Report from the NSF Workshop on Integrating Approaches to Computational Cognition

Matt Jones, University of Colorado
Richard M. Shiffrin, Indiana University
Joshua B. Tenenbaum, Massachusetts Institute of Technology
Alan L. Yuille, University of California Los Angeles
Jun Zhang, University of Michigan
Michael C. Mozer, University of Colorado
Bradley C. Love, University College London
Xiaojin Zhu, University of Wisconsin-Madison
Thomas L. Griffiths, University of California Berkeley
Charles Kemp, Carnegie Mellon University
Yann LeCun, New York University
Hongjing Lu, University of California Los Angeles
David A. McAllester, Toyota Technological Institute at Chicago
Ruslan Salakhutdinov, University of Toronto
Bernhard Schölkopf, Max Planck Institute for Intelligent Systems
Satinder Singh, University of Michigan
Robin D. Thomas, Miami University
Angela J. Yu, University of California San Diego

At a workshop sponsored by the National Science Foundation, 18 distinguished researchers from the fields of Cognitive Science (CS) and Machine Learning (ML) met in Arlington, VA in May 2013 to discuss computational approaches to cognition in human and artificial systems. The purpose of the workshop was to identify frontiers for collaborative research integrating (a) mathematical and computational modeling of human cognition with (b) machine learning and machine intelligence. The researchers discussed opportunities and challenges for how the two fields can advance each other and what sort of joint efforts are likely to be most fruitful.

There are several reasons to believe that theories of human cognition and of machine intelligence are currently in position to greatly benefit from each other. The mathematical and computational tools developed for designing artificial systems are beginning to make an impact on theoretical and empirical work in CS, and conversely CS offers a range of complex problems that challenge and test ML theories. ML systems would likely be more successful if they were more flexible and adaptive, like human cognition, and if they were built on richer representations that (in some sense) embody meaning or understanding as opposed to black-box statistics. The synthesis is also timely because CS researchers are starting to work on large datasets (e.g., Amazon's Mechanical Turk and web-based corpora), and a major focus of ML in the last decade has been to develop tools that can succeed in complex, naturalistic situations.

The remainder of this report summarizes the main conclusions from the workshop regarding opportunities for research integrating CS and ML. The first section discusses potential benefits to society, including a new generation of intelligent artificial systems, improved decision-making both for individuals and in public policy, advances in adaptive and personalized education and training, and new computational frameworks for understanding the brain. The remaining sections explore specific collaborative research areas where the next breakthroughs might occur, followed by mechanisms that NSF and other funding agencies could pursue to encourage greater collaboration. In the Supplementary Material[1] to this report, the workshop participants describe recent research they believe exemplifies the potential of bridging these two fields.

Societal Contributions

Recent developments in CS and ML have put these fields in a position where increased collaboration and integration over the next several years may produce technological solutions to many important current societal issues. The following subsections describe some of the most promising areas of development and practical impact.

---

[1] The Supplementary Material can be obtained at: http://matt.colorado.edu/compcogworkshop/supplement.pdf

Human-Like Artificial Intelligence

Collaboration between CS and ML could soon produce the next generation of artificially intelligent (AI) systems. These developments could transform everyday life by creating a new augmented reality, with computer systems that can have many of the desirable properties of human cognition and with which humans can efficiently interact and communicate.

Making ML systems more cognitive could enable them to interact with humans more efficiently, for example by taking natural-language instructions or queries instead of requiring detailed programming for any task. They could also take on many of the strengths of human cognition, including robustness, flexibility, adaptiveness, analogical reasoning, and the ability to learn and generalize from a small number of examples.

Potential applications of this next generation of AI include manufacturing, transportation, business, usability, human-computer interaction, forensics, and database search. Also important, especially for systems interacting with humans, is social cognition or theory of mind. For example, to build a self-driving car, it is essential to have an internal model of how humans reason and behave, including deeper theory-of-mind reasoning.

Building cognitive principles into ML systems may also lead to systems that do not need to be programmed for individual tasks. Instead, they can learn new tasks from experience, interpret ad hoc instructions, combine knowledge derived from these two sources (instructions and experience), and acquire abilities to perform multiple alternative tasks depending on context.

