mozer, M. C., & Hinton, G. E. (1989). TRAFFIC: A model of object recognition based on transformations of feature instances. In D. S. Touretzky, G. E. Hinton, and T. J. Sejnowski (Eds.), *Proceedings of the 1988 Connectionist Models Summer School* (pp. 452–461). San Mateo, CA: Morgan Kaufmann.

Menn, L., Markey, K., Mozer, M. C., & Lewis, C. (1993). Connectionist modeling and the microstructure of phonological development: A progress report. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, and J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 421-433). Boston: Kluwer Academic Publishers.

Mozer, M. C. (1993). Neural network architectures for temporal pattern processing. In A. S. Weigend & N. A. Gershenfeld (Eds.), *Time series prediction: Forecasting the future and understanding the past* (pp. 243–264). Redwood City, CA: Santa Fe Institute Studies in the Sciences of Complexity, Proceedings Volume XVII, Addison-Wesley Publishing.

Mozer, M. C. (2004). Lessons from an adaptive house. In D. Cook & S. Das (Eds.), *Smart environments: Technologies, protocols, and applications* (pp. 273–294). J. Wiley & Sons.

Mozer, M. C., Kinoshita, S., & Shettel, M. (2007). Sequential dependencies in human behavior offer insights into cognitive control. In W. Gray (Ed.), *Integrated Models of Cognitive Systems* (pp. 180–193). Oxford University Press.

Mozer, M. C. (2009). Attractor networks. In T. Bayne, A. Cleeremans, & P. Wilken (Eds.), *Oxford Companion to Consciousness* (pp. 86-89). Oxford, UK: Oxford U. Press.

Mozer, M. C., & Lindsey, R. V. (2017). Predicting and improving memory retention: Psychological theory matters in the big data era. In M. Jones (Ed.), *Big Data in Cognitive Science* (pp. 34-64). New York: Routledge.

## Other Book Chapters

Anderson, J. A., & Mozer, M. C. (1981). Categorization and selective neurons. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 213–236). Hillsdale, NJ: Erlbaum Associates.

Mozer, M. C., & Gross, H. (1986). An architecture for experiential learning. In T. M. Mitchell, J. G. Carbonell, & R. S. Michalski (Eds.), *Machine learning: A guide to current research* (pp. 219–226). Boston: Kluwer Academic Publishers.

Mozer, M. C. (1991). Connectionist music composition based on melodic, stylistic, and psychophysical constraints. In P. M. Todd & D. G. Loy (Eds.), *Music and connectionism* (pp. 195–211). Cambridge, MA: MIT Press/Bradford Books.

Mozer, M. C., & Behrmann, M. (1992). Reading with attentional impairments: A brain-damaged model of neglect and attentional dyslexias. In R. G. Reilly & N. E. Sharkey (Eds.), *Connectionist approaches to natural language processing* (pp. 409–460). Hillsdale, NJ: Erlbaum Associates.

Bachrach, J., & Mozer, M. C. (1995). Modeling and controlling finite-state systems. In Y. Chauvin & D. E. Rumelhart (Eds.), *Back propagation: Theory, architectures, and applications* (pp. 351-388). Hillsdale, NJ: Erlbaum.

Mozer, M. C. (1995). A focused back-propagation algorithm for temporal sequence recognition. In Y. Chauvin & D. E. Rumelhart (Eds.), *Back propagation: Theory, architectures, and applications* (pp. 137-170). Hillsdale, NJ: Erlbaum.

Mozer, M. C., Dodier, R. H., Anderson, M., Vidmar, L., Cruickshank III, R. F., Miller, D. (1995). The neural network house: An overview. In L. Niklasson & M. Boden (Eds.), *Current trends in connectionism* (pp. 371-380). Hillsdale, NJ: Erlbaum.

Mozer, M. C., & Sitton, M. (1998). Computational modeling of spatial attention. In H. Pashler (Ed.), *Attention* (pp. 341–393). London: UCL Press.

Mozer, M. C., & Miller, D. (1998). Parsing the stream of time: The value of event-based segmentation in a complex, real-world control problem. In C. L. Giles & M. Gori (Eds.), *Adaptive processing of temporal sequences and data structures* (pp. 370–388). Berlin: Springer Verlag.

Mozer, M. C. (1999). Explaining object-based deficits in unilateral neglect without object-based frames of reference. In J. Reggia, E. Ruppin, & D. Glanzman (Eds.), *Disorders of brain, behavior, and cognition: The neurocomputational perspective* (pp. 99–119). New York: Elsevier.

Tsoi, A. C., Back, A., Principe, J., & Mozer, M. C. (2001). Memory kernels. In J. F. Kolen & S. C. Kremer (Eds.), *A field guide to dynamical recurrent networks* (pp. 39–54). Piscataway, NJ: IEEE Press.

Mozer, M. C. (2001). Object recognition. In N. J. Smelser & P. B. Baltes (Eds.), *Encyclopedia of the Social and Behavioral Sciences, Vol. 16* (pp. 10781–10784). Oxford, UK: Elsevier Science.

Mozer, M. C., & Vecera, S. P. (2005). Space- and object-based attention. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of attention* (pp. 130–134). Academic Press / Elsevier.

