I am committed to *human-centric artificial intelligence (AI)*—AI that mimics and enhances human capabilities, understands and anticipates an individual's needs, and acts in synergistic coordination with individuals. My work explores the topics of:

- **Cognitively informed AI** : Incorporating insights from human perception and cognition into the design of AI architectures and machine learning methods;
- **Human optimization** : Developing software tools to improve how people learn, remember, and make decisions;
- **Cognitive modeling** : Building psychologically grounded models of human cognition that allow us to interpret behavioral and neurophysiological data; and
- **Intelligent environments** : Designing computer interfaces that are smarter, adaptive, and easier to use.

For each of these topics, I illustrate with past and current projects. All citations are to my work.

Cognitively informed AI

In machine learning, models are guided by computational theory. In neuroscience, models are guided by neuroanatomy and neurophysiology. I believe that there is an intermediate perspective with critical insight into the design of robust, flexible AI: the perspective of psychological process and cognitive theory. I seek insights from psychology that motivate AI and machine learning approaches. Here are two examples of current projects.

Leveraging theories of long-term memory to improve recurrent networks

People forget information over time. The rate of forgetting from long-term memory depends on an individual's exposure history: unfamiliar facts may quickly decay on first exposure, but the decay rate drops on subsequent exposure [2]. Some theories of memory explain changes in memory durability by positing that memories are stored in multiple traces decaying at different rates [63]. Contrast this account with a standard long short-term memory (LSTM) neural net architecture, which is designed to store memories indefinitely in a single trace and to forget only when triggered. We propose alternatives to LSTM-style recurrent nets that operate in a continuous time environment and adopt key theoretical assumptions from human long-term memory. In past work, we explored two models, one extending the gated-recurrent unit (GRU) to continuous time [54] and one with activation dynamics corresponding to time-scale inference of a Hawkes process [53].Current we are developing a third, the *Lifetime-Limited Memory* (LLM) [60]. LLM is premised on the assumption that—as in human memory—information has a finite lifetime, and therefore it is essential to store information about the age of the memory as well as to retrieve and erase memories based on age. We obtain age dependence by endowing each LLM cell with a bank of leaky integrators that have log-linear spaced time scales to ensure a wide dynamic range. Temporal localization can be performed via linear mixtures of these integrators, and their fixed decay incorporates forgetting yet ensures that gradients can propagate back in time regardless of the model parameters. LLM is designed to attain scale invariance in a continuous time environment, a problem in which I've had a longstanding interest [36]. Our aim is to predict *event sequences* produced by humans, consisting of discrete behaviors stamped with a real-valued time of occurrence (e.g., online product purchases, musical artist selections, and email communications).

Leveraging theories of consciousness to improve deep networks

Human consciousness is often viewed as a type of blackboard that enables modular components of the mind to communicate with one another. However, successful communication also requires a common language. We have studied information transmission in a modular dynamical neural net architecture [4, 27, 28] and define a notion of *well-formedness* that ensures successful communication. Well-formedness refers simply to the fact that the output of one component produces a representation that, through training, is familiar to another component. One can encourage well-formedness via attractor dynamics that 'clean up' representations [35, 40, 84]. We adopted this idea to denoise internal representations of deep networks. Internal attractor dynamics are trained through an auxiliary denoising loss to recover previously experienced hidden states from noisy versions of those states. We explored this approach in RNNs, where internal noise can accumulate over a sequence. We show that denoising serves as a valuable form of inductive bias, dramatically and robustly reducing generalization error in data-restricted environments [55]. This work is part of a larger agenda to characterize symbolic representations as strong interpretations of subsymbolic representations, and to explore the continuum of representations in between [1, 5, 29, 30, 34, 41, 47].

Human optimization

Cognitively-informed AI exploits our understanding of human cognition to improve AI methods. My complementary research on *human optimization* draws on AI methods to boost human cognition. This research leverages cognitive models to infer internal state and to perform surrogate-based optimization [14, 21]. Much of my present work is aimed at determining the most effective means of teaching and the manner in which to best present information for human consumption.