Decision-Making

Integration of CS and ML could lead to important advances in theories of decision-making, with applications to health care, policy, and commerce. A new theory of behavioral economics grounded in modern probabilistic formulations of human cognition would be context-sensitive and more applicable to complex, dynamic, time-sensitive environments. Current computational models of cognition take into account bounds on computational resources and other aspects of the cognitive architecture of agent in question. Using bounds from these more accurate cognitive models should enable better predictions from both prescriptive and descriptive theories.

Improved theories of human decision-making could benefit society through a better understanding of the relationship between policy and human behavior, as well as through new means for helping people to make better personal decisions (e.g., in health and finance). Improvements might also come for computerized decision-making systems, such as product recommenders and personalized search and advertisements.

Training and Education

The field of education has the potential to be transformed by the internet and intelligent computer systems. Evidence for the first stage of this transformation is abundant, from massive open online courses (MOOCs) to web sites such as Khan Academy (khanacademy.org) that offer online lessons and drills. However, the delivery of instruction via web-connected devices is merely a precondition for what may become an even more fundamental transformation, as computational methods from CS and ML are applied to personalize and optimize education. Specifically, CS models can be used to generate quantitative predictions about learning outcomes for individual students or groups, and ML methods can use high-dimensional data to infer latent states of those models and thereby to optimize and personalize the educational process.

Teaching can be personalized both at the level of learning styles and at the level of specific instruction, feedback, rewards, scheduling, and assignments. Optimization can occur both for the student, in terms of knowledge acquisition, understanding, attention, and retention; and for the teacher, in terms of efficiency of time and other resources, as well as real-time assessment and adaptation in response to student progress. These potential advances would apply not only to STEM education and other traditional educational settings, but also to job or military training, training of perceptual expertise (e.g., radiologists or security screeners), or training of reasoning, decision-making, and cognitive control.

Computational models of learning in CS have become highly complex, incorporating assumptions about knowledge representation, processing, memory, expertise, processing of instructions, and so on. These models often use rich (i.e. flexible, structured, hierarchical) conceptual representations that may be restructured as the student learns. They can take into account constraints and limitations of the human cognitive system, capture cognitive biases, and explain dimensions of individual differences. They can capture facets of learning specific to educational settings, such as differences between group instruction (e.g., classroom) and individual tutoring, and they can embody specific educational theories about pedagogy (e.g., scaffolding, multiple routes to solution). Applied to particular knowledge domains, CS models can assume complex and specific knowledge structures as possible starting points for students' knowledge. Finally, they can address psychological questions such as how to reward and motivate students (e.g., importance of play or exploration) and when a student is ready for learning a certain concept, because of developmental stages or individual learning trajectories.

Given such a specification of a learning situation and the cognitive processes of a learner (or class), ML methods could be used to maximize desired learning outcomes. ML has a variety of optimization tools that could be used to determine the parameters (e.g., content, timing, and feedback) that will lead to optimal learning according to a cognitive model. Moreover, hierarchical and nonparametric models can be used to estimate and accommodate individual differences among students. Data-mining techniques could be applied to the massive quantities of data coming available from MOOCs and from online learning experiments. Similarly, ML techniques could be applied to subtle aspects of a student's behavior—such as facial expressions, fixation sequences, response latencies, and errors—to make explicit inferences about the student's latent state of knowledge and understanding, which could then be used as real-time feedback for instructors or automated tutoring systems.

ML theories of learning per se are of course also relevant. There are formal characterizations of many learning tasks as ML problems, such as classification (Khan et al., 2011), and it will be important to determine how the principles from this research extend to richer (more cognitive) tasks such as concept acquisition and skill learning. Active learning (Castro et al., 2009; Settles & Burr, 2012) and semi-supervised learning (Gibson et al., 2013; Zhu & Goldberg, 2009) are recent foci of ML research that could be directly brought over to the education scenario. Formal distinctions between types of learning, such as model-based versus model-free, might also be applied to educational settings.

Understanding the Brain

Understanding how the brain works is arguably one of the most important challenges for contemporary science. Although this question is often considered the domain of neuroscience, CS and ML have much to contribute, by offering a top-down theoretical framework for interpreting the great quantities of data being produced by neuroscience. For example, neuroscience research has been heavily influenced by ideas from control theory as formalized in the ML framework of reinforcement learning (Sutton & Barto, 1998). Likewise, CS research on perceptual priming led to investigation of repetition suppression, which has become one of the most robust findings in neuroscience and neuroimaging (Grill-Spector et al., 2006). Whereas traditional constructs in cognitive psychology (e.g., representation, concepts) have not taken hold in neuroscience, the language of statistical inference and decision-making under uncertainty that unites computational CS with ML is a potential unifying language that bridges to work at the neural level.