Mozer, M. C. (2006). Rational models of cognitive control. In C. Calude et al. (Eds.), *Unconventional Computation: Fifth International Conference* (pp. 20–25). Berlin: Springer-Verlag

Hochreiter, S., & Mozer, M. C. (2007). Monaural speech separation by support vector machines: Bridging the divide between supervised and unsupervised learning methods. In S. Makino, T.-W. Lee, & H. Sawada (Eds.), *Blind signal separation* (pp. 405-421). Berlin: Springer-Verlag.

Mozer, M. C., & Wilder, M. H. (2009). A unified theory of exogenous and endogenous attentional control. In D. Heinke & E. Mavritsaki (Eds.), *Computational modeling in behavioral neuroscience: Closing the gap between neurophysiology and behaviour* (pp. 245-265). London: Psychology Press.

Knights, D., Mozer, M. C., & Nicolov, N. (2009). Detecting topic drift. *Recent Advances in Natural Language Processing V (a.k.a. Current Issues in Linguistic Theory)* (pp. 113-130). Amsterdam & Philadelphia: John Benjamins.

## Technical Reports

Mozer, M. C. (1983). *Principal component analysis using parallel computation* (ICS Progress Report, First Quarter). La Jolla: University of California, San Diego, Institute for Cognitive Science.

Mozer, M. C. (1984). *Inductive information retrieval using parallel distributed computation* (ICS Technical Report 8406). La Jolla: University of California, San Diego, Institute for Cognitive Science.

Mozer, M. C., Pashler, H., & Miyata, Y. (1991). *Recovering the “where” from the “what”: A connectionist mechanism to direct attention to objects*. Unpublished manuscript.

Rickard, T. C., Mozer, M. C., & Bourne, L. E. Jr. (1993). An interactive activation model of arithmetic fact retrieval. Technical Report 92-15. Boulder, Colorado: Institute of Cognitive Science, University of Colorado.

Mozer, M. C., Zemel, R. S., & Hungerford, M. (2003). Optimal adaptation of neural codes. An account of reptition suppression. Unpublished Manuscript.

Mozer, M. C., Mytkowicz, T., & Zemel, R. S. (2004). Achieving robust neural representations: An account of repetition suppression.

Richardson, S., Otte, M., Mozer, M. C., Diwan, A., Sweeney, P., & Connors, D. (2009). Discovering the runtime structure of software with probabilistic generative models.

Reid, S., & Mozer, M. C. (2011). Probabilistic pairwise classification.

Lindsey, R. V., & Mozer, M. C. (2013). Predicting individual differences in student learning via collaborative filtering.

Lindsey, R., Polsdofer, E., Mozer, M.C., Kang, S., H., K., & Pashler, H. (2013). Long-term recency is nothing more than ordinary forgetting.

## Selected Invited Presentations

Invited Participant, Attention and Performance XII, England. July, 1986.

Colloquium, Department of Psychology, University of Guelph. Guelph, Ontario. March 1988.

Tutorial Speaker, Rocky Mountain Artificial Intelligence Conference, Denver, CO. June 1989.

Invited Participant, Symposium on Cognitive Neuroscience. The Eleventh Annual Conference of the Cognitive Science Society. Ann Arbor, MI. August 1989.

Colloquium, Department of Psychology, University of Michigan. Ann Arbor, MI. January 1991.

Colloquium, Department of Music, University of Washington. Seattle, WA. February 1991.

Colloquium, Department of Computer Science, University of Massachusetts at Amherst. May 1991.

Colloquium, International Computer Science Institute. Berkeley, CA. June 1991.

Faculty Lecturer, James S. McDonnell Summer Institute in Cognitive Neuroscience. Dartmouth, NH. July 1991.

Colloquium, Program in Cognitive Science, Princeton University. Princeton, NJ. September 1991.

Invited Speaker, Fourth International Conference of the Society for Computer Science (Gesellschaft fuer Informatik). Munich, Germany. October 1991.

Colloquium, Siemens Nixdorf. Munich, Germany. October 1991.

Colloquium, Department of Psychology, University of Braunschweig. Braunschweig, Germany. October 1991.

Invited Speaker, Workshop in Time Series Analysis and Prediction. Santa Fe Institute for Nonlinear Studies, Santa Fe, NM. May 1992.

Colloquium, Xerox Palo Alto Research Center, September 1992.

Colloquium, Psychology Department, University of Denver, December 1992.

Colloquium, Department of Psychology, Carnegie-Mellon University. Pittsburgh, PA. February 1993.

Colloquium, Department of Computer Science and Engineering, Oregon Graduate Institute. Portland, OR. March 1993.

Colloquium, Department of Mathematics and Computer Science, Colorado School of Mines. Golden, CO. April 1993.

Invited Speaker, Connectionist Models Summer School. Boulder, CO. June 1993.

Colloquium, Department of Computer Science, University of Hamburg. Hamburg Germany, July 1993.

Faculty Lecturer, Summer School in Cognitive Neuroscience. Trieste, Italy. July 1993.

Colloquium, Department of Psychology, Oxford University. Oxford, England. February 1994.

Colloquium, Department of Psychology, University of California at San Diego. La Jolla, CA. June 1994.

Invited Speaker, Neural Information Processing Systems. November 1994.

Colloquium, Department of Computer Science, Colorado State University. Fort Collins, CO. January 1995.

Invited Speaker, Swedish Conference on Connectionism. Skovde, Sweden. February 1995.

