Reports on the 2013 AAAI Fall Symposium Series

Gully APC Burns, Yolanda Gil, Natalia Villanueva-Rosales, Yan Liu, Sebastian Risi, Joel Lehman, Jeff Clune, Christian Lebiere, Paul S. Rosenbloom, Frank van Harmelen, James A. Hendler, Pascal Hitzler, Krzysztof Janowicz, Samarth Swarup

■ *The Association for the Advancement of Artificial Intelligence was pleased to present the 2013 Fall Symposium Series, held Friday through Sunday, November 15–17, at the Westin Arlington Gateway in Arlington, Virginia, near Washington, D.C., USA. The titles of the five symposia were Discovery Informatics: AI Takes a Science-Centered View on Big Data (FS-13-01); How Should Intelligence Be Abstracted in AI Research: MDPs, Symbolic Representations, Artificial Neural Networks, or — ? (FS-13-02); Integrated Cognition (FS-13-03); Semantics for Big Data (FS-13-04); and Social Networks and Social Contagion: Web Analytics and Computational Social Science (FS-13-05). The highlights of each symposium are presented in this report.*

The Association for the Advancement of Artificial Intelligence was pleased to present the 2013 Fall Symposium Series, held Friday through Sunday, November 15–17, at the Westin Arlington Gateway in Arlington, Virginia, near Washington, D.C., USA. The symposia included Discovery Informatics: AI Takes a Science-Centered View on Big Data; How Should Intelligence Be Abstracted in AI Research: MDPs, Symbolic Representations, Artificial Neural Networks, or — ? Integrated Cognition; Semantics for Big Data; and Social Networks and Social Contagion: Web Analytics and Computational Social Science.

The Discovery Informatics symposium provided a forum for understanding the role of AI techniques in improving or innovating scientific processes. Researchers at the How Should Intelligence be Abstracted in AI Research symposium discussed and compared different abstractions of both intelligence and processes that might create it. The Integrated Cognition symposium explored the integration of mechanisms, capabilities, constraints, models, applications and levels. In the Semantics for Big Data symposium, participants

explored the many opportunities and challenges arising from transferring and adapting semantic web technologies to the big data quest. Finally, in the Social Networks and Social Contagion symposium, a community of researchers explored topics such as social contagion, game theory, network modeling, network-based inference, human data elicitation, and web analytics. Highlights of the symposia are contained in this report.

Discovery Informatics: AI Takes a Science-Centered View on Big Data

The Discovery Informatics: AI Takes a Science-Centered View on Big Data symposium builds on an AAAI Fall Symposium and NSF-funded workshop from 2012. We showcased AI techniques for scientific discovery across domains, (biodiversity, climate modeling, clinical interventions in lethal sepsis, neuronal classification, metabolic pathways, drug discovery, search for exoplanets, and detection of neutrinos). The symposium combined invited and paper presentations with a shared session (talk and round table) with the Semantics for Big Data symposium. Academe, industry, and government participants attended.

Discovery informatics can be thought to have three aspects: (1) accelerating science through AI techniques; (2) improving our understanding of the scientific discovery process; (3) advancing AI in support of these two goals. Symposium participants recognized the need of practical metrics for successes. One possible inclusion criterion could be whether AI systems enable the creation of testable scientific predictions (directly or indirectly). This sets a high bar but nonetheless was accomplished by many of the systems presented at the symposium. Describing the difficulties of using human scientific expertise in AI-centered systems showcased the importance of human computing.

