Artificial Intelligence in the Open World

Eric Horvitz

Presidential Address

July 2008

Chicago Illinois

Opening Session of the Annual Meeting

Association for the Advancement of Artificial Intelligence

Introduction by Alan Mackworth (Immediate Past President of AAAI)

Well good morning and welcome to AAAI-08, the Opening Session. My name is Alan Mackworth, and it is my pleasant duty to introduce Eric Horvitz. Eric is the current President of AAAI and now he will present his Presidential Address. Eric is well known in the field both for his scholarship and his leadership so I really don't need to say too much. But I'll start with just a few facts. Eric started in the joint PhD MD program in neurobiology at Stanford. But he saw the light and decided not neurobiology but computational theory mind was the secret to the universe, so he switched in his second year to studying that. In fact, he switched to computer science and decision theory, but still graduated from the combined PhD MD program. So, not only is he Dr. Horvitz, but he is actually Dr. Dr. Horvitz. And he really likes to be so addressed, so I encourage you to talk to him that way. He finished the PhD in 1991. He finished his MD in 1994. He was on staff at the Rockwell Science Center from 1991 to 1993 and involved with David Heckerman in a startup then that was acquired by Microsoft. So, since 1993 at Microsoft Research, he has been very active. Now, he is Principal Researcher and Research Area Manager. He's been enormously influential in establishing AI as a major component of the MSR activity. Perhaps his most important research contribution has been to build the links between AI and decision science. Eric was and is and will be a major player in the probabilist revolution that has swept Artificial Intelligence. Thanks in part to Eric's work, decision-theoretic concepts now pervade AI. For example, in his doctoral work, he coined the concept of bounded optimality, a decision-theoretic approach to bounded rationality. Through subsequent work at MSR, he has played a major role in establishing the credibility of Artificial Intelligence with other areas of computer science and computer engineering, including all of the great work he's done linking AI and HCI and even work in operating systems, working on caching. So I can't begin to summarize the rest of his research contributions, but I'm sure and I hope that Eric will do some of that in his talk. And throughout all of this activity, he's played a major service role. As Ron Brachman just emphasized in accepting his Distinguished Service award, if you're a citizen of a scientific community, you have a duty to help serve and play leadership roles in that community, and Eric has taken that lesson more than almost anyone else. He served not just in AAAI, but in UAI, the ACM, IEEE, DARPA, NSF, you name it. So, as an organization, we've been very lucky to have landed Eric as our President. So please join me in welcoming Eric Horvitz.

Presidential Address by Eric Horvitz

Thank you for the kind words, Alan. Good morning.

It's an honor to be here today with colleagues, all sharing an interest in advancing the understanding of the computational mechanisms underlying thought and intelligent behavior and their embodiment in machines. We're part of a rich and fascinating intellectual history of people who have wondered about the nature of mind over the centuries and, more particularly, those whom have been optimistic that we can learn new insights about thinking—that we can and likely will fathom the machinery, the principles underlying thinking and intelligent behavior. Hundreds of years of thinking in reflection, but more recently in the 18th century from de la Mettrie on to Babbage, on to Alan Turing, John von Neumann, Nobert Weiner and on to Alan Newell, Herb Simon, and the larger upswing in interest over the last 75 years.

Looking back, the excitement about possibilities was almost palpable in the 1940's when several founders of modern AI were enlivened by work on the theory of computability and by thinking that led to designs for general purpose computers. Here's John von Neumann, looking very proud in front of the EDVAC, one of the first machines to have a stored program. Remarkably, EDVAC ran quite reliably until 1961. von Neumann was passionate about the computational basis of thinking. He also pioneered *utility theory* and *action under uncertainty* in a collaboration with Oscar Morgenstern—a set of ideas that Herb Simon would later call one of the major intellectual achievements of the 20th century. von Neumann and Morgenstern give a formally axiomitized statement of what it would mean for an agent to behave in a consistent rational manner. It assumed that a decision-maker possessed a utility function and ordering by preference among all the possible outcomes of choice, that the alternatives among which choice could be made were known and that the consequences of choosing each alternative could be ascertained via a consideration of a probable distribution over outcomes. Tragically in 1955, just a few years after this picture was taken, John von Neumann was stricken with cancer. He struggled to complete his Silliman lectures and associated manuscripts on *Computers and the Brain,* but that was left incomplete.

Six months before John von Neumann's death, in the summer of 1956, a defining summer study was held at Dartmouth University, led by John McCarthy and attended by an extraordinary set of folks we now know well. The proposal and meeting were the place where the phrase *artificial intelligence* was first used. John McCarthy has mentioned that the proposal and meeting was put together to, "put the flag on the pole," about the lofty goals of the scientific endeavor of AI—and they were laudable. The proposal for the meeting is remarkable in its modernity, and could almost describe current research foci, seeming to raise the level of abstraction of thinking away from the kinds of work that John von Neumann and others had been pursuing on optimization and action under uncertainty: "Machine methods for forming abstractions from sensory and other data," "carrying out activities, which may be best described as self-improvement," "manipulating words according to the rules of reasoning and rules of conjecture," "developing a theory of complexity for various aspects of intelligence." A new paradigm was forming, branching AI away from the more numerical decision sciences and operations research into a world view that included valuable and novel focuses on high-level symbols, logical inference, and cognitive psychology as a source of insights and inspiration. Computers were used to explore heuristic models of human cognition as well as well-defined, structured, closed-world puzzle-like problems—game playing and theorem proving, including programs that could carry out proofs for theorems in Euclidian geometry.