Invited Symposium Speaker, Cognitive Neuroscience Society. San Francisco, CA. March 1995.

Colloquium, Santa Fe Institute. Santa Fe, NM. May 1995.

Faculty Lecturer, James S. McDonnell Summer Institute in Cognitive Neuroscience. Davis, CA. July 1995.

Invited Speaker, Lifestyle Technologies. Los Angeles, CA. August 1995.

Seminar, Department of Psychology, University of Toronto. Toronto, Ontario. October 1995.

Colloquium, Department of Psychology, McMaster University. Hamilton, Ontario. October 1995.

Colloquium, Department of Computer Science, Oregon Graduate Institute. Portland, OR. March 1996.

Invited Speaker, Apple Computer. Cupertino, CA. March 1996.

Invited Speaker, Conference on Neural Networks for Computing. Snowbird, UT. April 1996.

Invited Speaker, Montreal Workshop on Neural Networks. Montreal, Quebec. April, 1996.

Invited Speaker, Interval Research. San Jose, CA. May 1996.

Invited Speaker, Siemens Corporate Research. Princeton, NJ. June 1996.

Colloquium, Department of Cognitive Science, Johns Hopkins University. Baltimore, MD. June 1996.

Colloquium, Center for the Neural Bases of Cognition, Carnegie Mellon University. Pittsburgh, PA. March 1997.

Invited Speaker, Summer School on Adaptive Processing of Temporal Information. Vietri sul Mar, Italy. September 1997.

Colloquium, Institute for Research in Cognitive Science, University of Pennsylvania. Philadelphia, PA. October 1997.

Colloquium, Department of Psychology, University of Arizona. October 1997.

Colloquium, Systems Engineering, University of Pennsylvania. Philadelphia, PA. February 1998.

Invited Speaker, Neural Modeling of Brain and Cognitive Disorders Workshop, College Park, MD. June 1998.

Invited Participant. McDonnell Pew Program in Cognitive Neuroscience Annual Meeting, Montreal, PQ. June 1998.

Colloquium, Broadband Telecommunications Center, Georgia Institute of Technology. January 1999.

Colloquium, Department of Computer Science, University of Arizona. January 1999.

Colloquium, Department of Psychology, University of Iowa. March, 1999.

Colloquium, Department of Cognitive Science, University of California, Irvine. April, 1999.

Colloquium, AT&T Research Labs, Florham Park, NJ. June 1999.

Invited Participant. McDonnell-Pew Program in Cognitive Neuroscience Annual Meeting, San Diego, CA. June 1999.

Invited Speaker, International Joint Conference on Neural Networks. Washington, DC. July, 1999.

Colloquium, Department of Psychology, University of Pennsylvania. October, 1999.

Colloquium, Santa Fe Institute. Santa Fe, NM. February, 2000.

Colloquium, Department of Computer Science, University of Toronto. March, 2000.

Colloquium, Lucent Laboratories, Murray Hill, NJ. March 2000.

Invited Speaker, Fourth International Conference on Cognitive and Neural Systems, Boston, MA. May 2000.

Invited Speaker, Symposium on *Bayesian Models of Human Cognition*, Cognitive Science Society Conference, Philadelphia, PA. August 2000.

Invited Speaker, Workshop on *Network Models of Brain Function*, Banbury Center, NY. September 2000.

Invited Speaker, ESource Members’ Forum (Energy Industry Conference), Colorado Springs, November 2000.

Colloquium, Department of Psychology, McMaster University. November, 2000.

Colloquium, Microsoft Research, Seattle. January, 2001.

Lecturer, *Complex Systems Summer School*, Santa Fe Institute. June, 2001.

Invited Participant, NSF KDI Workshop, New Orleans, LA. April 2002.

Colloquium, Department of Computer Science, UC San Diego, June 2002.

Lecturer, *Ninth International Summer School in Cognitive Science,* New Bulgaria University, Sofia. July, 2002.

Invited Visitor, Center for Cognitive Science, Macquarie University, Sydney, Australia. September-October 2002.

Colloquium, Department of Psychology, University of New South Wales, October 2002.

Invited Speaker, ESource Members' Forum (Energy Industry Conference), Colorado Springs, November 2002.

Invited Speaker, *International Neuroscience Summit 2002*, Berlin, Germany. November 2002.

Invited Speaker, *American Neuropsychiatric Association*, Bal Harbor, FL. February 2004.

Keynote Speaker, *International Conference on Cognitive Modeling*, Pittsburgh, PA. July 2004.

Colloquium, Intel Research, Berkeley, CA. February 2005.

Invited Speaker, *Modeling Integrated Cognitive Systems* (AFOSR workshop), Troy, NY. March 2005.

Invited Speaker, *Computation in Neural and Machine Vision Systems*, Toronto, ON. June 2005.

Keynote Speaker, *Intelligent Environments '05.* Colchester, UK. June 2005.

Colloquium, Department of Psychology, Macquarie University, Sydney. July 2005.

Invited Speaker, Psychology Department, UCSD. January 2006.

Keynote Speaker, *Unconventional Computing '06*. York University, UK. September 2006.

Invited Speaker, Department of Cognitive Science (COGS200). University of California, San Diego. May 2007.