Combining CS and ML could thus offer a new, more principled foundation for studying the brain. ML techniques will be valuable for processing the rich datasets being generated, and CS will offer psychological theories for interpreting those data. A cognitively grounded computational understanding of the brain would also enable technological advances, such as neural interfaces for people with sensory or motor disabilities. From a policy point of view, such a top-down approach could be valuable in directing more targeted research towards mental health—because ultimately it is "mental" health we care about, not just brain health.

Promising Research Areas Integrating Computation and Cognition

At a more technical level, there are many research domains that are increasing in importance and for which progress will be greatly enhanced by synergy between approaches based in CS and ML. The two fields are just now poised to tackle many of these issues. We have chosen a few of these to highlight in this report.

Towards Better, Deeper Generalization

CS and ML take complementary approaches to learning. CS focuses on rich, flexible knowledge representations, and often on how people can acquire and use knowledge with very little experience. ML's focus is more on sophistication of learning algorithms, on representations that perform very well for a specific task (e.g., image classification), and on leveraging large volumes of complex or high-dimensional data.

One goal for integrating ML and CS is to extend ML methods to apply to richer knowledge structures, such as symbols, relations, graphs, analogies, rules, grammars, and hierarchical and compositional representations. Most advanced human thought (e.g., mathematics) is hierarchically structured, and motor and visual abilities arguably are as well, built on vocabularies of basic elements that are composed to form bigger patterns. Whereas much of current ML uses flat feature vectors or n-gram–style representations, inference over richer, more cognitive representations is crucial for the next level of AI, to enable flexibility and abstract generalization.

A second goal is a formal understanding of learning from very little data. Humans are able to make strong and abstract generalizations based on little experience, to a degree not approached by current artificial systems. Recent research on one-shot learning has offered initial answers in the case of category learning, whereby superordinate categories determine inductive biases for member categories, in terms of which dimensions are relevant for basic-level categories (e.g., number of legs is a strong cue for most animal species, even a species the learner does not yet know). These inductive biases enable rapid learning about new categories from just a few (or even one) examples (Salakhutdinov et al., 2012).

A third goal is autonomous discovery (or selection, tuning, or restructuring) of representations. Most ML systems start with hand-coded features that were selected for a given task. Approaches such as kernel methods and deep networks can be seen as taking an opposite approach, entertaining an infinite variety of possible features. Human cognition occupies a middle ground, with representations that are flexible in many ways but strongly resistant to change in other ways. This constrained flexibility is likely critical for humans' ability to learn rapidly in new tasks, stimulus domains, or dynamic environments. Another factor that might be important for learning of representations is the need to perform multiple tasks. Whereas most ML systems are designed for one or a narrow range of tasks, humans learn to reuse knowledge in many different ways, which likely encourages representations that are more robust and abstract. Development of ML systems that are capable of humanlike transfer learning might lead to more sophisticated computational mechanisms for representation learning.

Flexibility and Self-Programming

Related to the question of how people acquire new representations is that of how they generate strategies for novel tasks. Even when task instructions are given, there is the question of how people convert those instructions to an internal program to direct their behavior, as well as how they rapidly tune that program with experience. Better understanding of these psychological processes might be informative to research in computer science on program synthesis and on automatic programming from natural-language input.

As discussed in the previous subsection, a simple characterization of the brain/mind as universally more flexible than current artificial systems is incorrect. In many ways, the brain is quite inflexible, for example in how perceptual systems parse sensory input into primitive dimensions. Cognitive scientists have argued that humans' ability to learn from limited data derives from our reliance on strong prior expectations about the abstract structure of conceptual domains (e.g. in research on one-shot learning). Therefore the real computational problem is determining the combination of constraints and flexibility that lead to intelligent behavior.

Planning and Acting in Complex Dynamic Environments

Real environments have complex structure and are constantly changing, and human agents continuously interact with such environments with dynamic actions that reflect implicit and explicit planning and decisions. Although laboratory studies have traditionally focused on discrete actions (e.g., binary responses), CS can and should measure and model these more continuous aspects of behavior. To take some examples, one can measure the timing and spatial trajectory of responses (such as the movement of a hand from a start point to a decision response button), eye movement paths and timing, facial expression, body posture and movement, and neural responses, and this can all be done as the environment changes. Such measurements provide critical insight into the way humans interact with a dynamic environment by planning, dynamic decision-making, and continuous actions. CS researchers are beginning to exploit such rich veins of data, and ML researchers are starting to develop and use methods that produce normative ways humans can plan, act, and change in dynamic environments. A critical point is the human agent (or agents) is a part of the environment and all actions taken change the environment.

