The 2015 AAAI Fall Symposium Series Reports

Nisar Ahmed, Paul Bello, Selmer Bringsjord, Micah Clark, Bradley Hayes, Andrey Kolobov, Christopher Miller, Frans Oliehoek, Frank Stein, Matthijs Spaan

■ *This article contains the reports of the AI for Human-Robot Interaction, Cognitive Assistance in Government and Public Sector Applications, Deceptive and Counter-Deceptive Machines, Self-Confidence in Autonomous Systems, and Sequential Decision Making for Intelligent Agents symposia, which were held November 12–14, 2015 in Arlington, Virginia.*

The Association for the Advancement of Artificial Intelligence presented the 2015 Fall Symposium Series on Thursday through Saturday, November 12–14, 2015, at the Westin Arlington Gateway in Arlington, Virginia. The titles of the six symposia were as follows: AI for Human-Robot Interaction, Cognitive Assistance in Government and Public Sector Applications, Deceptive and Counter-Deceptive Machines, Embedded Machine Learning, Self-Confidence in Autonomous Systems, and Sequential Decision Making for Intelligent Agents. This article contains the reports from five of the symposia.

AI for Human-Robot Interaction

Human-robot interaction (HRI) is a broad community encompassing robotics, artificial intelligence (AI), human-computer interaction (HCI), psychology, and social science. In this meeting, we sought to bring together and strengthen the subset of the HRI community that is focused on the AI challenges inherent to HRI. As a field, HRI aims to develop robots that are intelligent, autonomous, and capable of interacting with, modeling, and learning from humans — goals that are shared at the core of AAAI. While general HRI work is seen across a variety of venues, AI-HRI serves as a gathering point for the AI-focused community within HRI.

The central purpose of AI-HRI is to share the most exciting research in this area while cultivating a vibrant, interconnected research community. We built on the success of the community-building accomplished by last year's AI-HRI symposium with the introduction of a heavier emphasis on sharing cutting-edge research results and devoting more time to the presentation and discussion of current work in the field. Accordingly, AI-HRI featured 10 keynote lectures, 24 short paper presentations, and 10 long paper presentations.

A major theme of the symposium included plan understanding and negotiation between humans and robots. The contributions surrounding this theme focused on constructing systems capable of recognizing human intent and using it as a heuristic for symbolic and motion planning, performing open world reference resolution, recognizing the purpose of objects in a scene during task execution, adapting to plan breakdowns, generating precise language for task-oriented dialogue, and performing preference-based task allocation and scheduling across human-robot teams. A second major theme involved introducing autonomy to socially assistive robots, particularly in educational or therapeutic settings. Multiple contributed papers addressed issues of autonomously promoting social collaboration between children, generating effective academic curriculums for tutoring robotics, developing personalized approaches to reducing pain anxiety in children, and moderating multiparty interactions.

A diverse array of keynote speakers presented their latest work on a list of topics spanning intelligent interface design, activity recognition, learning from demonstration, motion planning, task understanding, manipulation, reinforcement learning, smart controls for medical robotics, and human-robot collaboration. Themes introduced by these keynotes spurred discussion regarding the development of proper evaluations for autonomous robot systems that interact with humans, as it is often intractable to simulate the human presence for such work and it is often infeasible to collect hundreds of samples for large-scale analysis. It became clear that participants harbor a variety of sometimes conflicting viewpoints about expectations and framing for AI-focused HRI work, but all share the same goal of broadening HRI's reach into the spotlight of each of the many fields it depends upon for fulfilling its ultimate goal: understanding and developing autonomous systems that interact with humans in meaningful and positive ways.

This report was written by Bradley Hayes who, with Matthew Gombolay served as cochairs of this symposium. The symposium organizing committee consisted of Bradley Hayes (Massachusetts Institute of Technology), Matthew C. Gombolay (Massachusetts Institute of Technology), Brenna D. Argall (Northwestern University), Bilge Mutlu (University of Wisconsin-Madison), Julie A. Shah (Massachusetts Institute of Technology), Sonia Chernova (Georgia Institute of Technolgoy), Andrea L. Thomaz (Georgia Institute of Technology), Kris Hauser (Duke University), and Brian Scassellati (Yale University). The papers contributed to the symposium were published as AAAI Press Technical Report FS-15-01, and the papers can be found in the AAAI Digital Library.