Invited Speaker, Workshop on *Closing the gab between neurophysiology and behavior: A computational modeling approach.* University of Birmingham, UK. June 2007.

Colloquium, Department of Computer Science, University of Nevada, Reno. October 2007.

Invited Speaker, Department of Cognitive Science (COGS200), University of California, San Diego. November 2007.

Invited Speaker, Temporal Dynamics of Learning Center Annual Meeting, University of California, San Diego, February 2009.

Colloquium, Department of Psychology, Indiana University, October 2009.

Colloquium, School of Informatics, Indiana University, October 2009.

Colloquium, Department of Brain and Cognitive Sciences, University of Rochester, March 2010.

Colloquium, Department of Cognitive Science, University of California Merced, March 2010.

Colloquium, Department of Cognitive Science, University of California Irvine, April 2010.

Invited Speaker, Temporal Dynamics of Learning Center Annual Meeting, University of California, San Diego, January 2011.

Invited Speaker, Department of Cognitive Science (COGS200), University of California, San Diego, April 2011.

Invited Speaker, Temporal Dynamics of Learning Center Annual Meeting, University of California, San Diego, January 2012.

Invited Speaker, *Workshop on Optimal Teaching*, San Diego, May 2012.

Invited Lecturer, *European Summer School in Cognitive Science*, Sofia, Bulgaria, July 2012.

Invited Speaker, *Summer Symposium on Visual Search and Selective Attention*, Munich, Germany, July 2012.

Invited Speaker, *NSF Workshop on Computational Cognitive Modeling*, Arlington, VA, May 2013.

Cognitive Brownbag, Department of Psychology, UCSD, May 2013.

Colloquium, Google Brain, Mountain View, CA, October 2013.

Invited Speaker, Temporal Dynamics of Learning Center Annual Meeting, University of California, San Diego, February 2014.

Invited Speaker, Reasoning Minds, Houston TX, February 2014.

Invited Speaker, *Personalized Learning Workshop*, Houston TX, April 2014.

Invited Speaker, Temporal Dynamics of Learning Center Annual Meeting, University of California, San DIego, February 2015.

Invited Speaker, Machine Learning Group, Department of Computer Science, University of Toronto, June 2015.

Invited Speaker, *NIPS Workshop on Reasoning, Attention, and Memory.* Montreal, December 2015.

Invited Speaker, ICML Workshop on *Machine Learning for Digital Education and Assessment Systems*. New York, NY, June 2016.

Invited Speaker, *NIPS Symposium on Recurrent Neural Networks*, December 2016.

Invited Speaker, *NIPS Workshop on Machine Learning for Education*, December 2016.

Invited Speaker, *NIPS Workshop on Future of Interactive Learning Machines*, December 2016.

Invited Speaker, Openstax Foundation, February 2017.

Invited Speaker, ECE Seminar Series, Rice University, February 2017.

Invited Speaker, Intelligent Systems Program, University of Pittsburgh, March 2017.

Cognitive Brownbag, Department of Psychology, UCSD, May 2017.

Keynote Speaker, Learning Understanding Cognition Intelligence Data Science (LUCID) Conference, Madison, WI, August 2017.

## Statement of Research Interests

### 1. Research Philosophy

I develop formal theories of human perception, cognition, and learning. Such theories serve to explain and interpret experimental data and provide a mechanistic understanding of brain function. They also have potential practical implications for rehabilitation following brain injury, and instruction, learning, and performance for the general population.

I have worked on a diverse set of topics, picking domains that call out for modeling approaches or that afford opportunities for synergistic collaborations with experimentalists. My work is unified by the methodological bias offered by a *rational perspective*, which asserts that human cognition is optimized to the structure of the environment in which it operates, subject to possible constraints on the information processing architecture (e.g., memory limitations, hardware restrictions). Rationality provides a satisfying understanding of cognition by characterizing the objectives of the cognitive system in quantitative terms, thereby suggesting elegant computational principles that underlie cognition and providing parsimonious accounts of a experimental data. Rationality allows us to describe the cognitive system as a good engineering solution to a set of conflicting demands, and allows us to understand the trade offs in the design of a cognitive system. Nonetheless, rationality is not as strong a bias as one might imagine, because significant flexibility can be attained via the architectural constraints.

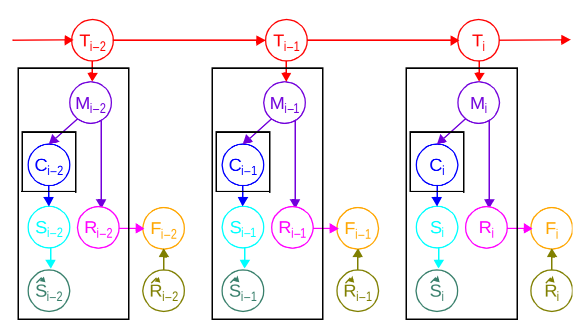
### 2. Computational Modeling of Human Cognition

I describe recent cognitive modeling efforts in four domains: executive control, visual attention, temporal dynamics of information processing, and improving human learning and retention.