It was during this time that Herb Simon communicated some of his thinking about the challenges of intelligent decision making in open worlds. Over the years, as I've wrestled myself with building decision making systems, thinking about probabilistic and logic-based approaches that make decisions under limited time, I've often thought about Simon's fanciful image of the ant interacting with the complex environment at the beach, wandering among the hills and nooks and crannies, in thinking about building computational intelligences. In Simon's parable, the ant's actions and cognition are relatively simple in comparison to the complexities of the surrounding world and the creature is buffeted by all the complexity, by the nooks and crannies of the world. So Simon posed a vision of intelligence as hopelessly bounded rationality, as satisficing rather than optimizing, given incomplete knowledge about one's own preferences, the state of the world, and outcomes of action. And this has been an enduring perspective and a valuable one. Our agents try to do their best immersed in complex universes with limited representations and limited time and memory to form computations so as to sense, reason, and act in the open world.

To date, our agents have largely been closed-world reasoners—even when it is clear that problem solving, and intelligence more generally, must wrestle with a larger, more complex world, an open world that extends beyond the representations of our artifacts. In formal logic, the open-world assumption is the assumption that the truth value of a statement is independent of whether or not it is known by any single observer or agent to be true. I use *open world* more broadly to refer to models of machinery that incorporate implicit or explicit machinery for representing and grappling with assumed incompleteness in representations, not just in truth values. Such incompleteness is common and is to be assumed when an agent is immersed in a complex dynamic universe. I also allude to the open world outside the closed world of our laboratories, where AI is pressed into real service, working with realistic streams of problem instances.

I'd like today to touch on a few key technical challenges that I see in moving toward open-world reasoning, and, more generally, *open-world AI*. I'll touch on directions with moving AI systems into the world where applications interact with the complexity of real-world settings. Then, I'll share reflections about our endeavor, the nature of our research in the open world, the larger community of scientists working collaboratively on the grand pursuit of an understanding of intelligence.

In the 1950's and early 60's, rich and very interesting work in AI blossomed, and logicists started thinking about placing theorem provers into the real world—into the open world. Perhaps making the closed-world assumption would be safe. What is not known to be true is false. The inadequacies of such an approach was readily seen, and thinking blossomed about extensions to logic heading into the 70's and 80's, as well as defining some hard problems that would likely have to be reckoned with as logical reasoners would be "stepping out" into the open world, knowing full well that they would have a very incomplete understanding of that world. The *frame problem* was defined along with the *qualification* and *ramification* problems. Logical reasoners seemed almost paralyzed in a formal sense. How could agents reason and take action if they didn't know what the scope of relevance and attention was? A bunch of work—and work to this day—ensues on the various approaches to relevance and focus of attention. Only certain properties of a situation are relevant in the context of any given situation or situation and action coupled together, and consideration of the action's consequences can and should be conveniently confined to only those relevant distinctions about the world and about thinking. Work today, and over the last several decades, includes non-monotonic logics where updates to assumptions are made in response to observations and such approaches as circumscription, which seek to formalize the commonsense assumption that things are as expected unless otherwise noted.

Some fabulous work in the 70's and 80's focused on pooling large quantities of interrelated expert heuristics, where rules that would be chained backward to goals and forward from observations. If our systems were complete, we might continue to fill them with human expertise, and make them more complete that way. But when such systems were used in some real world domains like medicine, the need for managing uncertainly came to the fore. I and a number of my colleagues during my graduate work at the time, wrestled with the use of such knowledge-based systems in time critical domains. We started to dive more deeply into the more fundamental probabilistic representations that captured uncertainty, for example, in a deadline or in a problem state. A community of like-minded folks looking back at the work of von Neumann and others started to form, coming to be known as the *uncertainty in AI community* or UAI. These ideas were also spread into other communities as well, into several subdisciplines of artificial intelligence. The early days of UAI were very exciting times of synthesis, looking back and looking forward. In many ways the thinking didn't necessarily start anew, but instead built upon some of the really interesting core work and contributions coming out of the earlier efforts and research in artificial intelligence. Soon the full brunt of reasoning about action, decision, reflection, going back to the early 20th century, came to the fore, to be used and applied in hard AI challenge problems.

Now uncertainty became an organizing principle in some communities doing research in this space. Incompleteness is inescapable, uncertainty is ubiquitous—uncertainty in the state of the world, the outcome of action, in problem solving processes themselves. The core idea is to push unknown and unrepresented details into probabilities and then to propagate them in a coherent manor. So at the core of some of this effort was machinery, and designing machinery and reflecting about machinery, for handling uncertainty and research limitations as being foundational in intelligence.

Now, as many people know, representations such as graphical models of various kinds came in this work, efforts including the Bayesian network or belief network representation. Judea Pearl introduced *d*-separation, providing a sound and complete algorithm for identifying all independencies entailed by these graphs. Here's a small probabilistic model that might explain why a car won't start, given observations about fuel, for example, and the status of hearing the sound from the turning over of the engine. Quite a few models were also created that could do decision theoretic inference directly, outputting ideal action and the expected utility of that action. A world state would emit evidential patterns so that actions might be considered. The utility model, capturing an objective function and goals, was also admitted into the decision model, and some evidence might be observed in advance of a decision. In the end, the expected utility of action depended on both the action taken and that hidden state of the world.