Thus, planning of actions must take the anticipated action consequences into account.

The complexities raised by the rich and changing data sets, and the difficulties of developing models of the planning and execution of actions in such interactive and dynamic environments, require joint research efforts from the ML and CS communities. Conversely, such joint efforts offer the hope of an enormously increased understanding of human cognition and action, as well as the ability to tackle many real-world problems of importance to society. Potential applications can be found in decisions and actions required in such arenas as medicine, business, forensics, and navigation.

There is enormous variation in the time scale over which dynamic actions, planning, and decisions take place in changing environments. Thus one can explore simple perception and memory retrieval taking place in a few hundred milliseconds, or planning, action and decisions that occupy a significant portion of the lifespan (e.g., buying a home, planning a business venture, managing investments, or converging upon a marriage partner).

A final considerations is that the factors important for planning and action go well beyond the simple assessment of physical stimuli that has been the focus of much past research. The environment, and the way that one's actions affect the environment, are so complex that planning and decision must often rely on high-level abstractions such as analogy: We find situations in memory that are sufficiently similar to the present situation to draw an analogy and guide future actions based on the remembered outcomes. This perspective brings all the constraints of human memory and pattern matching into the modeling of action, planning, and decision. A critical topic then becomes the way that a system designed by evolution and tuned by experience adapts to both the environment and cognitive constraints so as to optimize performance.

From Learning Algorithms to Learning Models

ML has many powerful learning algorithms, but in most cases there is no model of the learner per se, as an agent in which a given algorithm is embedded. In many applications, it is the learner as a whole that needs to be optimized. This is especially true for education: for optimizing learning and teaching, and for analyzing and experimenting with online education. In that case, one must take into account global characteristics of the learner such as curiosity, motivation, interest, active learning, exploration, and so on.

Even outside educational applications, ML could benefit from developing complete agents that are capable of multiple tasks. Such an approach would encourage integration of learning with decision-making, planning, and agent architectures. It would create a need for models of active, adaptive internal processes that are subject to bounded rationality, which are currently not present in most ML architectures. Extant production systems models such as SOAR (Laird, 2012) and ACT-R (Anderson et al., 2004) aim for this type of internal process modeling, although they might be better constrained by tying to ML theories of statistical optimality.

The perspective of the operation of the whole agent, in optimizing its learning and behavior, aligns with the notion of bounded rationality, which views rationality in the context of a finite agent or program, and as a property of the agent's full range of behavior (across many tasks). This approach might naturally lead to systems with rich, humanlike knowledge representations (Russell & Wefald, 1991). The approach would also benefit from enlarging the conversation to include AI and robotics, not just ML proper.

Theoretical Limits of Learning

ML offers formal theorems proving bounds for certain learning problems or algorithms, pertaining to what is possible in principle: what can be learned, how, and when (Langford & Shawe-Taylor, 2002; McAllester, 2003; Zhu et al., 2009).

These results are mathematically elegant, but they often give bounds that are far too weak for practical purposes. More problematically, they ignore the key point that humans almost never learn anything from scratch. A typical human learner has acquired a variety of knowledge about the world. When faced with a new situation, the learner can use priors (e.g., most coins are balanced, so I expect this new coin to be balanced) or can adapt existing models to the data, perhaps by analogy (e.g., physicists using models of the solar system to understand the structure of atoms). There have been some attempts to model this type of problem (e.g., Kemp & Tenenbaum, 2009) but little or no attempts to prove theorems. Perhaps CS could motivate formal learning bounds for these more realistic types of learning situations.

Using Big Data to Understand Cognition and Behavior

Recent years have seen an explosion of large datasets, some from controlled experimentation and some from uncontrolled observations and productions. Many of these datasets are produced by humans, indirectly or directly, and can in principle be used to understand cognition and behavior and to guide actions and decisions that depend on such understanding. The ML community has been developing techniques to mine large datasets. The CS community, which in the past focused on collecting and modeling small datasets in controlled studies, has started to face the need to deal with far larger amounts of collected data, and data that are generated in uncontrolled fashion. These facts point to opportunities for progress by combining the best forces of the two groups.