Learning and memory

Human memory is imperfect; thus, periodic review is required for the long-term preservation of knowledge and skills. We developed a method for efficient, systematic, personalized review that combines a psychological theory of memory with latent-factor models [25, 58]. Psychological theory characterizes basic mechanisms of learning and forgetting shared among members of a population, whereas latent-factor models use observations from a population to draw inferences about individuals. The method was integrated into a semester-long middle school foreign language course via retrieval-practice software. In a cumulative exam administered after the semester's end that compared time-matched review strategies, personalized review yielded a 16.5% boost in course retention over current educational practice and a 10.0% improvement over a one-size-fits-all strategy for spaced study. We are presently following up on this work by incorporating personalized review into environments with rich sensing to better infer a learner's state.

Another line of work aims to minimize the amount of practice required to induce perceptual categories, e.g., bird species. To determine the most effective training policy in an inductive category learning task, we conducted human experiments using Bayesian optimization to efficiently search a space of training policies defined by the ordering of exemplars [24]. To reduce the combinatorics, the spaces we search are guided by the psychological literature; we have focused on the sequencing of category labels—referred to as blocking and interleaving in the psych literature—and exemplar difficulty—referred to as fading [70]. To explore larger policy spaces, including ones that incorporate feedback from the learner, we are now performing surrogate-based optimization using hybrid neural net and cognitive models [74, 75]. This approach has identified teaching policies that empirically improve on the best policies discovered by psychologists.

Judgment

Psychologists have long been struck by individuals' limited accuracy in using rating scales to express internal sensations, impressions, and evaluations. Instead of using an absolute scale, individuals rely on reference points from recent experience. This relativity of judgment limits the informativeness of responses on surveys, questionnaires, and evaluation forms. Because the cognitive processes that transform internal states to responses are influenced by recent experience in a lawful manner, these processes can be inverted to *decontaminate* a series of ratings and obtain more veridical judgments. We have explored methods for decontamination using conditional random fields [64], latent-factor models [59], and matrix factorization methods [80], obtaining a roughly 25% reduction in rating error. Decontamination of human-labeled training data for ML classifiers (e.g., radiographs) can yield significant improvements in classifier accuracy [23].

Decision making

Individuals are often faced with temptations that can lead them astray from long-term goals, e.g., the calorie-laden treats that sabotage a diet. We're interested in developing interventions that steer individuals toward good behavior. In the realm of finance, an effective solution has been the prize-linked savings account, which incentivizes individuals to save by tying deposits to a periodic bonus-awarding lottery. In contrast to this one-size-fits-all approach, we investigate the effectiveness of customized bonuses [67]. We formalize a delayed-gratification task as a Markov decision problem and characterize individuals as rational agents subject to temporal discounting, costs associated with effort, and moment-to-moment fluctuations in willpower. We created an online delayed-gratification game in which the player scores points by choosing a queue to wait in and gradually advances to the front. Once the model is fit to data collected from the game, the instantiated model is used to optimize predicted player performance over a space of incentives. We demonstrate that customized incentive structures can improve goal-directed decision making.

Cognitive modeling

Cognitive models are formal theories of human perception and cognition that provide a mechanistic interpretation of experimental data. My collaborators and I have developed models in visual perception [32, 35, 49], spatial attention [33, 35, 66], awareness [4, 27, 28], deficits arising from acquired brain dysfunction [38, 43, 44, 50, 77], and executive control [45, 46, 71]. In this work, I pursued the traditional goal of explaining universal phenomena emerging from data aggregated across individuals and across individual noisy observations. However, in order to use models for human optimization, my current goals are to interpret and anticipate the behavior of a *specific* individual in a *particular* context. I describe work in three areas aimed at this granular variety of modeling.

Sequential dependencies

Human behavior at any moment is modulated by recent incidental experience. For example, when individuals drive in a realistic automobile simulator, their recent history of stops and starts has a dramatic and systematic effect on the subsequent latency to brake [7]: reaction times can be slowed enough to increase braking distance by 10 feet when traveling at highway speeds. Recent history is typically treated as noise to be averaged out of data in experimental psychology, but it can explain about 25% of the variability in response latency. Sequential dependencies are manifest in every aspect of behavior [57]. I have contributed to empirical and modeling efforts on sequential dependencies in perceptual grouping [83], visual search [3, 42, 65], allocation of spatial attention

[19, 20, 78], response initiation [9, 46], concept learning [52], word interpretation [15, 16, 17, 56], pain judgment [26], motor control [81], and executive control [71].