Soon people were thinking about more generalized decision problems extending over time where a sequence of actions would have to be computed over time, the world state would be evolving over time, and there was some utility function either globally or acutely. This was the decision-theoretic perspective on planning—very complicated problems. Work at the University of Washington, Toronto, Vancouver, MIT, Berkeley, and Stanford focused on various ways to decompose this problem and simplify it by factoring it into simpler problems, by abstracting it into higher-level actions and state spaces and so on, and this work continues to this day.

At the same time, parallel efforts were accelerating in machine learning. Discovering structure in concepts, in particular in the Uncertainty in AI community, there was quite an interesting set of efforts that continue to this day on discovering structure—actually building graphical models from data, more data. The basic idea is to apply heuristic search to reason about the likelihood of different models actually applying likelihoods themselves to structure and identifying the best model. What was very exciting about this work is that, not only could we identify the best models given our approach, but we also could actually reason about the potential existence of unknown variables, hidden variables. So, for example, we would know that there could be a hidden variable upstream of variable A and B—variable C—that was likely affecting both variables—to infer that hidden variable, thinking out of the box.

Now rather than resource bottlenecks being scary things during this time they became interesting sources of reflection about mechanisms of intelligent thinking and behavior, where interesting insights were arising in these types of situations. In my own dissertation work and beyond, I sought to move beyond traditional studies of bounded rationality—often associated with shaking one's head and turning to heuristics—by pursuing principles of intelligence in perception, learning, and decision making amidst incompleteness, uncertainty, and resource limitations—a pursuit of what I termed *bounded optimality.* Could we actually build systems that could do their best under constraints of resource in time and memory? What insights might we learn about intelligence by pursuing the optimization of thinking processes under resource constraints? These questions framed the potential value of taking an economic perspective on computational problem solving, leading to research on flexible or anytime procedures, as well as on principles of metareasoning, on how portions of reasoning resources might be ideally allocated to reflection about problem solving.

As we took our complicated, *NP*-hard algorithms to the real world, like healthcare, we started thinking: might there be a way to do incremental refinement rather than waiting for a final answer to an inferential problem? Methods that could actually generate increasingly valuable, or increasingly complete results over time were developed. I had called these procedures *flexible computations*, later to also become known as anytime algorithms. The basic idea was if we had a way to refine the value of the output of thinking, we'd have a more robust approach to uncertainty in our deadlines about when to stop. For example, a flexible procedure for generating well-characterized partial results within a medical decision support system might provide some value to a patient by enhancing a physician's decisions, rather than providing nothing helpful should a deadline come before the output of a computation is completed. In taking an economic perspective on inference under time constraints, we can build systems that consider the cost and the benefits of computation over time, and compute a net expected value of computation and an ideal stopping time that would be a function of, for example, the cost of delay in a particular context.

We also began to wrestle with the uncertainty in the output of computation. The output of computational processes could be as uncertain as the world we faced—both in the cost and the value of computing—leading to notions of the *expected value of computation*, computing how much it would be worth to think longer in a setting—and not just how much longer to think, but also, what was the next best computational action?

The expected value of computation provided a very nice framework for deliberating about reflection. Our systems now could look in a formal way at problem solving tasks as well as about the thinking process itself, and consider both in synchrony. Notions of the partition of resources—how much of the limited time for a computation in the real world should be applied to the base level versus to the metalevel, to the reasoning about the reasoning, to optimize what was happening at the base level—pointed to notions of the ideal partition of resources and a formal approach—decision theoretic control—to metareasoning. It's exciting to build tractable polynomial time metalevel reasoners that actually could control more complicated base-level domain inference, and to do this in such areas as time-critical medical decision making. Here's a situation where a patient is wrestling with an unknown respiratory ailment and the system here is helping a physician in an emergency room to figure out what's going on. The bounds on the probabilities of a patient being in respiratory failure are being tightened over time with evidence and computation, and we actually compute the expected value of the computation and it tells us when to stop and treat the patient now rather than waiting for more computation to occur. What's interesting in this work is that we actually could compute the expected value of adding a metareasoning component to a system over a system that only reasoned at the base level, for the first time providing us with some insights about the value of a metareasoner, the value of having this extra structure in our problem solvers.

Work also went on with learning more deeply about hard problems like satisfiability, a machine learning to understand, for example, how long a computation would run given some analysis of the problem instance and actually looking at the problem solving behavior over time early on in a computation to come up with an ideal restart policy, for example, but more generally, in many fields of AI the idea of actually trying to characterize the nature of problem solving and its uncertainty particularly for run times. Now with all these methods, though, using probability to push uncertainty in a coherent way around our systems, we really didn't yet have a deeper place to open-world AI. These methods helped us to deal with uncertainty in the time available for computation and how a system would work with uncertainty about world states, but in some ways we were still a relatively closed world. The big challenge up ahead here is how to open up our systems, even probabilistic systems, to being more open-world in their approach.