Cognitive Assistance in Government and Public Sector Applications

The concept of a cognitive assistant as a partner to help humans perform their work better dates to the early days of AI, including the writings of Douglas Engelbart and Joseph Carl Robnett (JCR) Licklider. Recent advances in AI and cognitive computing, such as IBM Watson, Deep Learning, and NLP, along with the vast increase in available data, are enabling renewed hope that we will soon be able to offer knowledge workers a partner in their efforts. Cognitive assistance in government presents opportunities and challenges — some in common with other domains, and some distinct, which we hoped to explore in this symposium.

The symposium brought together researchers from industry, government, and academe. The topics discussed covered cognitive assistants for law, intelligence analysts, cyber-security, contracting officers, health-care professionals, and office workers. The types of support considered ranged from enhancing creativity to supporting cognitively disabled individuals and those with dementia.

One major theme of papers presented at the symposium was the variety and complexity of government and public sector use cases. Because of the complex laws, regulations, processes, and procedures required by government agencies, cognitive assistants operating in this environment must be aware of and operate in compliance with all rules. These compliance requirements will vary by agency and use case. Presenters also discussed the wide variety of users that cognitive assistants might need to support — from citizens and new employees who know little, to subject matter experts to those with dementia. The

scale of the cognitive systems presented also varied from those supporting an individual such as a patient, to those supporting the public or a large call center (such as the IRS runs) that will have to operate at scale.

Another major theme, which also distinguishes government and public sector applications, is the issue of trust. Many of the cognitive assistant uses will be in mission-critical applications where lives may be at stake. Presenters discussed the need for cognitive assistants to adequately calculate and present to the user the confidence the system has in the output. There was discussion around trusting too little versus trusting too much, and how to be transparent in explaining the basis for decisions or recommendations. It was acknowledged that for the foreseeable future, there will be gaps in the trust factor.

The symposium also included two invited talks. A talk given by Jerome Pesenti (IBM Watson) focused on the developments of Watson since the Jeopardy Challenge in order to support the company's customers. Tim Estes (Digital Reasoning) described the developments at his company around knowledge representations, knowledge graphs, and their recent use of deep learning.

The symposium also included a panel discussion on workforce issues associated with use of cognitive assistants. Participants discussed how cogs affect some professions (such as eliminating the time-consuming discovery work of first year lawyers) but will create economic development in other areas (for example, personalized medicine). The symposium was useful in showcasing many different examples of cognitive assistant projects and bringing together those involved to share experiences. The participants share a common goal of developing cogs, and agreed that they would like to attend future symposia with the same focus as this one.

Frank Stein served as chair of this symposium. The organizing committee included: Chuck Howell (Mitre), Scott Kordella (Mitre), Lashon Booker (Mitre), Ed Rockover (NPS), Hamid Motahari (IBM), Murray Campbell (IBM), Jim Spohrer (IBM). The papers of the symposium were published as AAAI Press Technical Report FS-15-02, and are available in the AAAI Digital Library.

Deceptive and Counter-Deceptive Machines

This symposium was configured by its organizers, and increasingly by researchers working in deceptive and counter-deceptive machines (DCDM), to be part of a series of DCDM *n* conferences; indeed the 2015 symposium summarized herein was actually the second conference. The first was held at the University of Maryland in July of 2013 as part of the North American Computing and Philosophy (NACAP) conference. Accordingly, one of the main themes of the AAAI Deceptive and Counter-Deceptive Machines symposium was discussion about the future of the DCDM research program. Specific plans discussed (and in some cases affirmed) are beyond the scope of this brief report, but it is safe to say that another conference or symposium will happen, if for no other reasons than that malicious humans invariably exploit deception to produce the harm they produce, and that machines can aid in the unmasking of that deception. The DCDM area would therefore appear to have an active and important future. We return briefly to this future at the end of the present report.

Participants appeared to be unanimous in their affirmation of the proposition that in adversarial contexts, for instance war and espionage, at least certain forms of deception are not only permissible, but desirable. Given this, the goal to engineer deceptive machines would presumably be an attractive objective, and one that, if reached, would supply the defense and intelligence communities with technology potentially of great value to the United States (and its allies). In addition, unsurprisingly, there was agreement as well that machines able to counter deception are of great value.

Yet this agreement immediately gives rise to one of the main drivers of multiple presentations at the symposium: namely, What is deception?, where a bona fide answer to this question — given that it's being asked by those in, or at least seeking to contribute to, AI — would need to be both rigorous and computational. Presentations seeking to answer the question ranged across fascinating attempts to model forms of deception by harnessing: analogical processing; formal, logicist methods; cognitive architectures; and sophisticated NLP techniques. One clear takeaway from these talks is that deception covers a seemingly unlimited number of shades or subcategories, from lying (in a number of shades of its own), to paltering, to fraud, to deceptive talk, and more.