The meeting started with a discussion of socially intelligent science by Haym Hirsh (Cornell University), in which he described examples such as the Polymath and FoldIt projects. Anita de Waard (Elsevier) discussed modeling laboratory data in an electrophysiology laboratory and effective data sharing. Claire Monteleoni (George Washington University) described applying machine-learning ensemble methods to climate modeling data. Elizabeth Bradley (University of Colorado, Boulder) described reasoning about processes in paleoclimatology, a field related to climate modeling but with different informatics requirements. Nicolas del Rio (University of Texas at El Paso) described biodiversity forecasting with semantic web services. Peter Karp (SRI International) described widely used informatics tools that model and predict metabolic pathways across species. Semantic web technology is emerging as a powerful platform, and Michel Dumotier (Stanford University) described biomedical hypotheses within the Bio2RDF platform as our joint talk. Rinke Hoekstra (VU University Amsterdam) presented linked open data tools to discover connections within established scientific data sets. Louiqa Rashid (University of Maryland) presented work on similarity metrics linking together drugs, genes, and diseases. Kyle Ambert (Intel) presented Finna, a text-mining system to identify passages of interest containing descriptions of neuronal cell types. Francisco Osuna (University of Texas at El Paso) presented a system to facilitate the creation and execution of data management plans. David Jensen (University of Massachusetts) described developing systems to uncover causal relations in social science data by leveraging quasi-experimental designs in data. Zoran Obradovic (Temple University) described a clinical informatics system to predict outcomes in patients suffering from possibly lethal sepsis. Yolanda Gil (University of Southern California) described integrating visualization steps in workflows for data analysis. Finally, Kiri Wagstaff (NASA Jet Propulsion Laboratory) described discovering novel data signatures from the Kepler satellite's search for exoplanet candidates.

The symposium included a panel on impact featuring Kayur Patel (Google), who described some fallacies of big data science doctrine; Pietro Michelucci (Thinksplash), who described the importance of human computation to scientific discovery; and Barbara Ransom (National Science Foundation), who described the semantic and social challenges faced by the EarthCube geosciences project. During a joint round table discussion with the Semantics for Big Data symposium, participants discussed (A) using semantic metadata and ontologies, (B) discovery, and (C) how we can support scientists in the context of big data.

The symposium was chaired by Gully APC Burns,Yolanda Gil, Yan Liu, and Natalia Villanueva-Rosales. The papers of the symposium were published as AAAI Press Technical Report FS-13-01.

How Should Intelligence Be Abstracted in AI Research: MDPs, Symbolic Representations, Artificial Neural Networks, or —?

The field of artificial intelligence is large and divided into many subcommunities, each with its own approach to achieving the common goal of creating computational intelligence. These separate communities work with different abstractions of intelligence (for example, connectionist neural networks, Markov decision processes, and symbolic representations) and different abstractions of the processes that might produce intelligence (for example, reinforcement learning, evolutionary computation, logical inference, and statistical machine learning). Because practitioners of different AI subdisciplines rarely talk to

one another, the goal of the symposium was to bring together members from each of the major AI subfields to discuss our diverse abstractions, perspectives, inspirations, and methodologies, thereby encouraging a cross-pollination of ideas between various levels and types of abstraction.

To our delight, the symposium was attended by researchers from most subfields of AI, including natural language processing, computational linguistics, computational neuroscience, robotics, symbolic AI, evolutionary algorithms, deep learning, reinforcement learning, and developmental robotics. We also had attendees from a variety of other disciplines and backgrounds, including statisticians, engineers, biologists, and entrepreneurs.

One controversial topic of conversation throughout the symposium was deep learning, which is a bioinspired technique based on how the brain processes information that has recently shown impressive performance over a wide spectrum of machine learning benchmarks. The symposium included keynotes from Andrew Ng (Stanford University) and Randall O'Reilly (University of Colorado, Boulder) related to this topic. Andrew Ng provided background on deep learning and its accomplishments from a more abstract machine-learning perspective, while Randall O'Reilly took more direct inspiration from the architecture of the human brain.

The multidisciplinary group of researchers also included biologists like Georg Striedter (University of California, Irvine), author of the book *Principles of Brain Evolution*, who gave a keynote describing how brain functionality has been viewed historically, from Plato to the present. In another keynote, Risto Miikkulainen (University of Texas at Austin) described how the field of neuroevolution, which evolves artificial neural networks and creates cognitive architectures through bottom-up evolutionary design instead of by top-down human engineering. Pierre-Yves Oudeyer (Inria, France) described his work in developmental robotics in a keynote that focused on how robots can be motivated by a curiosity to explore their world. In the process they build a model of the world and develop impressive skill sets.

While the current winds of AI seem favorable to subsymbolic approaches, proponents of symbolic AI made convincing arguments for its continued relevance and promise. In particular, Gary Marcus (New York University) argued that nonsymbolic approaches to AI may help us create AI at a nonhuman animal level, but that symbolic manipulation (good old-fashioned AI) will be essential for human-level thinking such as understanding language and inference. John Laird (University of Michigan) highlighted the difference in goals and outlook between general symbolic AI and the more common approach of tailoring AI to achieve a particular goal.