So I thought I'd mention a few interesting challenge problems. Overall, I'd like to sort of characterize the challenge we have in front of us as building *situated flexible long-lived systems*. We seek methods that can provide flexible adaptations to varying and dynamic situations, given streams of problem instances over time, and challenges over different time frames, handling a broad variation of uncertainty and goals, time criticalities, and the availability of actions, while learning about new objects, predicates, goals, and preferences, and even about perception and reasoning.   Here's a depiction of our agent, immersed in the world, looking at many situations, trying to flexibly adapt to varying tasks and goals and environments. The agent faces a number of challenges. What is the current problem and what should I work on next? How do we coordinate sensing, reflection, action, and learning? A standing challenge is the refinement of methods for guiding the allocation of time and other resources via computation of the value of information, the value of computation, and the value of learning over time. We need more work at effectively interleaving multiple aspects of reasoning and harnessing them in concert for doing well and living the good life over sequences of problems over time.

Another challenge is life-long learning. How do we trade off the local costs of exploration and labeling for long-term goals? It could turn out, for example, that there's quite a bit of work up front in training up a model. For example, even in working with humans in a setting that requires a labeling effort by people, for longer-term gains and really amortizing the effort over time in a life-long way, we're considerate of those kinds of efforts over time and long-term enhancements given multiple challenges.

Another challenge is handling streams of problems over time. This includes work on policies for using all time available, including what might be called as "idle time" versus looking at our systems as solving a single problem challenge and then waiting for the next problem to arrive. We want to have the best use of time to solve all future problems and want to sometimes trade off current problem solving for future problem solving. Here's our robot in the world dealing with problems, as represented by this lens coming into view. Sometimes it's good to trade off the present for the future, slowing down on current work while working on future problems. What comes to mind is the vision of a trauma surgeon who's just finishing up with a patient, putting the last sutures in, when all of a sudden his attention moves to speakers in the room, where he hears that there's an ambulance on the way in, carrying several patients and he's listening to the nurse giving him a description of the situation, and he's slowing down what he's currently doing with the patient at hand to prepare for the forthcoming task, perhaps planning, asking via the intercom, "I want to set up operating room three, two, … I'll need these kinds of tools and so on." So he's slowing down what he's doing now and trading it off for preparing for the future.

A core challenge is the frame and framing problem. What goals, preferences, objects, predicates, relationships should be in a decision model? How can we build tractable, relevant models automatically and how can the system learn more about the frame? In this dream sequence of context-sensitive framing that I've often referred to, is a vision of what we might have operating some day. I actually used these slides in a 1996 panel at AAAI-96 in Portland on big challenge problems for AI and I still carry these few slides around. The basic idea is we'd like to somehow figure out what an appropriate goal is at any moment in time or sequence of moments and back chain the goal into relevant distinctions from a large fund of knowledge. We wish to have knowledge and computational machinery that selects those distinctions and wires them into that model appropriately, chaining through dependencies so as to build a propositional model that applies to the situation at hand. And then of course we wish to do this over time and reason about long-term, as well as acute value of utility.

Some interesting work in this space has come at the synthesis in a section of propositional probabilistic representations and first order logic representations. It's been kind of a remarkable time of creative effort on representations in this space. At the first UAI conference in 1985, Jack Breese presented some interesting work that meshed first order knowledge bases together with propositional probabilistic representations. Other people working in this space included Mike Wellman, Robert Goldman, Eugene Charniak, and Kathy Laskey. The basic idea was that a theorem prover would take a situation from evidence, and generate a propositional model that did uncertain reasoning and decision-making, and then provide inference and best actions over time. Unfortunately, the knowledge bases were handcrafted and the method was not tested in a realm where queries were assigned autonomously while an agent was immersed and active in an environment.

Over the last eight to ten years, there's been a very interesting and remarkable effort to combine learning and first-order representations, first order of probabilistic representations, creating representations that are more amenable to open environments, to learning in real time, and so on. These include plan recognition networks, probabilistic relational models, Markov logic networks, and probabilistic models with unknown objects, BLOG. BLOG is an approach to representing and reasoning about the existence of unknown objects and their numbers. Let me bring up a little radar screen here to capture the notion. Imagine a vigilant agent watching a radar scope; it might be a little bit noisy, and here's a blip on that scope, and another blip. How many objects is that? It might be one, or is it three? And this representation helps us to reason about what's going on in the world given potential for unknown objects.

In other work that we've been doing, taking large amounts of GPS data and predicting, computing a probability distribution of where someone is next traveling, turns out that we like to reason about, after several weeks of observation, that someone's going to a new location. So there's a rich knowledge base level here where we can learn the nuances of the relationship between driving to a known destination and to a new destination that gets into the notions of efficiencies in traveling toward different locations and so on. So, the idea of reasoning about the unknown can be a rich knowledge based learning problem.

Work has also been going on in multiple teams on extending incomplete models with the learning of probabilistic planning rules. In some very nice research on learning symbolic models of stochastic domains, Hanna Pasula, Luke Zettlemoyer, and Leslie Pack Kaelbling attack the problem of a planner immersed in a messy world. Having knowingly incomplete information about the result of actions, this is a messy blocks world, where things can slip out of an agents' grasp and where piles of blocks may fall over at any time. Look at the simulator that captures some of the messiness that an agent might encounter in a more realistic world. Now, messiness is a way that agents with limited knowledge might see a larger universe. Things just aren't as clean and closed world as the old blocks world was. Let's be careful there because that might fall and so this simulator actually has notions of friction and some physics that captures the kind of a world that might be a little messier than the clean worlds of just abstract blocks and stacks. Oops there we go; oh well, we lost that stack there.