Large datasets contain an exponential explosion of correlations of all sorts, but the critical goal, if one wants to understand the processes generating the data (e.g., cognitive processes), is inferring causal mechanisms. The ML community has been developing techniques for causal inference, in the form of Bayes nets and Bayesian analysis, both often hierarchical in nature, and in the form of generative modeling. Thus far the techniques have proven immensely valuable in the analysis of relatively small datasets, and progress in causal inference from larger and larger datasets can be expected in years to come.

There is a clear linkage between large datasets and human cognition and behavior. Text databases are produced by humans and thereby reflect not only all the aspects of human language, but also every sort of human behavior such as preferences, social interaction, decision-making, learning, memory, and developed knowledge. Similar rich information about human cognition and behavior is embedded in visual materials (e.g., vacation scenes, photos of people, movies) and auditory materials (e.g., music, speech) that are also accumulating in ever-larger amounts. Data from educational settings is also likely to explode onto a massive scale as online education ramps up and digital devices record student queries and responses. If the task of finding causal mechanisms in such data is daunting, it is made even harder by the fact that the various types of information are typically linked (e.g. captioned photos, words in songs). Cognitive scientists seeking to understand the mechanisms of human cognition and behavior are just beginning to face the problem of using such massive and unorganized data, and computer scientists developing techniques to mine such data are just starting to face the need for approaches that will scale up to massive data.

Massive amounts of data also arise in controlled scientific studies, and they are starting to do so in ever-increasing numbers. Researchers in CS have begun to expand their inquiries in ways that reflect the growing realization that human behavior goes far beyond what is revealed by simple static measurements of accuracy in laboratory tasks. Cognitive scientists are starting to collect massive datasets of dynamic variables, such as eye movements, pupil dilation, head position, body posture, hand and foot movements, heart rates, skin conductance, verbal and physical indications of ongoing decision-making, and social interactions. Just a few examples of research that has started to deal with such data include analysis of eye movements in viewing of photographs, decision-making that takes into account high-order task demands, kinematics of body motion to accomplish motor tasks, and analysis of hand trajectories to indicate responses, using such techniques as Gaussian process regression, generative modeling, and deep belief networks. The techniques to deal with such data are still in their infancy and will require a two-way interaction and cooperation between scientists in the two communities.

Applying CS and ML Together to Explain the Brain

As with the other domains discussed in the previous subsection, neural data (e.g., fMRI, PET, Squids, EEG, and much more) have reached a level of complexity that requires sophisticated ML algorithms to analyze them. These data must then be linked to behavior and cognition, the data describing which is also starting to accumulate in massive quantities. On the other hand, a theoretical understanding of brain activity will likely only be achieved via the language provided by CS—one needs to interpret activity in terms of perceptual and cognitive operations and representations.

As President Obama noted in his recent address to the National Academy of Sciences when describing his BRAIN initiative, the goal of brain mapping is the understanding of behavior and cognition, learning and memory, and social interaction. Thus the scientific problem of understanding the brain is one that requires integration of CS and ML. ML can provide tools to analyze data about the brain, for example to classify brain activities from EEG or fMRI data. CS can specify the language for modeling the brain and can produce theories of brain activity for ML to test. The means for identifying the causal processes linking neural and behavioral data are still in the first stages of development and will require scientists in ML, CS, and neuroscience to work together to make progress.

CS and ML are currently having important impacts in neuroscience in ways that could naturally be combined. First, internal variables from cognitive models are correlated with brain data to infer the information-processing operations carried out by different brain regions (e.g., Fincham et al., 2002). Second, ML techniques are applied to high-dimensional brain data to characterize what information is in the signal, most commonly in multi-voxel pattern analysis (MVPA; Norman et al., 2006). A promising integration of these approaches, which would integrate CS and ML methods, is to train MVPA and related techniques to predict variables of cognitive models (rather than experimental conditions), to obtain high-dimensional characterizations of the brain activity corresponding to different cognitive functions (Mack et al., 2012).

Vision is a good example domain wherein there is a close relation between CS, ML, and knowledge of the brain. There has been a long history of biologically inspired models of vision. In particular, it is possible to map hierarchical computational models onto the hierarchical structure of the visual cortex. In turn, CS and ML theories of vision offer models of algorithms and representations that the brain might be using, and which can be tested by experiments (e.g., using fMRI or multi-electrode recording). Research in vision thus offers a guide for how CS and ML can collaborate to understand other aspects of brain function, including higher-level functions (e.g., planning, reasoning, control, problem solving).

