INTELLIGENT COGNITIVE ASSISTANTS

Workshop Summary and Recommendations

Abstract

A workshop to identify the most critical research needed to create Intelligent Cognitive Assistants: platforms which augment human capabilities.

Workshop Websites:

https://www.src.org/calendar/e006057/

https://www.nsf.gov/nano/

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Human decisions makers face a daunting list of interrelated and individually complex challenges that require solutions. Pressing long-term and acute problems on the horizon include economic inequalities, natural resource utilization inefficiencies, healthcare failings, an aging population, global instability, and epochal climate changes. There are great opportunities ahead in developing and executing on creative solutions to these and other multidimensional and multidisciplinary problems. However, new forms of computational analyses and human-computer collaboration will be an important part of the pathways to making effective progress on these complex problems.

A workshop was held in May 2016 under the sponsorship of the National Science Foundation to identify the most critical research required to enable a foundational set of ‘intelligent architectures and capabilities endowed with common sense knowledge and reasoning’ which, if applied appropriately, could fundamentally enable humanity to address these challenges within the foreseeable future. We call the most broadly effective implementation of these tools “Intelligent Cognitive Assistants (ICAs).”

This workshop was held not only because of the rapidly growing interest in such ‘ICAs’, but also to address the growing concerns regarding their beneficial and responsible implementation. To discuss the complex challenges involved as well as to identify the required breadth of research to address them, experts from across the computational, social, and cognitive sciences were invited to participate. We discussed future research trajectories and pressing concerns with respect to such intelligent agents, including “moonshot” scenarios pitched by participants and a series of keynote addresses that sensitized the group to the history, trajectory, and social context of such agents. Ultimately, participants synthesized a wide variety of perspectives to formulate an interdisciplinary research agenda of pressing importance.

The group reached a consensus around the concept of Intelligent Cognitive Assistants that complement, rather than replace, human capabilities. These must respond and change flexibly to changing environmental and usage conditions, consider the human life course in their application, facilitate ‘natural’ interactions involving ‘common sense’ toolkits and intuitive interfaces, and ultimately cultivate trust in relations between humans and machines. They should leverage models of the intentions and goals of the people they are supporting. There is a great opportunity to leverage detailed models of human cognition, including an understanding of biases in judgment, and models of attention, memory, perception, and comprehension.

Three scenarios were described – life-long education, group work, and elder care – that incorporate sensitivity to these research parameters in complex social environments, and which require interdisciplinary research to fully address. These life cycle scenarios can also be used directly as a ‘roadmap’ to guide new research towards addressing the most challenging needs as humans individually evolve as well as interact collectively throughout their lifetimes.

Lastly, it was a clear consensus of the workshop participants that because of the breadth as well as depth of the technical challenges, no single entity has anywhere close to the level of resources required to address them by themselves. In addition, because of the breadth of the pending impact of the application of machine intelligence and more specifically, ‘Intelligent Cognitive Assistants’, on such a wide swath of the global public, government has a moral and ethical responsibility to ensure that such research is aimed at the common good.

In fact, there is a very high likelihood that some of the most historically significant twenty-first century struggles will be over the control of such intelligent assistant systems and the data that is used to teach them – struggles between individuals, corporations, and governments.

Therefore, **a public-private partnership within the United States between government, industry, and academia is absolutely essential to drive new fundamental research** towards the most appropriate, effective and publicly responsible 'Intelligent Cognitive Assistant' solutions needed to solve our most critical challenges.

Key Findings and Recommendations

The ultimate goal is to create the most effective and beneficial ICA. This workshop was a first step in defining the role and requirements for ICAs. These requirements must now be addressed by new fundamental research which includes not only the algorithms, architecture and devices to enable these ICAs, but also the critical social science research which is necessary to maximize the effectiveness and benefits for society as a whole.

After hearing about many examples of successful human-computer symbiosis, there was a consensus among the workshop participants that the fundamental goal, looking forward, should be **to develop systems which aim at <u>enhancing</u> human capabilities, rather than systems that are aimed solely at <u>replacing</u> humans in specific tasks.** Instead of driving towards any 'moment of singularity,' such systems can then facilitate the 'continuity of humanity' by sharing the following key objectives:

- Engender human-machine trust
- Mitigate the concentration of external control
- Ensure appropriate security and privacy
- Built-in flexibility and adaptability to changing physical environments, social context, user needs or cognitive conditions by learning throughout their work life with many fewer examples than humans need
- Produce and deploy architectures and capabilities endowed with common sense knowledge and reasoning
- Address a model of social good
- Use a multi-disciplinary research initiative to achieve these goals, including neuroscience, cognitive psychology, sociology, and computational science, including artificial intelligence and machine learning
- Pursue methods that can support new kinds of human-computer collaboration

Specific research recommendations

- Understand 'common sense reasoning' well enough to enable incorporation into ICA system designs which will require research into reasoning and development of an API or tool-kit
- Produce a modular ICA system architecture that designs for scalability and adaptability
- Address the problems of latency-appropriate, unsupervised, on-going machine learning
- Support natural language understanding, generation, and dialog for implementation in human interface scenarios
- Develop 'common-sense,' 'natural,' 'self-explanatory' interface modalities for a broad diversity of users
- Develop computational machinery to support fluid and fluent human-computer collaboration, including methods that understand the complementarity of human and machine contributions, and models of interaction that support mixed-initiative interaction.

Application Driver Examples

- Enhancement of Education and Training, beginning with early childhood development and continuing with preparing children for future careers. Life-long learning including retraining the adult workforce to address changing technology and economic landscape requirements. Such applications should address global audiences in scope, from under-resourced to well-resourced communities.

- Facilitating and enhancing workplace-related and other common goal-oriented group activities and applications, such as facilitating many-to-many interactions, building trust, and coordinating multi-modal communication, as well as providing information to enhance the capabilities of these kinds of groups.
- Elder care and 'Aging-in-place' assistance, including ethical and gentle support with unobtrusive data capture, sophisticated situational recognition, and assisting in facilitating interaction or taking corrective action in a state of human cognitive decline.

Next Steps

A follow-up workshop is proposed to do a deeper dive in to the social science aspects of these ICA challenges. A second workshop already scheduled for March 2017, will be focused on the research gaps and need for new algorithms and system architectures to support the same broad ICA system goals outlined in this workshop.

These follow-up events will serve to provide further details for possible research paths, but in the meantime we propose to **convene a meeting of industry and government stakeholders before January 2017** with the goal of drafting a new research program outline, and begin the process to obtain the commitments necessary to launch this new and critically important program.

WORKSHOP DESCRIPTION

Background

We are facing a daunting list of interrelated and individually complex challenges that require solutions not yet found. At the top of this list are domestic economic inequalities, resource utilization inefficiencies, healthcare failings, an aging population, global instability, and epochal climate changes.

A workshop was held in May 2016 under the sponsorship of the National Science Foundation to identify the most critical research required to create intelligent architectures and capabilities endowed with common sense knowledge and reasoning which, if applied appropriately, could fundamentally enable humanity to address these challenges within the foreseeable future. We call these collective architectures and capabilities "Intelligent Cognitive Assistants (ICAs)."