So in interacting with the world and making observations about the situation with local reference to objects, the system induces new knowledge, expanding its competency. Here are examples of two learned rules, two learned probabilistic rules. The first one captures the notion that when the empty gripper is asked to pick up a very small block X that sits on top of another block Y, that the gripper may erroneously grab both blocks with high probability. The second rule here applies when the gripper is asked to put his content Z on a block X, which is inside a stack topped by a small block Y. The work captures knowledge that has been learned -- that placing things on top of a small block is risky, that there's a reasonable probability that Z will fall to the table and a small probability that Y will follow in the Humpty Dumpty outcome of the attempt to stack Z on top.

There's also been very nice work on extending the perceptual proficiencies of our systems in the open world. It's critical for systems immersed in the real world to learn to perceive, to recognize known and unknown objects, and to learn to understand that objects can look quite different in different settings. Learning to perceive in an open world is a challenging and rich problem area. As an example of some work in this realm, Gal Elidan, Geremy Heitz, and Daphne Koller have explored the use of landmarks in canonical objects to create a flexible way to recognize variations of key objects in the world under many different circumstances. The work highlights the power of extending recognition processes into the open world by learning to recognize deformable prototypes at a high level of abstraction.

In other work on transfer learning, the key idea is to transfer abilities and skills from competency in one task to another. It's a critical challenge, and Rich Caruana, Sebastian Thrun, and many others have explored transfer learning in the context of real world settings -- the application of models learned in one situation or setting to related settings. In some recent work done by Ellen Klingbeil, Ashutosh Saxena, and Andrew Ng, a robot is trained to recognize doors and to learn to use a robot motion planner to open previously unseen doors, interpreting for example their configuration, how they might swing and swivel and hinge. Let me show you a little bit about this robot in action—the Stanford STAIR robot—coming up to a door it hasn't worked with before with a different kind of handle. Here it's orienting itself, and getting a sense for where its own effector is. Some other examples… It knows how to push that door open, bump it open. This sequence includes a very nice interaction where you're actually sitting inside the office and watching that robot poke in. It's an interesting experience to be sitting in an office someday, and how it feels for a robot to come looking in at you, saying hello. Let's see where that particular sequence is here at the end of this video snippet. Here we are, you're in the room, and say hello—very cute.

So overall, the big challenge is going to be to prosper in the open world and to develop explicit machinery for prospering in the open world. Lao Tzu: "To know that you do not know is the best." We need models and machinery that understand that. This includes: modeling model competencies, limitations and extensions; context sensitive failures and successes—learning about them, predicting them, expecting them, and reasoning about how to repair them; models of anomaly and surprise. We've built some models of surprise and models of future surprise and this is a rich area for machine learning. Also, understanding the value of prototypical shapes, concepts for transferring to other application areas or between application areas and situations, notions of using analogy effectively. But most importantly we build to learn objects, predicates, preferences, goals in noisy environments over time.

Our community knows the benefit from open world challenge problems, for example the AAAI CVPR semantic robot vision challenge, where robots have to perform a really tantalizing scavenger hunt in a previously unknown indoor environment. Key words are given to these agents in advance, with a time limit in advance of the challenge, and they have to go out to the web and understand how to convert words into images and then go find these objects that are actually scattered throughout an indoor environment. We all know very well the value of the DARPA challenge problems, to have automated vehicles grapple with previously unseen tours. It's a really great challenge problem for grappling with closed versus open world models. The first DARPA challenge problem actually was interesting in that it highlighted what happens when you have an incomplete model. If you've never seen what it looks like to see a model that's closed in some ways, smoking at a distance, this is what it looks like. In this case, this is the sand storm system. It was way ahead in the first challenge and when it happened to get caught up on a berm and it didn't understand where it was or what was going on, it just sat in one place trying it's best to keep on going and ended up just burning up tires. This is what it looks like to see a closed world model smoking from a distance.

Stepping into the world in terms of new directions, we see some really interesting application areas that highlight open world challenges. These include, robust services in dynamic settings, the area, rich area, exciting area of human-computer collaboration, pursuits of integrative intelligence, and work in the sciences. Let me talk a little bit about some work we've been doing on robust services in dynamic settings. We've been looking at the interesting traffic prediction and routing challenge of understanding how to route through an entire city system given changing flows based on changing patterns of traffic. The big challenge is we typically have sensed highway systems around the nation and even the world, but all the side streets, the surface streets are typically not sensed, so a prediction challenge that we were grappling with on my team was to use the highway system, the sensed larger highway of the structure as a sensor for all side streets because, to come up with the paths that route around traffic you need to really consider the flows through all street segments, running A* or Dystra , for example, to generate the path through the whole city system.

So to do this, over five years we've collected GPS data from volunteers who just put the GPS devise in their car and drove in an ambient manner given the way they're going in general, not having the device affect their general patterns of behavior. It's a quite detailed set of data, including about 300,000 kilometers of data throughout the Seattle region. The basic idea was to apply machine learning and reasoning techniques to take the sensed highway system along with a stream of weather reports, accident reports on the highway system, major events like baseball games and football games, and so on in the area, and to basically use the data about surface streets—full information—to weave together and create a larger predictive model for what we could generalize to all street segments in the greater city region. The system considered topological notions of distances from on and off ramps for surface streets and their relationship to the highway system. It also included notions about the properties of these surface streets. How many lines? How were they divided? By a concrete divider? Resources near by: Was there a bank, a mall, or farmland near a surface street and so on, and what its distance is from that street. The basic idea was to develop a system that could predict all velocities, do its best at generating road speeds for all street segments in city areas and then to use that for routing, applying a routing algorithm with those world weights.