Mechanisms of Collaboration

Despite their overlapping scope of investigation and theoretical frameworks, the CS and ML communities have many differences that can impede collaboration. Most obviously, CS is concerned with the workings of the human mind, whereas the goals of ML are methods that perform well regardless of their biological relevance. Likewise, ML is concerned primarily with performance or optimality, whereas CS often deals with aspects of cognition that are (or appear) nonoptimal.

More generally, as an engineering discipline, ML deals more with what can be done in the short term using current techniques. As a scientific discipline, CS seeks more fundamental theoretical advances that will have long-term impacts on our understanding. Nevertheless, theoretical developments can yield engineering results in the longer term. CS, and biology in general, can give theoretical inspiration to ML. Conversely, ML researchers have experience working in large datasets and have developed a powerful set of technical tools that can also be applied to cognitive modeling. The following subsections outline possible means for fostering this type of collaboration.

Shared Challenges, Databases, and Benchmarks

One goal of collaboration between CS and ML should be to refocus ML on what humans can do that machines currently cannot. The first step is to identify the tasks on which humans outperform current ML. Much of current ML works on vision, text processing, classification, and search, and it has produced an array of complex and powerful learning algorithms. Shifting the emphasis to other domains could encourage extension of these algorithms to operate on richer and more abstract representations and to perform multiple tasks with a shared knowledge base.

A challenge to such a redirection of ML is that the ML community is doing quite well under its present structure. It has been enormously successful in recent years in technological applications (e.g., search in large databases, computer vision, voice recognition). To get the attention of the ML community at large, it needs to be shown new problems where it could benefit.

A significant portion of the ML community is driven by performance on benchmark problems or datasets. ML researchers in this tradition work by incrementally beating each other on performance metrics on well-accepted databases (e.g., PASCAL or ImageNet). Even 1-2% annual increments can translate to 10-20% increments over a decade. Nevertheless, such benchmarks—if narrowly designed—can lead to a "tyranny of datasets" wherein considerable time is spent on a restricted set of tasks that may not generalize or encourage certain types of progress.

Thus, CS might best impact ML if the CS community can help construct databases, metrics, or challenge problems for comparing and testing ML algorithms that encourage development of more cognitive or humanlike methods. A core skill of cognitive psychologists is the ability to identify predictions of a model and to design experimental tests. This approach can be applied to ML systems, to identify where they will fall short and to shed light on how humans succeed in those situations. In other words, cognitive scientists can analyze existing ML models and try to determine what they could not do, particularly because they lack certain critical aspects of human cognition. This process could lead to new benchmarks that would require theoretical advances in directions currently neglected by ML, toward making artificial systems more cognitive and humanlike.

One possible strategy would be to design sequences of challenge problems that guide toward a complete cognitive model. A sequence could follow a developmental trajectory (either individual or phylogenetic), or it could be based on intuitions about the order of theoretical developments that is most likely to culminate in a major advance (although in the latter case there might be a danger of implicitly building a predetermined cognitive theory into the sequence). This type of challenge could encourage the ML community to perform increasingly complicated tasks and hence drive the field forward.

The Appendix presents some candidate domains for new cognitively inspired benchmark ML tasks. A goal of these challenge problems is to require "cognitive" as opposed to "engineered" solutions. It is important that solutions advance theory, in the sense that they lend themselves to solutions of other problems rather than being specific to the given task or dataset. Challenges inspired by CS should also encourage solutions that have the hallmark properties of cognition (e.g., flexible, structured representations). Nevertheless, engineered solutions might succeed in some cases. Such outcomes might be informative, in showing that certain aspects of human cognition that appear highly complex might in fact be based on simple mechanisms (coupled with sufficient experiential data).

Cognitive Science in Complex Domains

A complementary strategy would be to get CS models to work on bigger datasets and more-complex stimuli. This might lead to adoption of ML methods but with more cognition at the core. Moreover, the ML community might pay more attention to cognitive theories if there were more demonstrations that cognitive principles apply in complex settings.

To achieve this would require a shift of values in the CS community, which currently emphasize simplified and well-controlled experiments. Success of a model in a complex domain should be valued as evidence that it can scale up to natural cognition. Journals and funding agencies might be able to encourage such a shift.