The drive to create these 'ICAs' in a wide variety of forms and with an even broader set of capabilities began over 60 years ago, and were even envisioned centuries earlier. Progress towards the broadest and most general implementation of these tools has ebbed and flowed over the past several decades in this quest, but in the past half-decade has accelerated due to advancements in semiconductor technology as well as information access through the world wide web which have in turn enabled advancements across a variety of scientific and engineering disciplines. Collectively these advancements have not only enabled rapid progress in such hot topics as 'machine learning' but also in the basic understanding of the operation and characteristics of the human brain, the most efficient and effective cognitive 'tool' we know.

However, these rapidly accelerating advancements are also creating a critical inflection point. Within our generation, or certainly within that of our children, we are likely to witness two transformative events– the creation of machine intelligence, and the comprehensive connection of humans, their devices and machines via a common communication network. These transformations can further democratize and accelerate new discoveries and innovation, allowing us to address many if not all of the challenges listed above. We will discuss this point by way of examples later in this report.

This situation will also however create unprecedented challenges and risks such as: (1) the potential to create further disenfranchisement and widening of gaps between segments of society, both in terms of education and relevant training. This can lead to further acceleration of gaps in income, underemployment, and other similar negative macroeconomic impact on societies at large. Further (2), novel runaway privacy and security challenges will no doubt arise from attempting to walk the narrowing line between fostering open access to data and information on the one hand, and incorporating appropriate protections against both the accidental and malevolent actions of others on the other.

These challenges will be made even more complicated by the fact that countless devices will be connected to each other with limited human intervention or direct control by design, such as systems that include self-repair, self-regulation, or even self-replication. Such complex and self-managed systems will have at least two primary and related weaknesses: (1) Flaws which are initially minor and obscure could potentially escalate via unpredictable sequences towards much larger and more damaging failures; and (2) It will become increasingly difficult to identify, much less avoid, single points of weakness without adding potentially overly burdensome and energy inefficient protections. These situations leave such systems vulnerable to failure and/or attack, especially if access to these systems becomes concurrently more ubiquitous to populations at-large.

Further, these unprecedented and critically important challenges mentioned discount the possibility of developing fully 'conscious' machines, ones that have some sense of 'self' and self-preservation. This concept, which up to now has remained almost entirely within the realms of science fiction and philosophical discourse, is now becoming a topic of increasing technical plausibility. For some, it is no longer assumed to be a question of 'if,' but 'when'.

While all of this new capability and technological capacity can easily be seen as changing the world for the better, this newly discovered power must also come with greater recognition of associated and collective responsibilities. In the end, it will be our fundamental values that will matter most, and how we imprint these values on our technological creations will determine the consequences of our collective choices, and in turn determine our shared future.

From a Technology Perspective

We also are at the cusp of a new era of computing. Initially, machines were designed and built to accelerate the performance of basic arithmetic calculations, as compared to human 'computers.' Subsequently these machines, by then called 'computers' themselves, were programmed to run complex simulations, as well as enable globally connected networks. We are now beginning to explore the use of machines to enhance and augment human cognitive abilities.

Computers up to now have essentially been designed from the 'inside out' – programmed by humans to perform specific tasks which act on their environments. Current trends in computation add a new design vector by designing computers from the 'outside in' – the external environments directly influencing emerging computing system design via 'training' (as with 'deep learning'), embedding cognitive capabilities (as with '1-shot learning'), and better appreciation of human-machine environments (as in human-computer interaction).

We envision the most effective 'machine intelligence' as an active interface between humans and their environments, providing insight and guidance for problems that cannot be handled efficiently nor most effectively by the unaided mind or by computers alone. In fact, the performance of such intelligent systems should be measured by the 'goodness of fit' more than any other single parameter.

How to optimize this collaborative interaction between humans and these 'intelligent' machines is an open research question. In order to create the foundation for future intelligent systems that can most effectively and efficiently assist individuals, businesses and society at large, it is essential that this research question be addressed.

Workshop Outline

To accelerate progress toward developing 'intelligent' and 'responsible' machines that can serve as effective and robust cognitive assistants to improve human productivity and overall quality of life, a workshop on "Intelligent Cognitive Assistants" was held on May 12-13, 2016. (**https://www.src.org/calendar/e006057/**)

The goal of the workshop was to gather experts from research fields spanning psychology, sociology, artificial intelligence and machine learning, robotics, computer science and engineering in order to identify the highest priority research still needed to address the challenges of creating the most effective and beneficial 'intelligent agent-human symbiosis.'

The workshop opened with two keynote talks. Eric Horvitz from Microsoft began by reviewing some of the AI history, gave us glimpses of some state-of-the-art application demonstrations, and shared his thoughts on which directions to head. Cynthia Breazeal, from MIT and Jibo, Inc., then gave us a look at the challenges and opportunities for using intelligent agents within social environments, and described specific examples of applications where they can assist and augment human capabilities.

The workshop then moved to a number of panel sessions. First, a number of experts from a wide range of disciplines and backgrounds presented a brief (2-3 slide) synopsis of their visions of possible future applications and the potential for intelligent assistants – what we called 'moonshots' – followed by an open dialog with the audience.

Next we moved to second set of panel sessions, where we heard about and discussed some of the implementation challenges and possible tradeoffs required to realize these future visions. All workshop participants, including the panelists, were provided a list of example 'implementation challenge questions' to consider prior to the workshop, and this list was then used to drive both the panelists' presentations as well as their dialog with the audience.

Within this portion of the workshop agenda, there was a third keynote talk given by Blaise Aguera y Arcas from Google. He spoke about the strengths and weaknesses of deep learning algorithms using some specific and recent application examples. He ended by commenting on implementation challenges such as developing 'trust' between humans and intelligent systems, the risks of any concentration of data access and control, and tradeoffs like 'augmentation versus assistance' and those involved with the 'responsible' usage of AI as it relates to human interaction with the natural world.

The second day of the workshop was devoted to breakout discussions. Each of the three parallel breakouts were designed to include diverse cross-sections of our workshop participants, each of them bringing unique and complementary expertise, experiences, and vision to the table. Each breakout team then presented their findings and specific recommendations which will be summarized next in this report under "Workshop Outcomes."

Lastly, our proposed 'next steps' build upon these key findings, driving towards the creation of new research programs that could potentially address the many critical challenges identified during the workshop, to enable the delivery of the envisioned benefits of 'Intelligent Cognitive Assistants.' These 'next steps' were described earlier and are listed again at the end of this report.

WORKSHOP OUTCOMES

Breakout discussions on the second day of the workshop aimed to synthesize the broad perspectives presented on the first day into practical, achievable, and ethical principles for establishing a well-grounded scientific research program in ICA. All three breakout groups focused on how to produce assistants that do not supplant or replace humans, but rather complement them in key ways, drawing on insights from cognitive, social, and neuro sciences, as well as computing research agendas. All three groups also addressed the ethical and social implications of creating such assistants, especially ways to avoid the social unrest, job losses, and inequalities that may result from rapid ‘technologization’ of the workplace. We also produced and shared scenarios in which intelligent assistants could be most beneficial to society while avoiding pitfalls of automation, and outline a few of these examples below.

Guiding Values

To help address these and other related tradeoffs and challenges, it was suggested that **defining a clear ‘value statement’ to guide and regulate ICA design, operation, and usage**, including governing the computational algorithms which would implement such values within in ICA system should be done. Although we acknowledge that ‘values’ and ‘judgements’ vary across societies and even between segments of societies, it is nevertheless critical to aim at defining guiding values and principles upfront. This can also better assist with developing an integrated ability to allow for guided modifications to facilitate specific applications down the road.