Our research has identified the specific environmental statistics to which the brain adapts [79] and the neural loci of multiple distinct mechanisms of adaptation [8, 79]. With sensitive experimental methods and modeling tools, we have shown that the effects of incidental experience can span hundreds of trials, and the decay of past experience follows a power function [81]. These results contrast with the conventional view that sequential dependencies are short lived and decay exponentially. We interpret sequential dependencies as reflecting the operation of control mechanisms that tune cognition in a nonstationary task environment [45, 71]. We posit that the mind constructs a generative model of the environment, [65], and probabilistic inference on this model determines the optimal control settings [71].

Memory retention

Forgetting is typically studied in the laboratory, but we have analyzed a large-scale corpus of students learning foreign language vocabulary using online software to understand forgetting in the wild [72]. A generic power-law model of forgetting explains only about 10% of variability in the data, but when the model is individuated by including features describing a student's study history, our model's explanatory power doubles. Incorporating latent variables representing student ability and lesson difficulty improves the model further [58].

When material is studied across multiple sessions, the temporal distribution of study has a profound effect on retention: spaced study leads to slower forgetting than massed study [2, 10]. The optimal spacing of study is not fixed but increases with the interval over which the material must be retained. We developed a model, MCM, that accurately predicts the optimal spacing of study for a wide range of published results [63]. The model is strongly predictive in that it uses only the rate of forgetting following an initial study session to predict memory recall following two or more study sessions, whereas other models in the literature fit data only post-hoc.

When learners test themselves during study, memory retention is improved. We have built models to explain phenomena surrounding self-testing [51, 69]. Our current challenge is to develop models that incorporate feedback from learners as they self test. Existing models, including MCM, confound the causal consequences of successful retrieval with the inferential consequences (the inference that memory must have been strong in the past in order to succeed at test time). In collaboration with researchers at Quizlet, we are currently exploring a Bayesian variant of MCM that teases apart the causal and inferential consequence of student self-testing [62].

Cognitive skill acquisition

Models of cognitive skill acquisition observe learners as they practice a series of problems (e.g., algebra) and predict the point at which mastery is achieved for a given skill (e.g., simplifying fractions). Two models dominate the literature: BKT, a hidden Markov model, and DKT, a generic recurrent net for sequence processing that is trained to become a human meta-learner. BKT is interpretable but DKT tends to perform better. We have proposed extensions to BKT that brings it to the performance level of DKT without harming interpretability, including: automatic discovery of underlying skills using nonparametric Bayesian methods [22]; inclusion of latent factors representing student ability and skill difficulty [11, 13]; and incorporation of a forgetting mechanism [12]. We have conducted extensive comparisons of BKT and DKT, with the aim of understanding what type of information DKT exploits that BKT cannot [12, 31, 82]. We are moving toward hybrid models that combine the strengths of BKT and DKT.

Intelligent environments

Adaptive house

In the 1990's, I had brief popular-press fame after turning a home-renovation project into the world's first *adaptive house* [37, 48, 61]. My home was instrumented with 75 sensors detecting environmental conditions—temperature, ambient lighting level, sound, and motion—and actuators controlling basic residential comfort systems—air heating, lighting, and water heating. The system predicted occupancy patterns and operated comfort systems to anticipate lifestyle needs while conserving energy. The system was composed of neural net predictors and reinforcement learning controllers that minimize a sum of energy and discomfort costs. We found that subtle statistical regularities in the home occupants' behavior could be exploited for increased comfort and decreased energy costs [68]. The Nest thermostat came along twenty years later, but my adaptive home was more sophisticated in both control inputs and outputs and controller design. The adaptive home also performed predictive lighting control: lighting anticipated impending occupancy and occupant activities [6]. My current research is informed by two key lessons from this project: make the interface invisible and focus on tasks that individuals are least able to manage on their own [39].