Interdisciplinary Education and Interaction

Critical to fostering interdisciplinary collaboration is laying the social groundwork for exchange of ideas, through conferences, workshops, graduate courses, and summer schools. Some past meetings, including ones organized by participants in the present workshop, have been quite successful in this regard.

In 2005, Tenenbaum and Yuille organized a workshop on *Probabilistic Models of Cognition: The Mathematics of Mind*, hosted by the Institute of Pure and Applied Mathematics (IPAM) at UCLA. This workshop was followed by 3-week summer schools in 2007 and 2011. The goal of all three meetings was to convene experts from across CS and ML to discuss probabilistic modeling approaches and their potential to provide a unifying and rigorous theoretical framework.

In 2008, Daw, Griffiths, Tenenbaum, and Zhu organized a workshop on *Machine Learning Meets Human Learning* at the Neural Information Processing Systems (NIPS) conference in Whistler, Canada. The goal was to bring together the different communities that study ML, CS, neuroscience and educational science. The workshop sought to provide researchers with a common grounding in the study of learning, by translating different disciplines' proprietary knowledge, specialized methods, assumptions, and goals into shared terminologies and problem formulations. The workshop investigated the value of advanced ML theories and algorithms as computational models for certain human learning behaviors, including, but not limited to, the role of prior knowledge, learning from labeled and unlabeled data, and learning from active queries. Finally, the workshop explored insights from the cognitive study of human learning to inspire novel machine learning theories and algorithms.

In 2011, Oliva and Yuille organized a workshop on *Frontiers in Computer Vision*, sponsored by NSF and ARL, which convened prominent researchers to explore future directions of research in vision (www.frontiersincomputervision.com). Topics of the workshop included the relationship of computer vision to biological vision, and collaborations between humans and machines on vision tasks.

Mozer et al. hosted a 2012 NIPS workshop on *Personalizing Education with Machine Learning*. There was also a tutorial on educational issues at NIPS by Brunskill and Gordon, another sign of interest and acknowledgement in the ML community.

Perhaps the most effective of these categories would be summer schools, which can capture the interest of graduate students early in their careers before they become more specialized. Interdisciplinary graduate courses teaching the technical skills of ML and CS would serve a similar purpose. A great proportion of researchers in computational CS and in ML begin with interdisciplinary ambitions spanning the two fields, but these ambitions are too often set aside as practical career constraints set in. Early exposure to cross-disciplinary work could show students that this is indeed an option and give them a common language for communicating between the fields. Hopefully, the momentum built by the current workshop can precipitate such a summer school or other meetings.

Funding

NSF currently lacks an appropriate funding program that reflects the interests of computational cognition. CISE Robust Intelligence RI-Large is a collaborative mechanism, not suitable for single investigators. The INSPIRE mechanism may not be suitable because it is aimed at transformative research that has not been tried before, and in this case there is already a critical mass of people and research bridging ML and CS. More immediately, the goal described above of building cognitively inspired metrics and databases for ML could be advanced by a research infrastructure grant.

Conclusions

The fields of cognitive science and machine learning, having been split for many years, are moving into a period of greater interaction and synergy. Future collaboration between CS and ML could open up new theoretical territory and produce major breakthroughs relevant to technology and society at large.

CS and ML are united by the shared goal of developing a computational understanding of human-level perception and cognition. However, they have tended to focus on different aspects of this problem—ML emphasizing powerful statistical algorithms that scale to complex stimuli or tasks, and CS emphasizing structured and flexible representations that can apply to multiple tasks. The strengths of the two fields are thus naturally complementary.

The primary challenges to collaboration may lie in the pragmatics of how the two fields work. What counts as a useful advance in CS is not necessarily considered so in ML. For cognitive ideas to have a greater impact in ML might require demonstrating that they can provide quantitative improvements on objective benchmarks or that they suggest new benchmarks that can push theory forward. The two fields' infrastructures for training and dissemination are also largely separate. Future funding for conferences, workshops, graduate courses, and summer schools might go a long way toward creating a new generation of interdisciplinary researchers.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111,* 1036-1060.

Castro, R., Kalish, C., Nowak, R., Qian, R., Rogers, T., & Zhu, X. (2008). Human active learning. *Advances in Neural Information Processing Systems (NIPS), 22.*

Fincham, J. M., Carter, C. S., van Veen, V., Stenger, V. A., & Anderson, J. R. (2002). Neural mechanisms of planning: A computational analysis using event-related fMRI. *Proceedings of the National Academy of Sciences, 99*, 3346-3351.