One approach discussed was to begin by identifying the elements most common to a variety of constrained environment applications, easily applicable to more complex, flexible, and adaptable solutions that could be implemented across a range of environments. To that end, we include several high-level goals and principles that should direct ethical, measured, and exceptional scientific and technical research in this domain:

(1) Enhance, not replace, human capabilities.

After hearing about many examples of successful human-computer symbiosis, there was a consensus that the fundamental goal, looking forward, should be to develop systems which aim at **enhancing human capabilities** and not systems aimed solely at replacing humans in specific tasks. One group described this is a fundamental “pillar” of ICA research based on “**complementarity**” between humans and machines [FIG.1]. This means facilitating the continuity of humanity's unique characteristics rather than driving towards fully competent machine intelligence. It also means aiming not to supplant or ‘disrupt’ human skilled labor, but to meaningfully integrate ICAs within existing workplace environments across a range of classes and types of work.

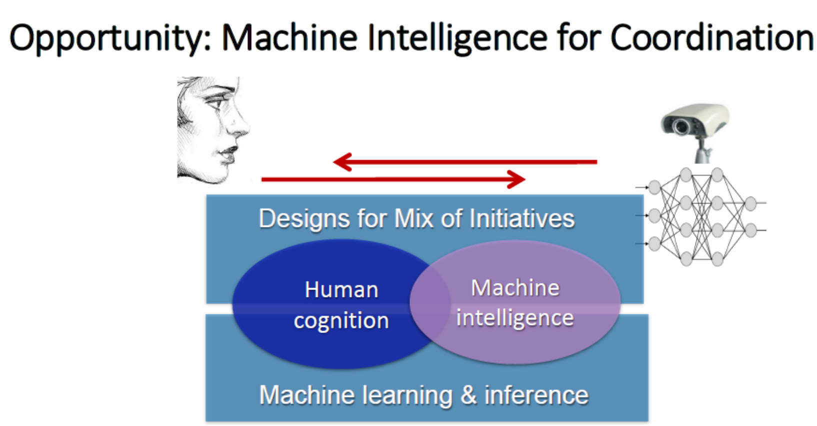

FIG.1 (Eric Horvitz)

Such systems would value enhancement, support, and assistance; they would locate continuity with our cultures and cultural environments; and they would eschew "chilling effects," uncanny valleys, or too much knowledge of their human users to engender a sense of creepiness. It is therefore worth investing in research into the evolving "sweet spot" in the division of labor between humans and machines, one that makes us feel more human as we work with such machines and which does not impinge on human autonomy or freedoms, both perceived and real.

Along with enhancement comes the possibility for addressing human biases. In many cases these are well understood as limitations of cognition, such as confirmation bias or inattention, which may occlude good decision-making or human cooperation: successful ICAs would not be a distraction but would help us to avoid these common errors. Breakout groups also discussed forms of implicit bias, and the importance of avoiding discrimination on the basis of race, class, gender, or geography in the development of ICAs. One group found consensus around phrasing that suggested ICAs "embrace a model of social good" and "empower minorities" or global underclasses to immediately address the developing global inequalities surrounding technological development and workforce replacement.

(2) Adapt with flexibility to dynamic, real-world environments

Along with complementing human qualities, there was also a general consensus that **flexibility and adaptability** are essential characteristics of ICA systems as they must integrate into dynamic real world environments [FIG.2]. This flexibility must respond to changing or uncertain physical environments, by learning incrementally throughout their work life, and doing so with ideally many fewer examples than humans require. Yet such systems must also respond to shifts in social context or changing user needs, due to aging or other physiological changes throughout the life course. As one group put it, this requires not only understanding immediate human goals but also being capable of operating in a larger, open, and dynamic world. This adaptable context-awareness must be among the fundamental features of any Intelligent Cognitive Assistants/agents (ICA) to create sustainable value in the long term.

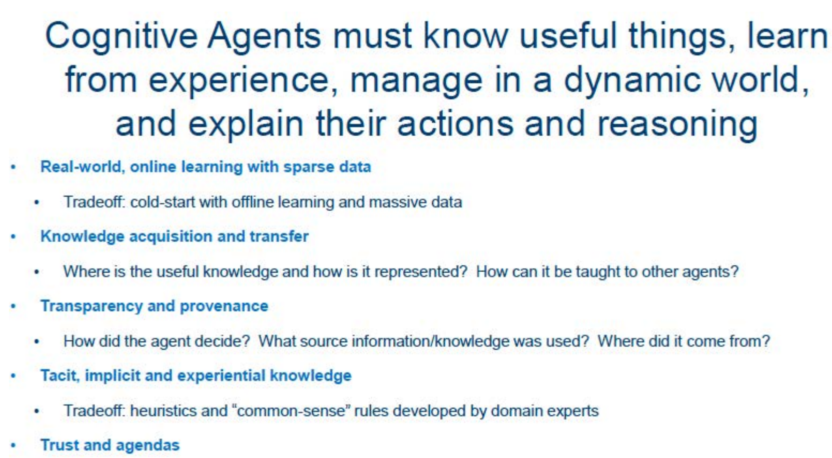

FIG.2 (Bruce Horn)

While context awareness is clearly essential for ICA development, the consensus across the groups was that such systems did not need to fully capture, model, or otherwise describe the environment or emotional context of interaction. Attention to dynamism and change in many cases precludes such a global model. Instead, the discussion groups focused on "building gracefully" on existing technologies and local understandings instead of looking for all-encompassing technical solutions. For instance, successful ICAs in line with the above two principles might not solve a problem for their human interlocutors but rather connect them with another person who has relevant expertise, and potentially provide additional support to that pair.

(3) Cultivate trust among humans and machines

Producing strong working relationships with intelligent assistants requires shared trust among human users of the system. This must be upheld as a design goal upfront in the development process and be maintained throughout any modifications/enhancements down the road. Trust requires attenuation to interaction patterns and accurate fulfillment of requests and system tasks, an attention to coordination of tasks and resources. But it also requires thoughtful care as to how and where data about such requests travels, how ICAs track or log user interactions, and intersections with IoT (Internet of Things) security and data protections.

Current research already demonstrates that any concentration of control or surveillance, especially through workplace technologies, creates an unsustainable and problematic situation and lack of trust. This problem compounds in systems that are designed to be ubiquitous in their application to society at-large. To that end, we recommend that research in the domain of ICAs address and fundamentally incorporate the extremely challenging privacy, security, ethical and regulatory challenges and tradeoffs associated with intimate digital systems. We must address these concerns both upfront and on-going throughout any ICA development process in order to maintain trust and valuable working relations among humans and ubiquitous machines.

(4) Facilitate "natural" interactions

To enable mutual trust, both intuitive interfaces and 'common sense reasoning' must be incorporated into ICA system design. It is of course a research challenge in itself to both understand and incorporate such a "common sense" concept into ICA design, and perhaps producing a common sense reasoning "toolkit" with an associated application programming interface (API) for research and application development offers one way forward. Such a toolkit should support domains of common sense such as for the physical world and for the social world. Research support is needed here both to develop the understanding of how humans reason, learn and collaborate -- with accompanying rich user modeling -- as well as in learning how to utilize that understanding to create the desired ‘toolkit’. This research may take quantitative or experimental approaches to the problem, or engage ethnography, contextual inquiry, reflective or experience-focused design practices, in order to produce computational systems that engage this problem.