Augmenting perception

We are developing methods to augment human perception on challenging visual search tasks like satellite imagery analysis. One technique leverages automatic classifiers. Although computer-assisted diagnosis systems have long existed which place boxes around potential targets, they have proven ineffective or harmful. We explore an alternative method of highlighting displays in which the saliency of regions of the visual field is modulated in a graded fashion based on classifier confidence. We show that this *soft* highlighting beats traditional *hard* highlighting, and yields human-classifier synergistic performance that beats either acting alone [18]. We use a similar approach to highlight displays with the spatiotemporal fixation sequences of experts, allowing novices to effortless learn where and how to look in images [76]. As an alternative to modulating images, we have proposed modifying the *task*, allowing novices with no training to perform like experts [73]. Instead of asking novices to classify images, say in a diagnostic dermatology domain, we leverage the human visual system's ability to judge similarity. Individuals are asked to compare a query image to sets of reference images and to select the best matching reference from each set. If the references have known classifications, this procedure yields an implicit classification of the query image. References are chosen to maximize expected classification accuracy under a cognitive theory of similarity-based choice. Novices attain expert-like performance, even without knowing what the classification task is.

Intelligent textbooks

We are developing intelligent textbooks through a collaboration with OpenStax, a nonprofit that supports open-access college-level digital textbooks. Our textbooks allow readers to highlight and tag text and to add questions and notes. Through these annotations, we obtain a window into a student's mental state during initial engagement with the material. We expect that these data will facilitate interventions, e.g., recommending other material to read, offering appropriately timed review questions, and early detection of comprehension difficulties. As a component of this project, we are building a *predictive scroller* for small screen devices. Given the visible content on screen and a user's history, the tool estimates reading time. Automatic scrolling frees the reader from swiping to advance the text, and mismatched reading-time predictions can indicate mind wandering and comprehension difficulties, triggering interventions.

References

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

[2] 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.

[3] 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 Academy of Sciences*, 110:10438–10445.

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

[5] Das, S. & Mozer, M. C. (1994). A unified gradient-descent/clustering architecture for finite-state machine induction. In Cowan, J. D., Tesauro, G., & Alspector, J. (Eds.), *Advances in Neural Information Processing Systems 6*, pp. 19–26, San Mateo, CA. Morgan Kaufmann.

[6] 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.

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

[8] 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.

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

[10] 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.

[11] 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 Kravcik, M., Santos, O. C., & Boticario, J. G. (Eds.), *Proceedings of the 4th International Workshop on Personalization Approaches in Learning Environments*, pp. 7–15. CEUR Workshop Proceedings, ISSN 1613-0073.

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

[13] 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 Stamper, J., Pardos, Z.,

Mavrikis, M., & McLaren, B. M. (Eds.), *Proceedings of the Seventh International Conference on Educational Data Mining*, pp. 99–106. Educational Data Mining Society Press.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] Kneusel, R. T. & Mozer, M. C. (2017). Improving human-machine cooperative visual search with soft highlighting. *ACM Transactions on Applied Perception*, 5:3:1–3:21.

[19] 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.

[20] 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.

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

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

[23] Lindsey, R. V. & Mozer, M. C. (2018). Training deep neural networks on datasets having conditionally dependent labels. In preparation.

[24] Lindsey, R. V., Mozer, M. C., Huggins, W. J., & Pashler, H. (2013). Optimizing instructional policies. In Burges, C. J. C. (Ed.), *Advances in Neural Information Processing Systems 26*, pp. 2778–2786. Curran Associates, Inc.

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

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

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

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

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

[30] 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.

[31] Montero, S., Arora, A., Kelly, S., Milne, B., & Mozer, M. C. (2018). Does deep knowledge tracing model interactions among skills? In *Proceedings of the Eleventh International Conference on Educational Data Mining*. Educational Data Mining Society Press.

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

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

[34] Mozer, M. C. (1991). Discovering discrete distributed representations with iterative competitive learning. In Lippmann, R. P., Moody, J., & Touretzky, D. S. (Eds.), *Advances in neural information processing systems 3*, pp. 627–634, San Mateo, CA. Morgan Kaufmann.

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

[36] Mozer, M. C. (1993). The induction of multiscale temporal structure. In Moody, J. E., Hanson, S. J., & Lippmann, R. P. (Eds.), *Advances in Neural Information Processing Systems 4*, pp. 275–282, San Mateo, CA. Morgan Kaufman.