Through the course of the discussion, many remaining challenges for AI became evident that cut across traditional boundaries. Different approaches had different strengths, weaknesses, and focuses. Their current abilities were clustered around three main divisions: building features from raw perception, making reactive decisions based on features, and higher-level cognitive reasoning. For example, deep learning focuses mainly on building features from raw perception, while reinforcement learning techniques focus on making decisions given such features. On the other end of the spectrum, symbolic approaches focus directly on cognitive reasoning. However, no approach seemed yet able to encompass all three levels.

The participants in general expressed interest in attending a follow-up conference and many reported that they had gained a greater understanding of AI as a whole, and in particular how the various questions tackled within subfields connect and complement each other.

Sebastian Risi, Joel Lehman, and Jeff Clune served as cochairs of this symposium. The papers of the symposium were published as AAAI Press Technical Report FS-13-02.

Integrated Cognition

Integrated cognition is concerned with consolidating the fundamental functionality and phenomena implicated in natural minds and brains or artificial cognitive systems, such as are key to building virtual humans, intelligent agents, and intelligent robots. It captures a grand challenge central to both artificial intelligence and cognitive science: how minds that are capable of yielding human-level performance in complex environments arise from the interactions among their constituent parts and mechanisms.

Integrated cognition spans not only the traditional cognitive aspects — such as planning and problem solving, knowledge representation and reasoning, language and interaction, reflection/metacognition and learning — that have been the focus of unified cognitive architectures, but also seeks a grand unification with the key noncognitive aspects, such as perception and control, personality and emotion, and motivation. It also concerns integration within and across multiple levels of processing and representation, from high-level social and rational thought and symbolic cognitive processes down to low-level biological, reactive, and subsymbolic processing.

In principle, several comprehensive conferences overlap substantially with integrated cognition, including AAAI's own annual conference. However, a focus on individual capabilities rather than integrated systems, and on methods of evaluation that are appropriate for the parts but not necessarily for the whole, has made them a less natural fit than they ought to be. A number of more specialized conferences have also arisen over the years that overlap with integrated cognition — such as Advances in

Cognitive Systems (ACS), Artificial General Intelligence (AGI), Biologically Inspired Cognitive Architectures (BICA), Behavior Representation in Modeling and Simulation (BRIMS), and the International Conference on Cognitive Modeling (ICCM) — but each has its own particular perspective and focus, resulting in a field that is fragmented into partially overlapping subcommunities. The goal of this symposium was to exchange research results across these subcommunities and perspectives, to attempt to leverage those results across levels and domains, and to discuss how best to facilitate such exchanges in the future.

In service of this goal, we recruited a relatively large organizing committee, with members in leadership roles across all of the specialized conferences listed above. In consequence, we did succeed in bringing together at the symposium researchers from artificial intelligence, artificial general intelligence, cognitive science, and neuroscience who work across a broad spectrum of aspects of, approaches to, and perspectives on integrated cognition. We had contrasting invited keynotes from John Laird (University of Michigan) and Chris Eliasmith (University of Waterloo), focusing respectively on integration from the AI/cognitive and computational neuroscience perspectives. We had invited opening and closing panels that provided alternative perspectives on integrated cognition and assessments of what the current consensus is concerning integrated cognition and what the biggest open issues are in attaining it. We also had sessions organized around accepted panels and papers, involving numerous aspects of integrated cognition: social, emotional, and motivational; metacognitive; neurocognitive; integrating language and cognitive architecture; knowledge and mechanism integration; and task-level integration. Significant discussion time was set aside as part of each of these sessions for across-perspective exchanges concerning both the specifics of what had been presented and the general topic of the session.

During the latter part of the symposium, a significant amount of side discussion was devoted to the question of where to go from here. There was little enthusiasm for the idea of creating yet another conference, even if it would be focused squarely on integrated cognition. However, several promising follow-on activities were proposed, including a summary article on the consensus and open issues regarding integrated cognition, a special issue of a journal, a multiweek summer school, and a redefinition of the existing conferences to expand their existing scope to provide a natural home for a broader range of work on integrated cognition. A number of these ideas are currently under serious consideration.