ICAs must be able to learn how humans work best (whether as an individual or with a team) and facilitate, complement, or augment their abilities with the most appropriate interfaces and interactions [FIG.3]. This requires several basic principles that could perhaps form part of the above-mentioned "toolkit." One significant challenge for research in need of more support is natural language understanding, as well as multi-modal conversational abilities including gesture, affect, and even interactions via sketching – all highly useful to address a wide variety of human-ICA application interface scenarios. Additionally, latency-appropriate, unsupervised, on-going learning is a key enabling feature for many if not most of the future, evolving applications envisioned. This is still an extremely challenging research goal in need of additional support. At times, people will also want to explicitly teach ICAs, and doing so requires research into how to make this process natural, efficient, collaborative, and capable of learning effectively from a small number of examples. In addition, what an ICA learns should be transparent, explainable, and easy to correct if undesired biases are unintentionally introduced. Finally, open questions about anthropomorphic qualities remain with respect to intentionally-designed interactive systems, and we therefore invite research into the design or attribution of personality, empathy, or other social qualities with respect to both embodied and disembodied cognitive assistants.

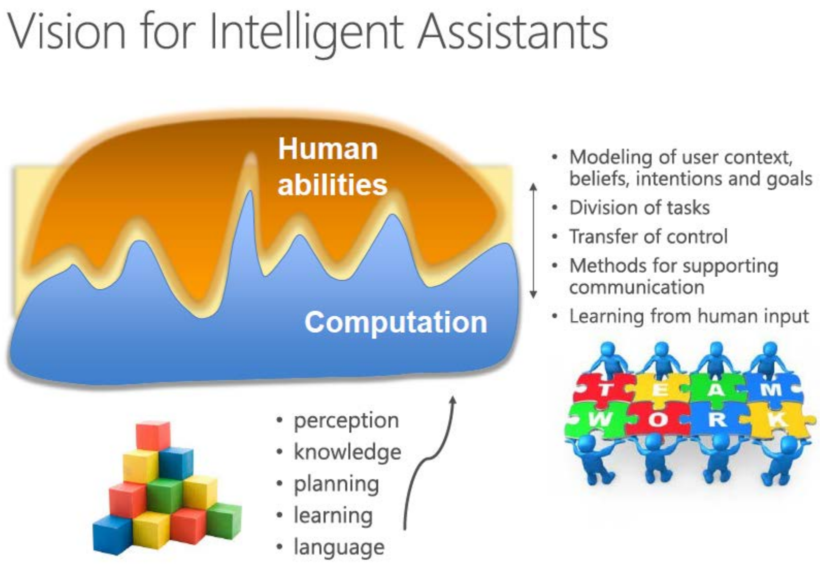

FIG.3 (Ece Kamar)

Coordinating human and machine multi-modal dialog, especially in groups, is essential – taking into consideration timing, pace, fluidity, and the costs of interruption. Agents must be developed with a "cadence" in mind of human interaction and the changing life course, as well as attentive to the cadence of near-futures: one group wondered if an ICA provided immediate gratification to the "now me" or instead remembered tasks so that they would not be forgotten by the "future me." Finally, a modular ICA system architecture may be most useful and flexible as a design feature for scalability, adaptability, and to enable design enhancement down the road.

(5) Incorporate multi-disciplinary perspectives

To address the above challenges, a multi-disciplinary perspective is essential. This must push beyond simply uniting sub-disciplines within computer science (i.e. machine learning versus natural language processing), psychology or the neurosciences, but reach across disciplinary divides and incorporate the social and behavioral sciences as meaningful interlocutors [FIG.4].

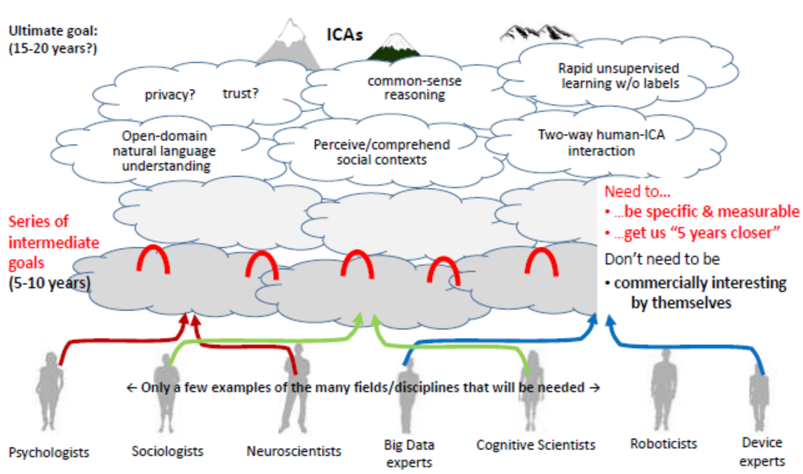

FIG.4 (Geoff Burr)

Research Driven Application Examples

One way to approach research in this domain is to address each of the above elements separately as unique research questions. We outline several of these distinct elements below. However, workshop participants found more value in the opposite approach: to focus on a variety of applications that require addressing and synthesizing across the above-mentioned domains. Such a comprehensive initiative would also facilitate direct collaboration across multiple disciplines aimed at a common set of goals. This initiative does not yet exist in this research community and is the most urgent and fundamental finding of this workshop.

It was the general consensus that focusing on a few application domains – especially those in which ICAs could potentially address some of the most critical and broadly impactful challenges facing society today – would enable a useful focus of attention and most efficient use of resources. Just as ICAs hold the potential to transform professional settings across a wide number of domains and industries, they could also transform the home and how we address a diversity of quality of life issues across ages and stages. Furthermore, from an ethical standpoint, the workshop participants acknowledged that ICAs could exacerbate inequity in society if we do not explicitly design to address the challenges and needs of diverse populations, including also the underserved and under-resourced.

We therefore recount here three application domains that were the most frequently mentioned during the workshop and discussed in some detail during the breakout sessions. Each domain covers the wide range of ethical and sociotechnical considerations discussed above, with plenty of opportunities to address specific scientific challenges. As one breakout group put it, this take the "**human life span as a roadmap**" for ICA development [FIG.5], with modalities that address **early development and education**, **group activities and working environments**, and **elder care**.

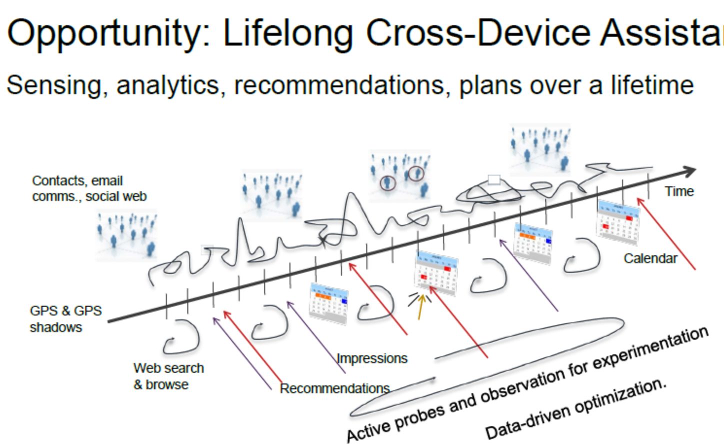

FIG.5 (Eric Horvitz)

(1) Enhancement of Education and Training

Just as periods of cognitive decline can produce opportunities for ICAs, periods of ramp-up or increasing capacity are also excellent candidates for Intelligent Cognitive Assistants that do not replace but rather supplement and support, or even enable augmented human capabilities. Education is fundamental to enabling opportunity: from **preparing our children** for future careers, **retraining the workforce** as career skills shift with the technological landscape, and **contributing to global stability** abroad.