Let's look at the *ClearFlow* system, which was fielded in April to 72 cities throughout North America, where every few minutes road speeds are being assigned across North America to 60 million street segments and being used in the routing service available to the public. I thought I'd bring up Chicago, Chicago's a relevant place right now, to show you what the status quo system would do given a tie up on the Kennedy Expressway, in this region that's black here. That's the blue route that basically says, considering the highway system as being backed up, I want to look at the surface streets and consider them at posted speeds. And most all traffic routing services now consider those surface streets at running at posted speeds. ClearFlow has a sense for pulling away from the highway system. It understands implicitly how diffusion might work off the off ramps and on ramps, so it provides an alternate route here using its road weights based on the predictive model about how highways interact with surface streets learned from data.

Now the current system considers that trips happen very quickly in that road speeds don't change during a route, but more generally we need to basically apply temporal models, and we're doing this in our laboratory, that forecast future speeds and uncertainties. So the idea is by the time I get to a street segment downstream in my recommended route, that road speed will be at a potentially different velocity, and depending on the uncertainties and variances along the way, I'll get to that road speed at different times. So we end up with a very interesting path planning problem while the ground is moving, a dynamic problem, so we have the opportunity to develop contingent plans. The basic idea is: I want to generate paths that allow an observer to observe and react and have an alternate flexible path they can go on that might say, for example, "When you get near this freeway, if things are slowing down take this route instead." And in the process of the search, the planning problem here, to actually reason about the accessibility of these alternate plans in advance of adjacent observation. This work is akin to work going on in several labs right now, including work by Christian Fritz on efforts to do planning in dynamic environments.

Another rich area for open world reasoning is in the realm of human-computer collaboration (HCI). There's a great deal of excitement right now in the HCI area about the applications of artificial intelligence principles and methodologies to HCI challenge problems. In many cases, it's applying well understood methods to problems in human-computer interaction, but in other cases there's some actual innovation going on in this realm that's actually leading to new AI methods and insights. One difficult and pretty interesting problem in HCI is the challenge of grounding, converging on shared references, beliefs, and intentions. This work has been studied in the realm of the psychology of conversation, but also in work on human-computer interaction. The basic idea is, in this case for example, the human is trying to communicate to a computer the need for assistance about a recent car accident. Here's some thinking going on, and an utterance is generated. It might be interpreted by that computer system with a basic knowledge. There might be some dialog to resolve uncertainty, but eventually there's some common ground reached and the system actually has a sense, potentially through sensing as well as listening, what's going on that puts the computer and the human on the same ground.

Now there's been interesting work on the grounding of beliefs in human computer interaction both for controlling displays and reasoning about when a person might be surprised. The reasoning system has a model of expectation in the world, for example, of a process, let's say, of traffic flows. It also has a model of what a human being might expect based on what it's learned from how people perceive and act. And by using estimations of the probability of distribution that might capture human beliefs and what it might believe to be our gold standard beliefs, knowing more about a situation, doing a deeper analysis, for example of traffic patterns, it might begin to predict when someone might be surprised by a current situation or understand which information would best benefit that human being to help debias, for example, a bias in judgment under uncertainty. There's been a rich area of study in the psychology of judgment about biases in judgment and in decision-making, and the idea of having systems that help to debias is very powerful and interesting.

Once we have common ground, we can reason about another interesting challenge and opportunity, which is mixed-initiative collaboration. Having systems that reason about the contributions for machine and human to jointly solve a problem together. It's a very interesting challenge problem in the general case where a problem here, represented as this blue blob, must be recognized, potentially decomposed, in this case into alpha and beta, into a problem that a machine can solve well and one that might be better left to the human being, to communicate that decomposition and the intent, and to go ahead and solve these problems together and jointly, either in sequence or in a more fluid, interleaved manner.   Research in this space, includes work on systems that schedule appointments from free content in email messages, that captures notions of decision-theoretic foundations of mixed-initiative collaborations. It's a very exciting area and a very open space for innovation.

In a related area, on *complementary computing*, we consider systems of people and computation together and think about ideal policies for coordinating these systems. Now, as an example, people that call into Microsoft Corporation work with an automated spoken dialog system that tries to work with them and figure out who they're trying to reach and so on. This can be a frustrating experience at times and people often will push zero to get right to the human operator. In general, we have a bounded number of operators who might get backed up and there might be a queue to actually get their attention, and the system is trying its best to help the caller in an automated manner. With complementary computing, we've created policies that we learn from data, understanding that at any point in the discussion, based on details and nuances of both how the speaker condition has been going and the turn-taking, about the ultimate outcome of that interaction with the spoken dialog system—and the time to get me to that outcome. Now compare that to the time of waiting in a queue to see the operator at the moment and we actually have a decision-theoretic policy that optimizes when the person will be transferred, so when things are very busy the automatic system might try harder to help the person out. When things are getting lighter on the human side, they'll be an earlier transition, especially given an inference that there will be a long-term interaction or a frustrating session ahead.