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

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

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

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

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

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

[43] 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.

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

[45] 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*, pp. 828–833. Erlbaum Associates, Hillsdale, NJ.

[46] Mozer, M. C., Colagrosso, M. D., & Huber, D. H. (2002). A rational analysis of cognitive control in a speeded discrimination task. In Dietterich, T., Becker, S., & Ghahramani, Z. (Eds.), *Advances in Neural Information Processing Systems 14*, pp. 51–57. NIPS Foundation, La Jolla, CA.

[47] Mozer, M. C. & Das, S. (1993). A connectionist symbol manipulator that discovers the structure of context-free languages. In Hanson, S. J., Cowan, J. D., & Giles, C. L. (Eds.), *Advances in Neural Information Processing Systems 5*, pp. 863–870, San Mateo, CA. Morgan Kaufmann.

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

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

[50] 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.

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

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

[53] Mozer, M. C., Kazakov, D., & Lindsey, R. V. (2016). Neural Hawkes process memories. NIPS Symposium on Recurrent Neural Networks, https://www.cs.colorado.edu/ mozer/Research/Selected%20Publications/talks/Mozer_NeuralHawkesProcessMemory_NIPS2016.pdf.

[54] Mozer, M. C., Kazakov, D., & Lindsey, R. V. (2017). Discrete-event continuous-time recurrent networks. arXiv:1710.04110 [cs.NE].

[55] Mozer, M. C., Kazakov, D., & Lindsey, R. V. (2018). State-denoised recurrent neural network. arXiv:1805.08394 [cs.NE].

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

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

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

[59] 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, F., & Weinberger, K. Q. (Eds.), *Advances in Neural Information Processing Systems 24*, pp. 1791–1799, La Jolla, CA. NIPS Foundation.

[60] Mozer, M. C. & Meirhofer, M. (2018). Lifetime limited memory networks. Working draft.

[61] 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 Giles, C. L. & Gori, M. (Eds.), *Adaptive processing of temporal sequences and data structures*, pp. 370–388. Springer Verlag, Berlin.

[62] Mozer, M. C., Mooney, S., Sun, K., & Stansbury, D. (2018). Disentangling inferential and causal effects of feedback on human memory state. Working manuscript.

[63] 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 Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., & Culotta, A. (Eds.), *Advances in Neural Information Processing Systems 22*, pp. 1321–1329, La Jolla, CA. NIPS Foundation.

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

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

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

[67] Mozer, M. C., Sukumar, S., Elliott-Williams, C., Hakimi, S., & Ward, A. F. (2018). Overcoming temptation: Incentive design for intertemporal choice. Submitted for publication.

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

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

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

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

[72] 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.

[73] Roads, B. D. & Mozer, M. C. (2017). Improving human-computer cooperative classification via cognitive theories of similarity. *Cognitive Science: A Multidisciplinary Journal*, 41:1394–1411.

[74] Roads, B. D. & Mozer, M. C. (2018). Optimizing training sequences for human category learning. In Preparation.

[75] Roads, B. D. & Mozer, M. C. (2018). Predicting the difficulty of human category learning using exemplar-based neural networks. In Preparation.

[76] 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.

[77] 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.

[78] Vatterott, D. B., Mozer, M. C., & Vecera, S. P. (2018). Rejecting salient distractors: Generalization from experience. *Attention, Perception, & Psychophysics*, 80:485–499.

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

[80] Wilder, M., Jones, M. C., & Mozer, M. C. (2014). Decontaminating human judgments: exposing and removing rating aberrations due to sequential biases. Unpublished manuscript.

[81] 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.

[82] Wilson, K., Xiong, X., Khajah, M., Lindsey, R., Zhao, S., Karklin, Y., Van Inwegen, E., Han, B., Ekanadham, C., Beck, J., Heffernan, N., & Mozer, M. (2016). Estimating student proficiency: Deep learning is not the panacea. In Baraniak, R. G., Ngiam, J., Studer, C., Grimaldi, P., & Lan, A. S. (Eds.), *Proceedings of the 2016 NIPS Workshop on Machine Learning for Education*. Barcelona, Spain.

[83] 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.

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