ICAs hold the potential to transform how, when, and where we learn – far beyond the instrumented classrooms and MOOCs of today [FIG.6]. For instance, personal tutoring is recognized as being far more effective than lectures in the classroom. ICAs could make deeply personalized learning scalable and affordable to many: from interactive learning companions for early childhood education, to ICA educational assistants to enhance learning in classrooms with teachers, while also reinforcing and extending quality learning opportunities in the home. Given that education is a longitudinal endeavor, this motivates research into ICAs that can engage people over extended time scales, modeling and personalizing the learner's state to identify the zone of proximal development for optimized learning, and to support the learner not only in terms of curricular objectives but other lifelong learning skills like mindset, curiosity, empathy, and more.

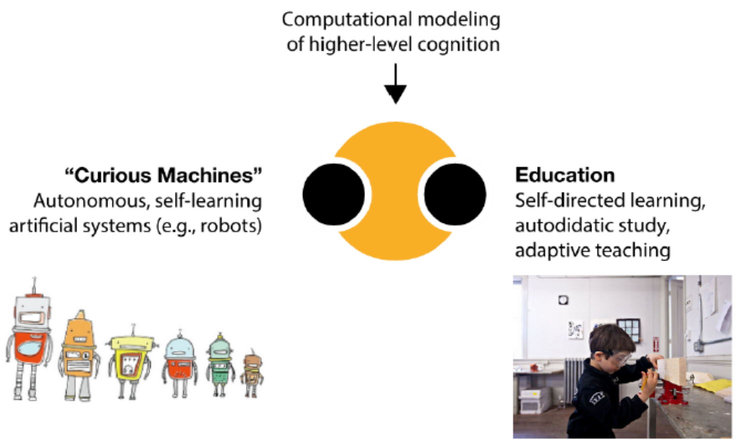

FIG.6 (Todd Gureckis)

Further, learning is not solely a cognitive pursuit. The ability for ICAs to engage with learners' social and emotional needs is also a hallmark of the best personal tutors. In addition, ICAs could provide much needed personalized educational experiences to those with cognitive or social impairments or other disabilities. Multi-disciplinary perspectives are necessary to designing and validating effective ICA educational interventions involving computational experts, educational practitioners (including teachers and clinicians/therapists), experts in how people learn from fields such as psychology/developmental psychology, neuroscience, cognitive psychology, sociology, and designers of highly engaging interactive technologies such as AR/VR, mobile, robotics, gaming, etc.

(2) Facilitating and enhancing common goal-oriented group activities and applications

What if an ICA could help you facilitate a meeting? Come to consensus on a difficult topic? Find mutual agreement in a tense situation? Or simply assist a group in finding collective goals, prioritizing among them, and producing joint work? Essential to understanding human capacity is the insight that we do not work alone but rather find our most powerful creative, productive and technical expression in groups. ICAs that do not interface simply with one-on-one human-machine interactions but instead support groups in their collaborative work can be most effective in meeting challenges that are bigger than any single individual in their local environment. One group even suggested a digital "Alex Trebek," host of *Jeopardy*, as a kind of model for ICA-as-facilitator. Work on automated facilitation was demonstrated in one of the presentations.

Not all groups are the same: we do not organize a company the same way we do a scientific collaboration, a classroom or even a collective. Such ICAs will therefore have to be well-schooled not just in individual human cognition but in group psychology, conflict management, organizational behavior, computer-supported cooperative work, and distributed cognition. They will have to not only organize and keep track of group tasks but also assist in prioritization, evaluation, and group completion of these tasks. As miscommunications and misunderstanding breed mistrust, such systems will not only need to develop trust in the system itself but also, concomitantly, trust among group members through facilitating strong working relationships, communicative channels, and visibility into distant group members' requirements and needs. In short, such systems will require "scaling up" the above considerations of human enhancement, engendering trust, facilitating interactions and incorporating interdisciplinary perspectives to the next level in order to support tasks requiring group attention and collaborative work [FIG.7].

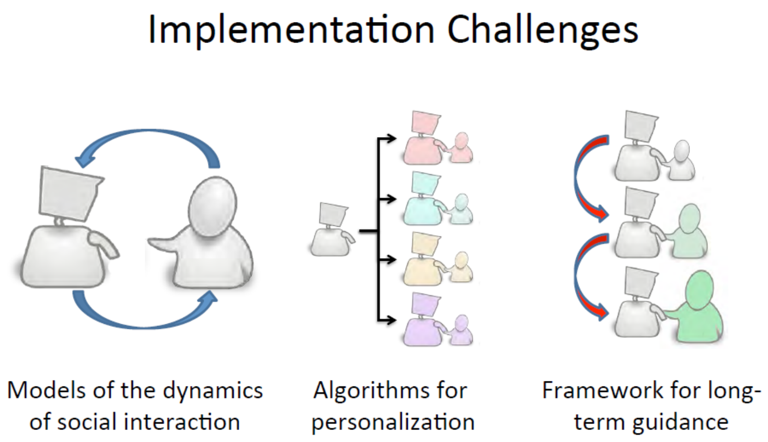

FIG.7 (Brian Scassellati)

(3) Elder care or facilitating "aging-in-place"

All three groups independently discussed elder care as a fundamental application domain for Intelligent Cognitive Assistants. Care for the elderly in an increasing global population is a wide-ranging social problem facing Western and other societies. Elder care requires addressing fundamental questions about human cognition, especially those that develop in a period of cognitive decline or change, asking not how we supplant or replace humans but instead how we can gently work to assist them with specific tasks or needs as their abilities gradually wane, and how we can support the broader human-technology care network. This requires addressing questions of social, behavioral, physical, and cognitive change, invoking system flexibility and learning. It requires "naturalistic" interactions and ease of understanding between humans and machines, and the development of trust through close interactions and support. It will also require a multi-disciplinary perspective, engaging not only computational experts but also perspectives from medical practitioners, nursing and palliative care, neuroscience, cognitive psychology, sociology, and potentially robotics.

A general 'process flow' was described for the most effective ICA solutions aimed at this domain. This includes the (1) unobtrusive capturing of data; (2) recognition of the environment, situations or common scenarios; (3) cognition or some system understanding of what needs to happen next, and (4) facilitating or taking action or corrective action. Such ICAs may focus on basic tasks like cleaning services and emotional awareness, managing household finances and medication regimes, personal care and accident prevention. Research required to enable effective ICAs in this domain includes topics such as multi-modal context awareness, communications capture, the building of relationship models, protection of sensitive content, and determination of intent and goals – all leading to the establishment and maintenance of mutual human-ICA trust and rapport. This also requires emotional awareness and support, whether through models of familial relationships or through situated communication, as well as the ability to detect intentionality and goals and possibly to develop a "personality" for the ICA in the very human-scale, intimate context of the familial home.