What's kind of interesting is that in this work we can start thinking about systems that actually adapt to changing competencies so our dialog system might be learning over time for example, it's competency might be becoming enhanced over time, the load on a staff might be changing—you might have more or less employees for example doing automated reception—and we have a system that's complimentary computing policy that understands how to ideally weave together the human and computing resources given changing competencies and resource availabilities. There's also the notion of a task market someday thinking more broadly about the prospect someday, that human and machine resources—sensing resources, effecting resources, robotic components—might be all available through networks' advertising and having planners that know how to assemble these pieces into plans that can actually be executed to resolve overall solutions. It's a very interesting space for future innovation. It's also a great opportunity to augment the abilities of human beings in their remembering, attending, judgment and so on, to augment human cognition, both for people who are healthy as well as people who might be more challenged, such as those facing degenerative conditions.

Twentieth century psychology is best characterized as coming up with understandings of the limitations and bottlenecks in cognition. As an example, many of you are familiar with the results of George Miller , on results with studies of recall where he found that people can hold about seven, plus or minus two chunks in memory at one time. Today we have access to bodies of fabulous work in cognitive psychology spanning different specialties of cognitive psychology, including work in attention, memory, learning, and judgment. This little jellyfish-shaped schematic shows reasoning efficiencies on the *y*-axis. Across the x-axis, I depict with sets of nooks and crannies, the biases, bottlenecks, and limitations of human cognition, some of which have been identified and characterized in cognitive psychology. There's great promise in constructing machine perception, learning, and reasoning methods that extend our abilities by harnessing an explicit understanding of the nooks and crannies in human cognition. There have been efforts aimed at achieving these goals in the realms of attention, and learning, and memory. Here is one example of work, done as a collaboration with Ece Kamar from Harvard, during her internship at Microsoft Research. The idea has been to construct and use simultaneously several predictive models, including a model of forgetting--predicting what someone will likely forget; a model of the cost of interruption, that predicts someone's current cognitive workload and workload over time; and a model of the context-sensitive relevance of the information that may have been forgotten---the value of being reminded about something. We can actually learn rich predictive models from data about the kinds of things we might forget in a setting—a model of the context-sensitive relevance, value of knowing, and the cost of interruption. These Bayesian models can be used in a symphony to control the emission of reminders and their scheduling so that they come at ideal times.

Moving beyond human-computer interaction, another interesting and challenging area that is coming to the fore these days for catalyzing work in open world AI is *integrative intelligence*. Work in this realm focuses on opportunities to build more comprehensive intelligences that can be successful in the open world by composing multiple competencies together into a symphony of sensing, learning, and reasoning. We seek to weave together several methods, both in doing new theoretical work that bridges lines of research that have been traditionally separate, and in weaving together set of components that have typically been developed separately, largely independently, and in a vacuum from one another—such as components from natural language processing, planning, vision, robot motion and manipulation, localization, speech recognition, and so on. An example of an effort in integrative intelligence, is work at Stanford by Andrew Ng and his team on pulling together vision, manipulation, navigation, and learning in robotic applications, and to study how these components work in concert with one another. Here's an example of some work with the STAIR system here, coming up to assist someone with a stapler. The STAIR robot is going out and searching for a stapler. It hasn't seen this particular model in the past, but it's trained up on what staplers look like. Future models might actually learn about staplers just by hearing the utterance and going off to web resources to try to figure out what staplers even are to begin with. This system also understands, has a model of grasping previously unseen objects that it's trained up on. Grasping is actually a remarkably challenging task for robots, robot controllers, and it's been a focus of attention of this team. The system's navigating back to the requester of that stapler. It's interesting to see how this task harnesses multiple competencies. We'll hear when that goal is reached--with a snap, and that's that.

Another example of a challenging integrative intelligence project is the Situated Interaction project on my team at Microsoft Research, where multiple components are drawn together--including machine perception, learning, and decision making-- with the goal of endowing a system with abilities to converse and collaborate with multiple people in open-world settings. One of the challenge applications is the *Receptionist* aimed at creating a system that can perform the tasks of a receptionist in a corporate environment, such as the receptionists that work in the lobby of buildings at Microsoft, where they field many questions and challenges that people might have as they enter the building. Task include gaining entry, contacting people, getting shuttles arranged for them to travel to different buildings on campus, and so on. We've woven together multiple components, including a model of user frustration and task time, dialog management and planning, behavioral controls, and conversational scene analysis. This is actually the computational view of a situation where there is reasoning about the role, goals, and relationships of each person recognized in a scene, noting for example, this person is engaging right now, this person is waiting, but part of the group, and so on. In this case we control an avatar that takes on some of the tasks of the receptionist. To see how this works, I'll just play this video of a scenario here. The red dot is actually the gaze of the avatar. Gaze is a powerful signal for grounding communication and guiding engagement with people in multiparty settings, based on who is being communicated with, when someone else is noticed as waiting, and so on. As the scenario unfolds, watch the person in the back at the right. He's going to be recognized during the conversation as waiting – there, he's being recognized now. We actually are timing his wait, and employ a model of his frustration. The avatar will get ready to engage this new person walking up to be assisted. This is the start work on a set of increasingly sophisticated interactions, and includes work on systems that might one day take care of multiple tasks and optimize workflow among many peoples' needs and goals.