#### *2.1. Executive Control and Sequential Dependencies*

Executive control refers to the flexible deployment of memory and attentional resources based on task demands. My work in executive control is distinguished from other approaches in two key respects. First, control is traditionally studied by examining how individuals respond to changing task demands in the context of a fixed stimulus environment (e.g., Wisconsin card-sorting task). In contrast, I believe a useful way to study control is to examine how cognition adapts to a *changing* environment when performing a *fixed* task. Second, modern neural theories of control are framed in terms of reinforcement learning: control operations are performed so as to maximize reward, produce few errors, etc. My perspective on control posits that optimal control operations can be inferred directly from a probabilistic model of the task and environment. Whereas the reinforcement learning approach requires hundreds of trials to learn to perform a simple task, my approach suggests a means for directly translating task instructions into appropriate behavior and then rapidly fine tuning behavior with subsequent task experience.

To examine control processes in the context of a changing stimulus environment, my recent work has emphasized modeling *sequential dependencies* in human cogntiion. Sequential dependencies are observed when an individual repeatedly performs a task, and performance is evaluated conditional on the statistics of the recent trial history. Sequential dependencies occur across a wide range of domains and experimental paradigms (Mozer, Kinoshita, & Shettel, 2007).

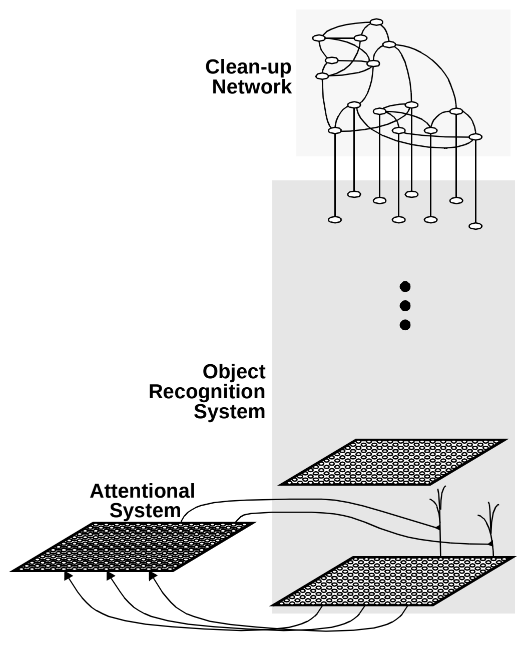



With collaborators, I've developed models of sequential dependencies in six domains: a simple choice task, where response latency depends on recent responses (Mozer, Colagrosso, & Huber, 2002); word reading where latency depends on recent item difficulty (Mozer, Kinoshita, & Davis, 2004); oddball detection, where latency depends on features of recent targets (Mozer, Shettel, & Vecera, 2006); ordinal categorization, where judgements are contingent on recent exemplars (Mozer, Jones, & Shettel, 2007); visual search, where performance depends on statistics of recent trials (Mozer & Baldwin, in press; Mozer & Wilder, in press); and a working memory task, where updates to working memory depend on the recent stimulus sequence (Reynolds & Mozer, in preparation).

These models share a theme: A generative model of the task environment is constructed, which specifies a probability distribution over possible states of the environment. The experience garnered from each trial is incorporated into statistics maintained by the model. Probabilistic inference on the model of the environment can determine the optimal setting of control parameters directly. Thus, each trial in a series leads to an update of control parameters. In this view, sequential effects reflect the operation of control processes at work. This approach also provides a framework for understanding how task instructions are mapped into initial control settings, and the manner in which these settings are refined with subsequent experience.

## 2.2. Visual Attention

My Ph.D. thesis presented a model of visual perception and spatial attention called MORSEL (Mozer, 1991). MORSEL includes an object recognition system that can identify multiple shapes in parallel and in arbitrary locations in the visual field, and an attentional system that determines where in the visual field to focus processing resources. MORSEL can explain a large corpus of experimental data, including perceptual errors that arise when several shapes appear simultaneously in the visual field (e.g., Mozer, 1983; Mozer, 1989), facilitatory effects of context and redundant information, and attentional phenomena.



MORSEL

MORSEL has proven itself by making experimentally-confirmed predictions (Mozer, 1991; Mozer & Sitton, 1998), but its greatest contribution is understanding the neuropsychological syndrome of hemispatial neglect. Lesioning MORSEL yields neglect dyslexia, a reading disorder that implies covert word recognition without awareness (Mozer & Behrmann, 1992). MORSEL also replicates the detailed pattern of patient performance on line bisection tasks (Mozer, Halligan, & Marshall, 1997). Indeed, the data fits are so good that MORSEL might serve as a diagnostic tool to characterize the nature of a patient’s attentional deficit.

Neglect patients fail to process visual information on one side, typically on the left. Much experimental work has focused on the question: with respect to what frame of reference is neglect of the left manifested? When a neglect patient shows a deficit in attentional allocation that depends not merely on the location of an object with respect to the viewer but on the extent, shape, or movement of the object itself, the inference has been made that attentional allocation must be operating in an object-based frame of reference. Through MORSEL, I show the highly counterintuitive result that this inference is not logically necessary (Mozer, 2002): object-based attentional effects can be obtained without object-based frames of reference. The psychological reality of various reference frames is crucial to distinguishing among theories of object recognition; my simulation results lend support to view-based theories.