Additional research goal characteristics

Across these scenarios we identify the following unique and bounded areas for investigation. A selection of these should be addressed in grounded research focusing on the above domains or any other complex situation for ICA development and implementation:

- Dynamically-updateable scene comprehension
- Language comprehension, generation, and multi-modal interaction
- Reasoning capabilities, decision-making and pragmatics (a 'common sense module')
- Dynamic social comprehension systems for determining social context and group deliberation
- Ethical, emotionally-aware, value-sensitive and socially-aware situated intelligences
- Support for hardware and software services systems over long durations, across platforms and over time
- Studies and ethnographic work on human-ICA environments, augmented cognition, and/or "assistive" roles
- Modular structures for scalability and modalities
- Foundational scientific research and/or broadly-usable toolkits, APIs, datasets for system training, or ethnographic and qualitative insights into situated behavior
- Collaborative, proactive, and/or mixed-initiative agents (or multi-agents)
- Hardware or system innovations specifically geared for intelligent assistant systems, such as lightweight augmented reality equipment, chipsets, and all associated infrastructure support
- Ambient or embodied agents
- Resilient, redundant, upgradeable hardware for long-term use

NEXT STEPS

- Briefing at NSF in Washington, D.C. to review the highlighted workshop findings and specific recommendations (October 27, 2016)
  - Several NSF Division representatives and leadership will be invited along with DARPA, NiH, and IARPA Program Directors
  - Briefing will include a proposal to create a new research program framework

- **Industry and NSF stakeholders meeting** to draft a program outline which will be used to solicit funding commitments (before January 2017)

- Follow-up workshop which will focus on research gaps and needs for new algorithms and system architectural platforms to address the targets proposed in the ICA Workshop report (potentially aligned with the 5th Annual Neuro Inspired Computational Elements (NICE) Workshop, March 6-8, 2017, San Jose, CA.)
  - We are working with the NICE workshop organizers to determine whether this annual event will adequately address the 'technology-focused' follow-up workshop that was proposed during this ICA Workshop
  - This workshop would also deliver a report, which will identify specific research gaps, and then recommend whether an additional and synergistic research program to the one proposed by an ICA stakeholders meeting should also be created.

APPENDIX A

Potential future applications of Intelligent Cognitive Assistants

Personal tutor, which would help its human user learn (be it extremely young, school-age, disabled, adult, etc.), with the capability of understanding the human learner's perspective, and recognizing when its human subject didn't understand or wasn't paying attention, in order to optimize both speed and depth of the human's learning.

Group tutor, which would need all the capabilities of the personal tutor, but would also need to recognize social dynamics amongst the human learners, both in order to ensure that all members of the group were learning but also to help leverage the presence of peers and augment the overall learning experience.

Team assistant, which would serve as a personal assistant to teams in professional contexts, with a similar need to recognize complex social dynamics, in order to maximize team performance by augmenting collaboration, discussion, and decision-making among the members of a team, both for real-time performance as well as for long-term productivity.

Discourse moderator, which would serve a similar role as the team assistant, but in contexts other than professional teams, in order to help improve discussion and decision-making in these scenarios (public hearings, socio-political discussions, open forums, etc.).

Personal assistant/butler, which would augment its human user by sharing his/her perception of the world, answering queries, providing information, performing tasks, anticipating needs and goals (both short-term and long-term, explicit and implicit), understanding when interruptions and reminders were salient, and helping compensate for cognitive deficiencies in its human user (in terms of memory, attention or other aspects).

Elder-care/disabled/health-care assistant, a particular manifestation of the personal assistant designed to assist humans with particular health-care needs.

APPENDIX B

Critical challenges and tradeoffs

1. **Trust**
   a. **"Accuracy"**
      - Is the output offered by Intelligent Cognitive Assistants (ICAs) – e.g., the information shown or recommendations offered, and decisions made or actions taken, etc. – sufficiently accurate, and adequately sourced and explained, that the human user(s) can/will choose to trust it?

   b. **"Fairness"**
      - Is the output offered by the ICA sufficiently free of bias by other parties? (companies, governments, other ICAs, other humans)
      - Social justice and ethical considerations are essential

   c. **"Privacy"**
      - Is the data gathered by the ICA kept private?
      - If it is used to help train better ICA, is this done in a sufficiently anonymized way?

2. **Required capabilities**
   - ICAs will need to understand its surroundings – what they see, read and hear – not just **perception**, but also common-sense understanding of associated context, including **social context** (short- and long-term dynamics of human interaction).

3. **Maximizing desired outcomes while avoiding undesired outcomes**
   - Provide economic gains **without** exacerbating economic or social inequality
   - Augment "with-ICA" human capabilities **without** degrading "without-ICA" human capabilities
   - Improve collective capabilities **without** degrading individuality
   - Help make humans better **without** undesired changes in "what it means to be human"

APPENDIX C

Detailed notes from the first day of the workshop

In the **first keynote** [K1_Horvitz], Eric Horvitz began by tracing the long-term dream of computers that could serve as "a colleague whose competence supplements your own," starting from the interactive Whirlwind system from the 1950s, through the "Man-Computer Symbiosis" article of J.C.R. Licklider, and the first computer mouse from SRI. He introduced three critical pillars for bringing about the necessary "tighter coupling between man and machine." The first pillar involved "**Complementarity**" in which, by considering what machines can do to extend our capabilities in a number of ways, machine intellect serves to augment or fill in the gaps in human cognition as identified by leveraging results from cognitive psychology.

The second pillar was "**Coordination of Initiative**", where man and machine can tackle different components of a particular problem yet communicate in a fluid and competent way, through a conversation: a continuous process of back-and-forth contributing, signaling and monitoring. Eric's third pillar was "**Building computational models of people and the world**", requiring both probabilistic models of the world, but also models of how humans are viewing, thinking and reasoning about the world around them. An example of this was not just prediction of traffic patterns, but prediction of what would be unusual or surprising to human users, along with the costs and benefits of interrupting a user with alerts or reminders. Another aspect discussed were life-long cognitive assistants that share, understand, and appreciate the human's memory landmarks.

Eric discussed tools available to researchers to help close the gap between man and machine, including increased computational capabilities, improved methods for learning, inference and representation, work on causal inference, perception, Natural Language Processing, improved capabilities for multisensory fusion, and integrative pipelines (image interpretation & captioning) [FIG.8]. He discussed opportunities for machine intelligence for coordination of machine learning, for coordination of physical tasks (such as surgery), to recognize commitments to others, to leverage large-scale datasets, and to address specific health-needs. He described life-long cross-device assistance enabling long-term planning, and integrative AI that merges speech, vision, and NLP to enable situated systems: human-centric computation anchored in the physical world.