I thought I'd mention that the area of science itself is a very big opportunity for our work in artificial intelligence, particularly the open world work, doing scientific discovery and confirmation, learning and inference, even a triage of experimentation. Some of the work, highlighted by the Dendral work in the 1960's and 1970's, is very exciting, understanding, for example, the scientific pipeline planner, a hypothesis generated confirmation. Much of the work back then—this is Bruce Buchanan, Ed Feigenbaum, and Josh Lederberg—can be mapped in many ways to the recent work going on in learning about structure and function from large-scale datasets, for example, datasets that capture regulatory genomics. In work by Nir Friedman and Daphne Koller along with Eran Segal and others, modules are inferred, and they really help us to understand components in the operation of biological systems, various kinds of metabolic components, and DNA and RNA processing, and so on. Working in this space has really benefitted from that coalescence of the first-order probabilistic languages with the older probabilistic inference work.

I've had a long-term interest in neurobiology, as you heard from Alan earlier, and there's quite an interesting possibility to now turn our methods inward and to learn more about neurobiology, both through new kinds of probes and data collection abilities. The recent work—actually it was described in *Science* about a month and a half ago by Tom Mitchell and others on his team—is very exciting in that we're seeing some interesting links to potential insights about representation of concepts in the human nervous system, where in this case words and word lattices and relationships to other words derived from a corpus, a language corpus, are used to make predictions about words and how they actually will affect when they're heard, the activity in different foci throughout the brain. It's work worth looking at in terms of a direction and what might be done someday.

There's also a coming era of neuroinformatics where now it's becoming possible to actually see actual, individual neurons working in harmony with other neurons. This is some data collected by the Clay Reid lab, which casts some light explicitly on goings on in the cortex of a cat as different bar patterns are displayed in its visual field. We're actually taking some of this data from a variety of animals – this is actually an invertebrate here – and applying methods we're calling the computational microscope. We seek to use this data on the activity of multiple interacting neurons, in procedures that provide neurobiologists with inferences about connectivity and function. This kind of tool is likely to be an ancestor of tools that will one day serve as workbenches that enable scientists to consider inferences about the relationships among neurons, the existence of modules of neurons, and other kinds of abstractions of the nervous system.

I'll now move from principles and applications into the open world more broadly. We as scientists and practitioners have a set of responsibilities in the open world: responsibilities to provide systems and methods and insights that are applied in ways that have social value, that enhance the quality of life of individuals and society overall. These methods and techniques also have implications for privacy, democracy, and freedom. Some of the papers at the meeting coming up this week are actually focused and show us ways that these methods can be used to enhance privacy, for example, in working with the web. There's also a long-term set of opportunities and concerns about potential long-term AI futures, including potential disruptions, good and bad, that might come from the application of our methods in the world. It is likely that advances in artificial intelligence will lead to disruptions, largely on the good side. We have formulated a panel that will be starting this fall, the AAAI Presidential Panel on Long-term AI Futures, bringing together a group of dedicated people, leaders in our field, to deliberate about and reflect about concerns, long-term outcomes, and, if warranted, on potential recommendations for guiding research and on creating policies that might constrain or bias the behaviors of autonomous and semi-autonomous systems so as to address the concerns. What might we do proactively, coupled with might happen, to make sure that the long-term future for society is a good one?

Finally, a couple of comments about the AI research community itself as it steps into the open world and works in the open world. If we think about it, AI has really branched over the years into a subset of disciplines, which is a very beautiful and rewarding thing to see happen. Typically, we have communities, like the UAI community doing uncertain reasoning, the logic community, like the SAT community, communities doing machine learning and diagnosis, cognitive science, knowledge representation at the KR meetings, for example. We also have application areas that have become specialty areas in their own right – vision, speech, user modeling, intelligent user interfaces, search and retrieval. Now, what's happening is, of course, we have communities with their own fashions and aesthetics and friendships, collegial association, and sometimes disparate approaches to problems. There's been some discussion about how this fracturing is not so good for the discipline of artificial intelligence. I'd like to differ with that and suggest that we actually nurture and come to be excited about the multiple communities of effort and work going on. I'd like to allude to a classic piece about the nature of the organization written in 1947 by Herb Simon. Herb used a parable to capture the varied beliefs, intentions, and goals that different people can have while coordinating effectively with one another: Three bricklayers were asked what they were doing, and they gave three different answers. In the story, one bricklayer answered, "I'm laying bricks," the other "I'm building a wall," and the third, "I'm helping to build a great cathedral."

To extend Herb Simon's metaphor of the pursuit of a soaring cathedral to our intellectual goals, rather than being a hindrance, the diversity of our subcommunities and the different focuses of attention are a rich source of related ideas touching on different aspects of a hard challenge, and they will come together over time, where synthesis makes sense. People may have different subgoals, but there's an ongoing construction of new arches, domes, and connecting spandrels linking things together, where different ideas come together, such as new representations coming from the intersection of first-order logic and propositional probabilistic representations and inference, or in the principles of bounded optimality rising where decision theory meets bounded rationality.

So, over the past few minutes I've reflected about directions, highlighting some bricks and some arches of future cathedrals, and perhaps some missing pieces in our existing blueprints about open-world AI that will require some hard but exciting research to fill in. The soaring structures that we seek are still off in the mist. I've no doubt that our communities working together and separately will assemble them over time and that we'll have in hand answers to long-held questions about intelligence and about the nature of mind, and that on the path to such an understanding, we'll continue to create and field technologies that will enhance the quality of life for people and society overall.

What a great endeavor we’re all part of.

Thank you very much.

(Applause)