MORSEL required only bottom-up attentional selection. Recently, we have focused our modeling efforts on the top-down control of attention, i.e., how task instructions modulate the deployment of attention (Mozer, Shettel, & Vecera, 2005; Mozer & Baldwin, in press). Our key insight is to define visual saliency as the probability that a location contains a target given the visual features in its neighborhood. Thus, attention is conceived not as a primitive, knowledge-independent process, but as fundamentally task driven and experience dependent. This perspective has allowed us to integrate the three major classes of attentional theories into a unified framework (Mozer & Wilder, in press).

### *2.3. Temporal Dynamics of Cognition*

I am attempting to develop a cognitive architecture that supports a theory of the temporal dynamics of information flow in cognition. The architecture is based on four minimal conjectures about neocortical computation. First, human cognition can be characterized by coarse-scale functional *pathways* that perform operations such as visual word recognition or response selection. Second, cognition arises from dynamic, task dependent interconnections among pathways. Third, the operation of a pathway exhibits a *speed-accuracy trade off*. Fourth, with each experience, a pathway tends to produce its response more rapidly.

#### *2.3.1. Specific modeling projects*

This *pathway* architecture has proven valuable in a broad range of modeling efforts. We explored the puzzling neuropsychological disorder of optic aphasia, which is marked by a difficulty in naming visually presented objects in the absence of visual agnosia—a perceptual deficit—or anomia—a naming deficit (Sitton, Mozer, & Farah, 2000). We explain optic aphasia in terms of relatively minor damage to two pathways, such that a cognitive task involving one damaged pathway or the other still yields near-normal performance. However, a task involving both damaged pathways—in this case, visual naming—is severely impaired due to interactions between the loci of damage. Our work offers a novel class of explanations for neuropsychological disorders involving multiple functional lesions to the cognitive architecture.

We applied the pathway architecture to the higher cognitive task of anagram solving (Grimes & Mozer, 2001). The architecture provides a way of conceptualizing the interface between symbolic (discrete) and subsymbolic (continuous) processing, because the asymptotic output of a pathway has a symbolic interpretation, although the underlying processing mechanisms are subsymbolic. We find that computational benefits are obtained by switching between symbolic and subsymbolic domains during problem solving, and have modeled experimental data concerning anagram solution times.

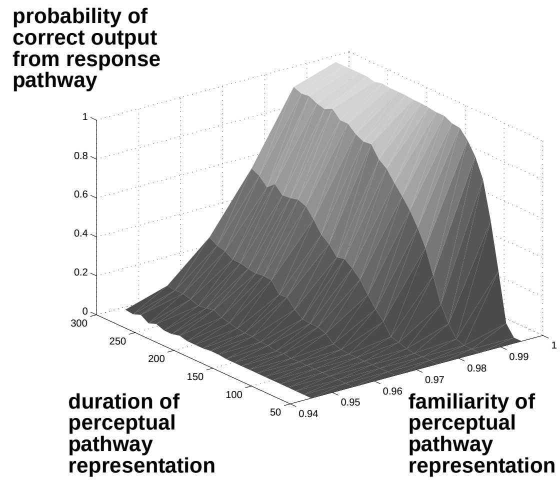
The pathway architecture is natural for explaining long-term priming phenomena (Mozer, Colagrosso, & Huber, 2003). It interprets priming facilitation in terms of two different mechanisms, one that raises the prior probability of primed stimuli, and the other that increases the transmission probability of primed stimuli. In explaining key data from the long-term priming literature, our model appears parsimonious and theoretically crisp relative to those of our competitors. One interesting finding is that a power law of practice emerges from our account, and explains the greater priming facilitation for relatively unfamiliar stimuli.

#### *2.3.2. General properties*

We have explored the pathway architecture at three levels of biological abstraction. At the *computational* level, we have modeled a pathway as a dynamic belief network in which nodes represent probability distributions over alternative states, and inference is performed by Bayesian updating. At the *psychological* level, we have modeled a pathway as a connectionist attractor network. At the *neurobiological* level, we have modeled a pathway as an associative mapping between layers of integrate-and-fire neurons. We aim to show that properties of the architecture hold true regardless of the implementation, and—by bridging across levels—to characterize the computation being performed by the biological hardware.

Using each of the three pathway implementations, we have explored simple tasks involving the coordination between a *perceptual* pathway that maps visual input to a semantic representation, and a *response* pathway that maps semantics to a classification response. These pathways in cascade exhibit certain characteristic behaviors, including:

- To ensure the correct behavior from the response pathway, the output of the perceptual pathway must be *stable*—well-formed and temporally persistent (see Figure next page). Stability, along with appropriate connectivity among pathways, therefore supports *arbitrary responses* to a stimulus and the *flexible control of behavior.*
- Stability becomes less necessary as the task becomes simpler and as the two pathways acquire coordinated practice on a task. Stability is thus essential for *complex decision making* and *decision making with limited domain expertise*.




The properties associated with stability are also associated with awareness, leading to a claim that *stability is the computational correlate of awareness*. If one accepts this claim, the behavioral properties of awareness—accessibility and reportability of mental states—drop out of the framework; there's no need to suppose special mechanisms of consciousness. We were originally resistant to this eliminativist view, but it seems to do a fine job of explaining scientifically-describable phenomena. We have modeled a wide variety of data from the literature related to subliminal priming, perception, and higher cognition, as well as qualitative features of awareness (Mathis & Mozer, 1995, 1996; Colagrosso & Mozer, 2005). Beyond the accounts of specific phenomena, the primary contribution of our computational approach is to unify a disparate collection of results, and to demystify the nature of conscious and unconscious cognition (Mozer & Borchert, in preparation).