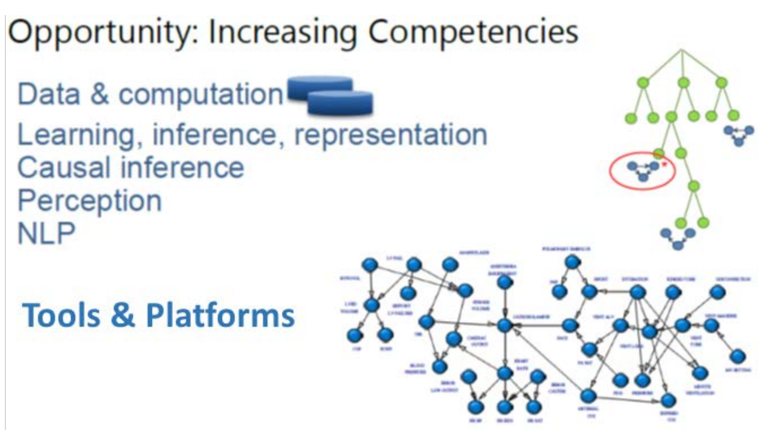

FIG.8 (Eric Horvitz)

At the end of his talk, Eric posed a few questions: How can we complement & extend peoples' abilities? How can we assist people with perception, learning, recall, inference, decision making, and planning? How might we leverage models of human cognition? What are key clarifying & illustrative scenarios? For whom and for what purposes? What new functions & services are most promising for composing into personal assistants? What sensing, effecting, and communication hardware would enable core capabilities? [FIG.9]

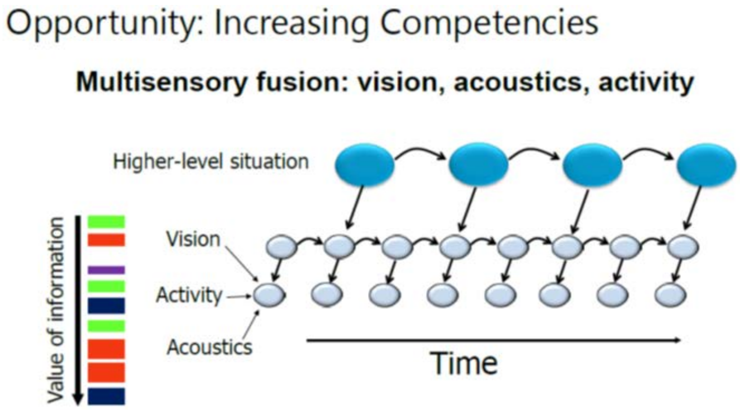

FIG.9 (Eric Horvitz)

Discussion included comments (Gary Marcus) on how accuracy requirements vary between applications (low accuracy ok for search, bad for self-driving cars), and the need for strong natural language understanding and common-sense reasoning. Eric's response was that progress in domain-specific systems would help lead to the advances needed for systems capable of wider-domain. Another commenter (Liz Spelke) raised the point that humans could also change as they interact with these systems, and that research into this will be both important and challenging. A third comment (Jaron Lanier) raised the importance of developing systems that can understand and respect the "edge cases" (for example, individuals whose preferences don't correspond to what most people prefer). A final comment (Ron Kaplan) raised the idea of "forgetting" things (that aren't needed or are painful) as well as reminders.

In the **second keynote** [K2_Breazeal], Cynthia Breazeal aimed to broaden perspective on the **role of cognitive assistants in society**. She began by observing the strong competencies of humans for collaboration, which she traced not just to the analytical thinking capabilities of the human brain, but even more so to its "social thinking" capabilities. And she pointed out that since we are wired to harmonize with each other, such social interactions are a large part of what makes us human – thus how Intelligent Cognitive Assistants interact with us and support us matters. After re-raising the importance of respecting cognitive diversity, she spent some time talking about potential impacts on the future history of our society and its economics, drawing from the book "The Second Machine Age" by E. Brynjolfsson and A. McAfee. Here, digital technologies were described as a "General Purpose Technology" – a major economic inflection point which interrupts and accelerates the normal march of economic progress by significantly boosting productivity and outputs across many, if not all industries. Driving forces for this include exponential improvement (Moore's Law, etc.), pervasive use to produce more with less (including with fewer humans), recombinant innovation through reusable building blocks, and global interconnectivity on a massive scale. While this creates "bounty" or wealth, it also seems to increase "spread" – differences among humans in wealth, income, mobility and access to the advantages of this bounty. Cynthia raised the question of whether ICAs would exacerbate or mitigate "spread," since some jobs will be displaced while others are enhanced. She asked "Who benefits from our brilliant machines?"

Cynthia then discussed key societal challenges that could be addressed with ICAs. These included inequality of education, chronic health conditions (cost from 2016-2030 projected to be $42 trillion), and global aging (leading to fewer caregivers-per-senior-citizen). After mentioning the importance of identifying long-term research that could complement the large commercial investments being made in personal assistants, she then provided an overall characterization of the potential applications that were proposed for the workshop (termed as "Moonshots"). The target group-size ranged from 1-on-1 relationships between an ICA and a primary human user, to ideas that involved a single ICA aiding a small group (project team, family). Cynthia also felt that large scale applications, perhaps for environmental challenges or self-governance, might call for large teams of ICAs working with large populations of people. Cynthia pointed out that length of engagement was another aspect for defining the ICA ideas, ranging from short-term "Focused engagements," to mid-term engagement over weeks and months, to ICAs capable of life-long interaction. She then further distinguished these ideas into ICAs intended for "normal healthy" adults or kids, to those focused on humans with developmental issues, with disabilities, or experiencing age-related issues.

Cynthia then described three categories of ICA ideas: the "**Assistant/Butler**", the "**Cognitive Orthotic**", and the "**Mentor/Coach**" [FIG.10]. She described a number of challenges: privacy and security, the use of these capabilities for criminal purposes, the unintended consequences (how will our biological brains cope with sudden superhuman abilities), social justice (exacerbating social inequity through limited access to those wealthy enough), ethics (our responsibility to each other as being part of a community vs. letting a machine do it), and who really benefits? In parting, she asked "How can our ambitions of future ICAs support our human values to enable a positive path for society that we will be proud to make happen?"

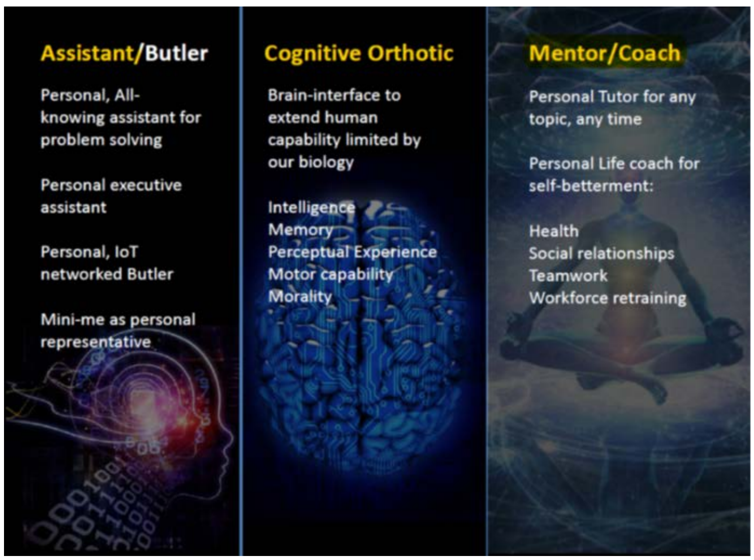

FIG.10 (Cynthia Breazeal)

The first commenter (Brian Scassellati) pointed out that most recent improvement in personal assistants has come by getting better at getting input into the system, as opposed to getting better at reasoning about that information once it is inside. Our systems are better at Natural Language Processing, but we still use it for scheduling. Second commenter (Winfried Wilcke) asked, on the topic of bounty and spread, what portion of humanity might be happy if bounty were sufficient that they did not need to work. Cynthia pointed out that people want to feel that they are doing something meaningful, implying that there will still need to be a balance between what machines do for us, and what we as humans still want to do. Later, Jaron Lanier pointed out that the large datasets these systems depend upon are contributed by humans, and he suggested that such contributions could potentially be compensated by micro-payments, if the economics could be worked out.