Christian Lebiere and Paul S. Rosenbloom served as cochairs of this symposium. The papers of the symposium were published as AAAI Press Technical Report FS-13-03.

Semantics for Big Data

To extract knowledge and insights out of big data requires novel methods that can cope with the heterogeneity, diversity, noise, and complexity of the data, while its volume and velocity forbid solutions available for smaller data sets. For instance, the manual integration and curation of large-scale data sets from different sources is no longer feasible. Semantic web technologies promise to deal with these challenges and, thus, jointly with work on data mining, high performance computing, and knowledge infrastructures, becomes a pillar of sorts to effectively tame big data. The symposium on semantics for big data aimed at providing a forum to discuss the relation between semantic web and linked data research and the currently unfolding landscape of big data analytics.

The symposium was structured around five major topics that emerged from the submitted papers. Each of these topics was represented by a presentation session followed by a panel or breakout group.

The first topic was devoted to crowdsourcing and cognitive aspects. It also included a panel on the role of human competencies for big data analysis in combination with intelligent systems. The participants noted a gap between the usage and success of crowdsourcing in other research fields compared to artificial intelligence and the semantic web in specific. In this context trust and provenance as well as their interplay were discussed. Another interesting notion centered on physical-cyber-social systems that aim at a more holistic view on the interplay of humans and computers. With respect to semantics and the processing of big data, these systems can help transform and aggregate abstract data, patterns, and analysis into actionable information; thus converting the constant stream of data into human understandable and tailored pieces of information.

The second topic was concerned with making various semantic web technologies, such as reasoning services and information extraction algorithms, scale so as to be suitable for big data problems. For instance, one case study discussed the difficulties of using semantic web technologies and the SPARQL query language to analyze network data from a massive multiplayer online role-playing game. In this context, one of the core challenges is the partitioning of large graphs for parallel processing without creating extensive communication overhead between the concurrent processes. The most intensive discussion during the symposium, however, emerged around three proposed laws derived from research on web-scale reasoning. Are there laws in the computer and information sciences, or are laws reserved to the natural sciences, for example, physics? How can they be evaluated and what phenomena do they explain? In this context, the participants compared the proposed laws to Richard Feynman's treatise *The Character of Physical Law* (Cambridge, MA: The MIT Press, 1965).

A third theme addressed the evaluation of seman-

tic web technologies for big data analysis and the need for standardized test beds as baselines. For instance, establishing semantic interoperability via alignments between different standards and data formats in the oil and gas industry was discussed as an interesting test bed for the suitability of semantic technologies.

The fourth topic was concerned with the combination of deduction and induction approaches. There was a broad agreement that an efficient handling of big data has to enrich bottom-up, data-driven methods known from data mining and machine learning with a declarative top-down layer provided by various semantic web technologies as well as ontologies that provide formal specifications of constraints and background knowledge. IBM's Watson was discussed as one example where knowledge and large amounts of unstructured data came together. Along the same lines, the participants discussed current trends in learning axioms from data as well as the increasing interest of social network researchers in semantic tagging.

Finally, the fifth topic dealt with big data integration. The participants noted that the focus on storing, accessing, and processing big data has drawn attention away from equally important challenges such as how to search for, normalize, transform, aggregate, integrate, and summarize massive amounts of data from highly heterogeneous sources. This, however, is a prerequisite to arrive at a more holistic understanding of many major questions in research and society that typically span across multiple domains — global change being a typical example that involves data and theories from economics, political sciences, geography, climatology, and so forth.

While most of the symposium focused on semantic technologies, one interactive session was entirely devoted to the role of ontologies. More concretely, the participants discussed the sweet spot for ontologies as a communication layer between numerical and statistical methods and models on the one hand, and the human user on the other hand. This included questions about the relation of semantic web ontologies to the ANSI data model stack as well as the appropriate level and granularity of axiomatization. Following the well known slogan that a little semantics goes a long way, the participants debated the question of how much a *little* actually is.

The symposium also included presentations by three invited speakers. Jennifer Golbeck from the University of Maryland presented her work on how to compute trust and building trust with users' social media data. Peter Fox from the Rensselaer Polytechnic Institute reported on the mediation and integration of geodata through formal vocabularies and highlighted the role of semantic heterogeneity. Finally, Stanford's Michel Dumontier discussed the generation of biomedical hypotheses using semantic web technologies. This last keynote was organized jointly with the Discovery Informatics symposium and was followed by a joint discussion session.