### *2.4. Improving human learning and retention*

In collaboration with Hal Pashler of UCSD Psychology, I have been developing computational models to understand human fact learning, e.g., foreign language vocabulary learning. Our ultimate goal is to suggest practical techniques for improving learning in educational settings.

Through psychological studies and simulation models, we have examined four phenomena (Cepeda et al., submitted; Mozer, Howe, & Pashler, 2004; Pashler, Mozer, & Cepeda, submitted):

- *The spacing effect*. When an individual has multiple opportunities to study an item, memory retention is better if the interitem spacing is larger, i.e., if study is *spaced* rather than *massed*.
- *Benefits of testing during study*. In accordance with intuition, material is better learned when individuals test themselves while studying.
- *Forgetting curves*. As the retention interval between study and test is increased, performance drops. The exact shape of the forgetting curve has been the subject of much debate, but it currently appears that forgetting is best characterized by a power function.
- *The implications of guessing*. When individuals are forced to guess at an answer, subsequent performance is—surprisingly—not affected. However, whether or not individuals are willing to venture a guess is a strong predictor of the ease of learning, even when the guess is wrong.

Our work to date has primarily been on exploring unrelated mechanisms to explain data pertaining to these four phenomena. Our ultimate aim is to integrate these various mechanisms into a unified model, which could be of great use in designing instructional programs. The model would be fed a sequence of items to be studied, and could predict resulting performance. If the model accurately predicts human performance, it could then be used to *optimize* human performance. That is, given the constraint that a learner has a fixed amount of time to study certain material, the model could be used to structure the material (specify what order of items to study, how often to repeat items, when to test, etc.) so that retrieval accuracy would be maximized at the point when retrieval of the information was required.

## 3. ADAPTIVE, INTELLIGENT COMPUTER SYSTEMS

I dabble in using machine learning techniques to develop intelligent, adaptive computer systems. I highlight work on problems in several areas of artificial intelligence.

*3.1. Integrating symbolic and subsymbolic computation*

The cognitive science community has recognized the importance of integrating symbolic and subsymbolic (connectionist) models. One popular approach is to design hybrid models that have both symbolic and subsymbolic components. We have pursued an alternative approach in which symbolic mechanisms and representations are used to constrain the design of a connectionist network—its architecture, dynamics, and training procedure. The idea is to design a network that—with certain weights—can exactly mimic a symbolic system. However, when the network is given more freedom in exploring weight space, it can behave as a continuous dynamical system. This approach can take advantage of expressive, powerful symbolic mechanisms and representations, yet still utilize connectionist gradient-based learning techniques. The somewhat nonintuitive conjecture underlying this work is that it is fruitful to explore a subsymbolic search space even if the ultimate goal is a symbolic solution.

I have developed five models that test this conjecture: SLUG (Mozer & Bachrach, 1991) infers the structure of large, regular finite-state grammars using a compact symbolic encoding known as an *update graph*; RuleNet (McMillan, Mozer, & Smolensky, 1992) learns explicit condition-action rules over categorized instances; the demon model (Mozer & Das, 1993) learns rewrite rules that—in conjunction with an external stack—allow it to parse strings in context-free grammars; DOLCE (Das & Mozer, 1998) induces finite-state grammars by means of discreteness constraints on its internal state space; and a template-based approach to rule extraction (Alexander & Mozer, 1999) that casts neural network weights as symbolic *n*-of-*m* rules. In all cases, we show that this *symbolically-constrained subsymbolic approach* achieves more robust solutions than does a generic neural net approach, that it compares favorably to existing symbolic learning approaches, and that the resulting models can often be interpreted directly in symbolic terms.

*3.2. Recurrent architectures and algorithms*

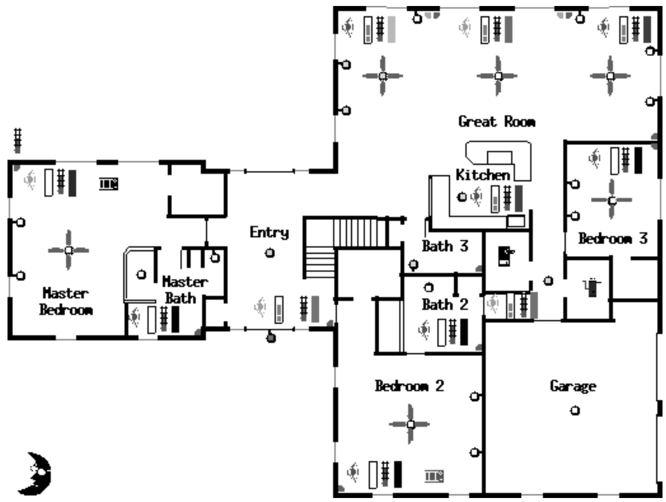
Cognition often involves the detection of regularities in temporally-extended sequences. However, recurrent neural network algorithms are notably poor at discovering relationships among events spanning long time intervals (Mozer, 1992, 1993). We have developed approaches to connectionist learning of structural regularities at longer time scales, using the domain of music composition, and have shown that they can solve problems beyond the capabilities of standard algorithms (Mozer, 1994; Schmidhuber, Mozer, & Prelinger, 1993). Our approaches are based on architectural constraints that bias learning. We have recently recast several such constraints from the neural net domain into a belief network formulation, with significant benefits in learnability and interpretability (Hochreiter & Mozer, 2001).