The symposium ended with an extremely lively panel debate, which pivoted around the following facts as context, and a key question:

> *The Facts:* Recently, the Office of General Counsel, U.S. Department of Defense, published an updated and comprehensive manual on the Laws of War, last issued in 1956. Chapter 6 covers the constraints on which weapons are prohibited lawful; and paragraph 6.5.9 is "Autonomy in Weapon Systems." The document states unequivocally that autonomous systems are mere weapons and hence cannot in and of themselves have obligations. By virtue of this fact, autonomous agents cannot be held responsible.
>
> *The Question (Q):* If deception necessarily involves norm-violation, then artificial agents cannot themselves carry out deception; hence only human beings can deceive.

Panelists articulated their position on Q, which was all it took for the fireworks to begin. Many key issues

arose, and were analyzed; we mention but one prominent one here: While AI's core business is to try to engineer artificial agents whose intelligence and autonomy ultimately approach that of humans, the vast majority of attorneys and policy analysts/policy makers apparently view the artifacts being produced as mere shallow tools (where a weapon is a type of tool), and therefore as things that could never be the bearers of obligations, including those to refrain from deceiving, and those to deceive in order to gain advantage in adversarial contexts such as war and espionage. This fertile clash will in all likelihood be one of the key drivers of DCDM III, and arguably one of the key drivers of large, society-scale debate that promises to grow in reach and intensity as AI produces ever-smarter machines.

There are of course other planned drivers of a third Deceptive and Counter-Deceptive Machines meeting, including two issues that were lightly but tantalizingly touched upon at the 2015 AAAI event: (1) What is the role and acceptability of deception in affective computing, where robots and computing machines can increasingly exploit affective pretense to manipulate humans? (This should presumably be a central issue in so-called robotherapy, where medical ethics, affect, and manipulation intersect.) (2) What is the role of machine deception specifically in cyber-security? Inevitably, we (or our digital extensions) will mislead, deceive, manipulate, and use social-engineering tricks on the machines around us. If these machines are naive or socially ignorant, then these systems will fail.

Micah Clark, Paul Bellow, and Selmer Bringsjord wrote this report. The papers of this symposium were published as AAAI technical report FS-15-03, and are available in the AAAI Digital Library.

Embedded Machine Learning

The organizers of this symposium did not submit a report for publication by press time. Papers from the symposium are available on the AAAI Digital Library (technical report FS-15-03).

Self-Confidence in Autonomous Systems

Modern applications of autonomous unmanned and robotic systems have created a demand for sophisticated AI that can be integrated seamlessly with human collaborators. However, it is extremely challenging to guarantee desirable levels of safety, performance, and human-autonomy interaction in practical applications. Autonomous agents are ultimately programmed by imperfect human designers to work with imperfect human users, so the risks of over- or underrelying on autonomous intelligence can never be completely removed. This in turn has sparked interest toward better understanding how mutual communication of intent and the perceived capabilities can affect human-autonomy coordination. This symposium sought to explore these issues from the standpoint of instilling machines with a sense of self-confidence. A key goal was to establish formal algorithmic and computational definitions of machine self-confidence, and thus identify key issues that emerge from allowing autonomous systems to be introspective about the limits of their own capabilities and knowledge.

The symposium attracted healthy participation from a highly diverse set of researchers working in AI, expert systems, natural language processing, military systems, human factors, robotics and aerospace engineering. Three key elements of machine self-confidence (which are also relevant in the context of human-generated self-confidence) constituted the symposium sessions: (1) task competency (that is, how does an autonomous agent know what it can actually do, and what it ought to do?); (2) information adequacy (that is, is sufficient information available to assess the situation and take appropriate action?); and (3) quantification and expression (that is, how can or should self-confidence be expressed, and what are the consequences for users?).

The invited talks and contributed papers helped illustrate the complex interplay between these three elements, and highlight a rich set of motivations and tools for studying machine self-confidence. Two invited speakers (Danette Allen, NASA Langley Research Center; and Kalmanje Krishnakumar, NASA Ames Research Center) described how recent advances in intelligent machine autonomy have fundamentally reshaped the way future aerospace robotic systems will be designed, especially to carry out missions that are impossible to perform with human crews (for example, scientific exploration of the icy moons of Jupiter under extreme communication delays). Another invited speaker (Jeffrey Morrison, Office of Naval Research) echoed similar themes from the perspective of defense applications, but emphasized how decision-making needs can change much more rapidly and unpredictably in this domain — to the point where autonomy should be designed to help get human users into the ballpark of correct decision making, rather than replace them completely. The presentation of the diverse roles to be fulfilled by autonomy led to much discussion on the importance of considering the limits of certainty in machine reasoning and its impact on supporting good decision making, as well as on acceptance of autonomy by different types of users.