Frank van Harmelen, James A. Hendler, Pascal Hitzler, and Krzysztof Janowicz served as cochairs of this sympo-

sium. The papers of the symposium were published as AAAI Press Technical Report FS-13-04.

Social Networks and Social Contagion: Web Analytics and Computational Social Science

Computational social science is rapidly becoming a topic of sustained interest in the computer science and artificial intelligence communities. These systems science approaches are drawing in researchers from domains such as public health, social science, marketing and advertising, and others. In these domains, the ubiquity of networks and contagionlike processes on these networks is leading to advances in network science, epidemiological modeling, and simulation science. These advances are facilitated by the availability of vast quantities of social media data.

This symposium brought together researchers from multiple disciplines, including public health, game theory, network science, social science, and business, with the common goal of developing interdisciplinary solutions to these problems.

One major theme of the symposium was the use of large-scale, data-driven simulation tools for problems at the intersection of public health, disaster resilience, and social and political science. These synthetic information models are created by combining multiple sources of data to create a high-fidelity representation of interactions within a social system, and are being used for modeling phenomena ranging from epidemic outbreaks, to social unrest, and disaster response.

The symposium included two invited talks. William Rand (University of Maryland) presented work on trust, influence, and urgent diffusion in social media, and Lise Getoor (University of California, Santa Cruz) spoke about entity-based data science.

Another major theme was the use of game theory for security in domains where networks and contagionlike phenomena occur, such as crime and transportation networks.

A third theme was the use of social networks and survey data to model the spread of social behaviors such as adolescent smoking and drinking behaviors. This work elucidated the need for taking into account the coevolutionary nature of peer networks and the spread of behaviors on these networks, since these phenomena are driven by both homophily and influence. More generally, as another presented paper discussed, the interactions between agent-internal cognitive networks (of beliefs, attitudes) and the agent-external networks (of communication, influence) can lead to surprising nonlinear phenomena.

The symposium also included six short presentations by student scholarship winners, on topics ranging from studying bullying on social media to modeling hospital-acquired infections.

Other topics that were covered by the talks included finding spatially clustered communities in networks, contagion on hypergraphs, models of web-browsing behavior, and network perspectives on software development in groups.

The variety of topics and breadth of discussion served as a reminder of the diversity of applications and approaches in computational social science. Symposium participants discussed possibilities for crossover and collaboration, and agreed on the importance of having a forum such as this symposium for encouraging new ideas.

Samarth Swarup served as the chair of this symposium. Madhav Marathe, Kiran Lakkaraju, Milind Tambe, and Cynthia Lakon were coorganizers. The papers of this symposium were published as AAAI Press Technical Report FS-13-05.

Gully Burns is project leader of the Biomedical Knowledge Engineering Group, Information Sciences Institute, University of Southern California.

Yolanda Gil is the principal investigator and project leader of the Interactive Knowledge Capture research group in the Information Sciences Institute and also a research professor in the Department of Computer Science, University of Southern California.

Yan Liu is an assistant professor in the Department of Computer Science, University of Southern California

Natalia Villanueva-Rosales is a research assistant professor in the Department of Computer Science and Cyber-ShARE Center of Excellence, University of Texas at El Paso.

Sebastian Risi is an assistant professor in the Center for Computer Games Research at the University of Copenhagen, Denmark.

Joel Lehman is a postdoctoral researcher in the Department of Computer Science at the University of Texas at Austin.

Jeff Clune is an assistant professor of computer science at the University of Wyoming.

Christian Lebiere is a member of the research faculty in the Department of Psychology at Carnegie Mellon University.

Paul S. Rosenbloom is a professor in the Department of Computer Science at the University of Southern California and a project leader at the Institute for Creative Technologies.

Frank van Harmelen is a professor in knowledge representation and reasoning at the Vrije Universiteit Amsterdam.

James A. Hendler is a professor for artificial intelligence at the Rensselaer Polytechnic Institute.

Pascal Hitzler is an associate professor for semantic web technologies at Wright State University.

Krzysztof Janowicz is an assistant professor for geoinformatics at University of California, Santa Barbara.

Samarth Swarup is a research scientist at the Virginia Bioinformatics Institute, Virginia Tech.