To address the challenge of constructing recurrent networks with attractor dynamics, we formulated a *localist attractor network* in which knowledge about the attractors is encoded locally in connections. This formulation solves problems associated with distributed attractor networks: they are trivial to wire up, spurious attractors are avoided, interpretation of their parameters and behavior is straightforward. We proposed a statistical interpretation of localist attractor net dynamics, which yields a convergence proof and a formal interpretation of model parameters (Zemel & Mozer, 2000).

*3.3. Applying machine learning techniques to difficult real-world problems*

The field of machine learning has exploded, in part because researchers have moved from toy problem domains to challenging, large-scale, real-world problems. I have been involved in ambitious projects of this nature in home automation, computer systems, and industry.

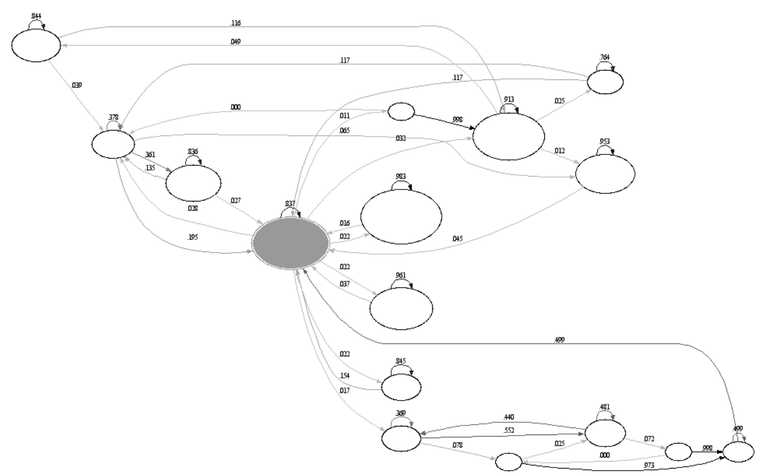
### 3.3.1. Home automation




Although the prospect of computerized homes has a long history, home automation has never become terribly popular because the benefits are seldom seen to outweigh the costs. One significant cost of an automated home is that someone has to program it to behave appropriately. We have explored an alternative approach in which the goal is for the home to essentially *program itself* by observing the lifestyle and desires of the inhabitants, and learning to anticipate and accommodate their needs (Mozer, 1998, 2004; Mozer & Miller, 1998; Mozer, Dodier, & Vidmar, 1997). We constructed a prototype system in an actual residence using neural networks for prediction and reinforcement learning for control. The residence was equipped with sensors to provide information about environmental conditions (e.g., temperature, ambient lighting level, sound, and motion) and actuators to control basic residential comfort systems—air heating, lighting, ventilation, and water heating. By predicting lifestyle patterns of the residents, the system could infer rules of operation that anticipate inhabitant needs while conserving energy.

The adaptive house project has received significant publicity, including stories in the *Chronicle of Higher Education*, *LA Times*, *New York Times*, *Boston Globe*, *The Economist*, *Electronic House* magazine, HGTV's *Extreme Homes* show, and newspapers around the world. It is one of the few "intelligent environments" that was in day-to-day use, and the first that did any sort of learning.

### 3.3.2. Understanding the performance of modern computer systems



Modern computer systems have become so complicated that computer-systems researchers have a difficult time identifying bottlenecks in program performance, which could arise from multiple levels within the computer, from hardware to firmware to operating system to virtual machine to the application software. As a result, systems researchers have essentially become cognitive scientists, attempting to infer the underlying operation of the computer system by instrumenting it and collecting statistics (e.g., cache misses, instructions executed per cycle). I am involved in a collaboration with systems researchers to reverse engineer these time-varying statistics to understand what is going on in the system. This task essentially involves building cognitive models of the system which are interpretable to a human analyst. We have projects to infer causality (Huang, Mozer, & Diwan, in preparation), and to infer the finite-state structure of the program execution (Richardson, Otte, Mozer, & Diwan, submitted).

### 3.3.3. Commercial ventures

I have been involved in four venture-capital-backed companies that rely on machine learning techniques. I am a co-founder of **Sensory Inc.**, in San Jose, which produces low-cost embedded speech recognition technology for consumer electronics, toys, and telephony applications. We've sold over twenty million chips embedded in about ninety different products, and we have about 85% market share in dedicated speech-recognition hardware (Mozer, 1996).

I served as chief scientist at **Athene Software** until its demise in 1999. Athene developed customer relationship management software for telecommunications carriers to predict profitability of customers and to reduce the likelihood of churn. Significant cost savings can be achieved using nonlinear prediction techniques and decision networks to identify potential churn and suggest interventions (Mozer, Wolniewicz, Grimes, Johnson, & Kaushansky, 2000; Mozer et al., 2002; Yan et al., 2003).

I currently serve on the technical advisory boards of several start ups. **AnswerOn** is concerned with churn prediction and remediation in the ISP industry. **Cognilytics, Inc.** estimates value and risk associated with various asset classes (e.g., residential mortgage portfolios). **Imagen Technologies** performs automated classification and analysis of medical imagery.