These ideas were complemented by the final invited speaker (Marek Druzdzel, University of Pittsburgh), who described how the concept of Bayesian surprise could be used by probabilistic expert systems as a way to assess confidence in recommendations made based on uncertain evidence, incomplete models, and limited training data. This idea also fed a

broader discussion on the utility and ethics of self-confidence presentation. While it was agreed that self-confidence can serve as a useful "shortcut" to building trust and coordination, the question also arose of whether machines should always provide honest self-confidence assessments (that is, whether deceiving users about the capabilities/knowledge of autonomous agents in certain cases is ever beneficial in certain cases).

As highlighted by the contributed papers, the question of machine self-confidence can also be practically applied to low-level task analysis, planning and design of resilient AI in various problem domains (robotic manufacturing, natural language understanding, adaptive flight control, character-based reasoning for AI, to name a few). Most papers focused on general statistical techniques for assessing information adequacy, building on the concept of Bayesian surprise and the related notion of information volatility/currency, while others focused on the idea of assessing self-confidence through plan resilience (through counter planning and plan repair strategies) and low-level task competency. Yet, all contributions made it clear that we have barely begun to scratch the surface in terms of fully understanding the implications and potential impact of machine self-confidence in both theory and practice.

Two interesting takeaways from the symposium highlight some of the core ideas driving future research in intelligent autonomous systems. First, machine autonomy is fundamentally about task delegation, rather than task relegation (which is the object of automation). Second, intelligence is necessary but not sufficient for autonomy. In contrast conventional brute-force techniques for validation and verification, new procedures for building safe and trusted autonomous systems will also have to be developed, especially to account for their ability to learn, adapt, and possibly explain their behavior in nondeterministic settings. Will future autonomous machines of the future need to be certified in the same way human pilots and vehicle operators are licensed today (through knowledge-based exams and skills-based tests)? If so, what feedback mechanisms could or should be provided as assurances to human users, and how does the notion of human trust in autonomy come into play?

The symposium concluded with a joint session with the AAAI Fall Symposium on Cognitive Assistance in Government and Public Sector Applications. Participants and organizers shared highlights of discussions from their respective symposia, and were quite excited to find many similarities in themes concerning the practical use and widespread acceptance of cognitive assistants and autonomous systems. An engaging discussion on the idea of machine self-confidence pointed to many exciting overlaps in these topic areas, and both groups of symposia participants indicated strong interest in jointly organizing future symposia to continue the conversation.

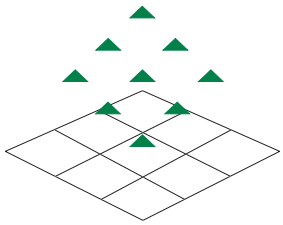
Slides of the invited speaker talks and other presented material from the symposium are available at scas2015.recuv.org. The authors of this report are Nisar Ahmed and Christopher Miller. The organizers of the symposium were Nisar Ahmed (University of Colorado Boulder), Christopher Miller (Smart Information Flow Technologies), Nicholas Sweet (University of Colorado Boulder), Ugur Kuter (Smart Information Flow Technologies), Andrew Hutchins (Duke University), and Mary Cummings (Duke University). The papers of the symposium were published as AAAI Press Technical Report FS-15-05 and are available in the AAAI Digital Library.

Sequential Decision Making for Intelligent Agents

The Sequential Decision Making for Intelligent Agents symposium provided a dedicated forum for researchers of computational sequential decision making under uncertainty, an area that has gained significant traction in AI. In many applications, dealing explicitly with uncertainty regarding the effects of actions, state of the environment and possibly the behavior of other agents is crucial to achieve satisfactory task performance. Decision-theoretic planning models like the Markov decision process (MDP), the partially observable MDP (POMDP), and their many multiagent extensions have emerged as the dominant paradigm for this purpose.

The program emphasized applications through talks by academic and industry speakers, ranging from technical problem domains such as aircraft collision avoidance, multirobot systems, smart energy grids, and dialogue systems to societal challenges in education and HIV prevention. The emphasis of the discussion was often on identifying the added value of sequential decision making over myopic methods. Theoretically, the advantages of looking far ahead when planning a sequence of decisions are well understood. However, this theory is based on an implicit but crucial assumption: the model of the world is either known or can be learned with high accuracy at reasonable cost. Hence, an important question is how sequential decision making can benefit from data science and machine learning in domains with large amounts of available data.