Gibson, B. R., Rogers, T. T., & Zhu, X. (2013). Human semi-supervised learning. *Topics in Cognitive Science, 5*, 132-172.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Science, 10*, 14-23.

Heider, F., & Simmel, M. (1944). An experimental study in apparent behavior. *The American Journal of Psychology, 57*, 243-259.

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review, 116*, 20-58.

Khan, F., Zhu, X., & Mutlu, B. (2011). How do humans teach: On curriculum learning and teaching dimension. *Advances in Neural Information Processing Systems (NIPS), 25*.

Laird, J. E. (2012). *The Soar Cognitive Architecture*. Cambridge, MA: MIT Press.

Langford, J., & Shawe-Taylor, J. (2002). PAC-Bayes and margins. *Advances in Neural Information Processing Systems (NIPS), 15*.

Mack, M. L., Preston, A. R., & Love, B. C. (October 2012). Model selection using multivariate measures of brain response: Exemplar vs. prototype theories of categorization. *Talk presented at Society for Neuroscience Annual Meeting* (New Orleans, LA).

McAllester, D. (2003). PAC-Bayesian stochastic model selection. *Machine Learning, 5*, 5-21.

Norman, K. A., Polyn, S. M., Detre, G. J., Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Science, 10*, 424-430.

Russell, S., & Wefald, E. (1991). *Do the Right Thing: Studies in Limited Rationality*. Cambridge, MA: MIT Press.

Salakhutdinov, R., Tenenbaum, J., & Torralba, A. (2012). One-shot learning with a hierarchical nonparametric Bayesian model. *Journal of Machine Learning Research: Workshop and Conference Proceedings, 27 (Workshop on Unsupervised and Transfer Learning)*, 195-207.

Settles, B. (2012). Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning, 6*(1), 1-114.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Yarbus, A. L. (1967). *Eye Movements and Vision*. New York: Plenum Press.

Zhu, X., & Goldberg, A. B. (2009). Introduction to semi-supervised learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning, 3*(1), 1-130.

Zhu, X., Rogers, T. T, & Gibson, B. (2009). Human Rademacher complexity. *Advances in Neural Information Processing Systems (NIPS), 23*.

Appendix: Candidates for Cognitively Inspired ML Benchmark Tasks

1. <u>Compositional image search</u>. Current image search systems, such as Google Images, can search on simple conjunctions (e.g., *black dog*), but they cannot handle queries with real relational structure (e.g., *three big black dogs sitting on a box*).

2. <u>Flexible interaction with a structured database</u>. ML systems should be pushed beyond searching, to interpreting. Promising approaches might be to make Google's Knowledge Graph more cognitive, or to build from Percy Liang's question answering system. Queryable vision systems for the blind would be one application. A sequence of challenge tasks could start with simple questions (*Is there an X?*) and increase in complexity (*Where is the X? What is the X doing?*). Similarly, questions could start without context, followed by questions that require use of context or the global structure of a scene.

3. <u>Nonverbal social perception and sentiment analysis</u>. Stimuli could be videos in the style of Heider & Simmel (1944), brief clips of silent movies, or videos of natural social interactions (e.g., at a restaurant table, airplane row, park bench, or business meeting). Questions could be taken from psychological studies of humans' interpretations of complex scenes (Yarbus, 1967), such as the following:

What does X want?
How old is X?
How rich is X?
How educated is X?
How nice is X?
Does X trust Y?
Does X like Y?
Is X afraid of Y?
Is X convincing Y?
Does X work for a major international corporation?

4. Open-mind tasks, with true/false questions. Performance could be compared between systems using n-gram and knowledge-graph style representations.

5. SAT or GRE reading comprehension questions. The challenge could also be to summarize a text or to answer questions such as *Who did what to whom?*

6. IQ test problems.

7. Picture caption evaluation. The challenge would be to build an integrated representation of the image and text. Tasks could include determining whether the caption is appropriate, or perhaps whether it is humorous.

8. Recognizing affordances in the environment. There would be no predetermined task goal. Rather, the task is to recognize what useful things could be done in a given situation.

9. Robotics and games. Challenges could be built from multi-agent video games or from children's playground games, requiring robotic systems to implement a form of social cognition or theory of mind.