A second important issue is that traditional approaches assume that an agent receives quantitative reward signals for its actions, but in some important application domains this is not the case. Examples that came up at SDMIA included (1) dialog systems, where the agent gets no explicit numeric rewards for what it says; (2) influencing the behavior of homeless youth through their social network to reduce the spread of HIV, where the resulting change in behavior is difficult to quantify objectively; and



Save the Date!
2017 Spring Symposium Series
March 27–29 2017

The 2017 Spring Symposium Series will be held March 27–29, 2017 at Stanford University. The call for proposals is available at www.aaai.org/Symposia/Spring/sss17.php. Proposals are due June 15. The Call for Participation will be available in August. Submissions will be due to the organizers on October 21, 2016. For more information, please contact the symposium cochairs, Gita Sukthankar and Christopher Geib, at sss17chairs@aaai.org or AAAI at sss17@aaai.org. A preliminary list of symposia will be available at the SSS-17 website in late July.

(3) a collision avoidance system for aircraft, whose policy depends on costs assigned to various conflict-resolution outcomes. In applications such as influencing behavior in social networks and collision avoidance system design, a promising approach is to construct a (possibly subjective) reward system for the problem that empirically results in the desired outcomes, as demonstrated in extensive simulations or field studies. For problems where constructing such rewards is too error prone or laborious, such as conversational systems, there are other feasible approaches, such as imitation learning, that allow the agent to learn a policy directly from a set of example conversations (trajectories through the state space) and thereby remove the need for explicit reward signals.

Besides applications and their associated issues, a varied range of topics was discussed and reflected upon, from relatively recent developments such as solution methods that are able to exploit the computational power of graphics processing units (GPUs) or deep reinforcement learning to new insights on established notions such as commitments or coordination graphs. A fundamental issue that was raised was that the lack of structure in general multiagent problems makes them both excessively and unnecessarily hard to solve. However, most practical multiagent scenarios have a lot of structure in agents' behavior that could be useful from a computational standpoint. In particular, the case was made that it is acceptable if some of that structure is provided by the system designer, as opposed to learning it from scratch. Another fundamental issue that merits attention is a high-dimensional action space, which is often encountered but has received little attention (unlike planning with a high-dimensional state space). Currently, practitioners often resort to problem-specific techniques for coping with large action spaces.

The symposium facilitated sharing of algorithmic ideas, insights into problem domains, and exciting recent results in an open and informal atmosphere. Overall, the symposium generated a lot of enthusiasm and excitement that more and more practical applications of sequential decision making in intelligent systems are coming into reach.

Matthijs Spaan, Frans Oliehoek, and Andrey Kolobov wrote this report and served as chair of the symposium. The organizing committee consisted of Matthijs Spaan (Delft University of Technology); Frans Oliehoek (University of Amsterdam); Christopher Amato (University of New Hampshire); Andrey Kolobov (Microsoft Research); and Pascal Poupart (University of Waterloo). Papers from the symposium are available on the AAAI Digital Library.

Nisar Ahmed is an assistant professor of aerospace engineering Sciences at the University of Colorado Boulder.

Paul Bello is the head of the Interactive Systems Section at the Naval Research Laboratory's Naval Center for Applied Research in Artificial Intelligence.

Selmer Bringsjord is the director of the Rensselaer AI and Reasoning Laboratory and specializes in building logicist AI systems with human-level intelligence and in the logico-mathematical and philosophical foundations of AI.

Micah Clark is the program officer for cognitive science, artificial intelligence, and human-robot interaction at the U.S. Navy, Office of Naval Research. He is also a research scientist at the Florida Institute for Human and Machine Cognition (IHMC).

Bradley Hayes is a postdoctoral associate at the Computer Science and Artificial Intelligence Laboratory at the Massachusetts Institute of Technology.

Andrey Kolobov is a researcher at Microsoft Research.

Christopher Miller is chief scientist of Smart Information Flow Technologies (SIFT) in Minneapolis, MN.

Frans Oliehoek is a lecturer at the University of Liverpool and a researcher at the University of Amsterdam.

Frank Stein is the director of the Analytics Solution Center at IBM.

Matthijs Spaan is an assistant professor at the Algorithmics group, Delft University of Technology, Delft, the Netherlands.