Another commenter (Bruce Horn) asked about the possibility of an "arms race" between cognitive assistants with opposing intentions (spam-blocking vs. delivering marketing content). In response, Cynthia raised concerns about ICA uses that are good for large companies vs. good for society at large. Jaron Lanier also pointed out the potential importance of helping to bridge language and cultural differences in helping people address their aging and chronic disease issues when those people are themselves recent immigrants. Mike Roco asked Cynthia if her different categories of assistants would require completely different types of assistants. Cynthia felt that commonalities could potentially emerge as initial ICA systems were developed, but that a single general-purpose AI might not be necessary. Ece Kumar asked about computer assistants for humans at age 0-3, and how one can test the impact that such systems are having, and ensure that these impacts are appropriate. Cynthia felt that it would be critical for all involved parties (companies, researchers, parents, etc.) to keep both benefits and drawbacks in mind, with a tight iterative development and improvement cycle. In a parting comment, Gary Marcus suggested that universal literacy by age 4 – an outcome for which the benefits are likely to far outweigh drawbacks – should be quite doable.

Panel 1A then opened the first of two panels on "**Moonshot Proposals**", or potential usage scenarios for ICAs. Saul Perlmutter proposed "**Assisted Deliberation Tools**" [O1A_Perlmutter], intended "to scaffold deliberations, discussions, decision-making and design processes in ways that are likely to yield productive, interesting, multi-party-satisfying outcomes." He proposed this for groups at all scales, ranging from search committees up to international policy decisions. Saul felt this would be useful to parse factual & logical issues from values, goals, desires, and fears, and that it would important to bring in the right parties to the discussion (representative samples of concerned parties and minority subgroups, as well as an effective panel of experts). He envisioned an interface that would bring new parties "up to speed" rapidly. Concerns he listed were the need for transparency (all parties would need to feel the ICA was being fair) and confidentiality, but Saul felt that ICAs offered the potential for impartiality, confidentiality and scalability that would not be possible with human moderators or with crowd-sourcing. He listed a number of activities that such an ICA could help with or take over, from clustering and simplifying common issues, to finessing logjams either by identifying focus areas, by introducing new information or even by imposing new modalities (video, in-person discussions).

Ken Forbus then proposed ICAs as "**Assistants for Learning Science**" [O1A_Forbus], to address the need for more teachers, tutors, and teammates with the right availability and personal insight. Such ICAs would serve as tutors, coaches, partners, and mentors that support people who want to learn any area of science, at any level, and at any time. Ken described these ICAs as stretch goal for 2050, as part of a suite of tests to replace the Turing Test. Challenges he listed included the need for such systems to rapidly learn by reading a small number of examples, to be able to digest multiple modalities (text, images, diagrams, gestures), to be able to have dialogues with students that could include culturally-relevant examples (which would change as pop culture changes), and to be able to improve the learning transfer from ICA to human by building upon relationships built up over weeks, months and years. Ken also pointed out the current educational opportunities for humans to learn as they teach the ICAs.

Bruno Olshausen [O1A_Olshausen] then spoke about **how our brains absorb and make sense of all the information** that comes into it, focusing on the visual system. He spoke about the difference between the sheer amount of data arriving into the brain from the optic nerve and the small amount that makes its way into persistent representations in the areas beyond the V1/V2 visual system. He proposed an ICA that would capture and retain sufficient information so that a person could "replay" their experiences at a much later date in the future at a high level of detail.

Janet Vertesi [[O1A_Vertesi] then proposed an "**assistant that would understand team dynamics**, organizational structures like hierarchy, and other processes for supporting collectives and organizations." (Note that the WebEx did not capture her slides as she presented them – her slide-deck is available from the same SRC site as the recording, linked to her name in the agenda. See also the citation list at end of this section.). Instead of a one assistant per person model, Janet asked what a socially-aware system would need to know to support teamwork. Such an ICA would need to understand authority and hierarchical organization structures, and to interact over short and long distances and time-scales. She described a few experiences studying teams, ranging from those oriented towards consensus under a single leader, to those oriented towards fairness, and addressed the role of affinity between human and robot. She described experiments with hierarchical and consensus teams in which the hierarchical teams accomplished more (found more objects, covered more ground, issued more "command sets"), but the consensus teams made fewer mistakes. She also observed that how the team informally organizes is just as important as, if not more important than, the way it may be formally organized.

Janet proposed that any ICA must capture the variation in these teaming structures, providing support by aligning group social norms with technical design to achieve collaborative goals, and by “patching the holes” (in the group) without adverse impact on immediate user needs and workflow. Group social norms can interact with technical design from hierarchical (favor doctors’ orders over nurses’), to lateral teams (integrating knowledge while preserving autonomy), to facilitating consensus-building. “Patching holes” includes reducing errors induced in top-down decision making, balancing input from and outcomes for all the human contributors independent of gender or race, depersonalizing conflict by making pressures in other locations visible to others, and avoiding “group-think” induced by strong social ties.

John Laird [O1A_Laird] then proposed a “**Neural Intelligent Cognitive Orthotic**”, in order to reduce the latency and need for physical action (pushing a button, etc.) in interactive communication between human and an ICA. Such orthotics would have access to everything our brain perceives, and would communicate seamlessly with the brain. While John described this as “going to an extreme,” he felt it would be the best way to completely achieve the desired complementarity between human and machine.

During the comments, Gary Marcus commented that he felt that we would need AI to understand the neuroscience well enough to be able to build such a cognitive orthotic. John Laird responded that he felt such an orthotic would not necessarily require a complete, “molecule-for-molecule” understanding of the brain. Eric Horvitz pointed out that there are professional researchers who are already studying how humans interact with and are influenced by information, either on the web or with medical information.

Janet talked about a “critical news-bot” that she was working on that incorporates news from a diversity of sources, and tied this to Eric’s comment and Saul’s presentation, saying that “good decision making” does depend on having information from a variety of source. She stressed the problems imposed when ICAs “make too many decisions about what we’re going to like and what we’re going to see,” because that can “lock” humans into a single pre-computed frame of reference through “confirmation bias”.

Blaise Aguera y Arcas expressed concern about the difficulty of moderating debate when a subset of the participants do not share the same sets of common principles, in particular in terms of no longer needing to base their positions and opinions on actual facts. He then pointed out that e-democracy “hackers” have helped mediate conflicts by “hacking the forum rather than by hacking the semantics.”

Saul expressed optimism due to the success of “deliberative polling” in facilitating deliberation among small groups chosen randomly from the population. Janet connected this to the importance of organizations in decision-making and deliberation. Mike Roco pointed out that one of the issues with the Cognitive Orthotic is that some observers would view this not as the use of cognitive machines to aid humans or even to improve humans, but would have a strong and visceral negative reaction, viewing the Cognitive Orthotic as such a fundamental change that is would be effectively “redesigning” humans. (In later discussion, Mike made clear that prior research programs containing a small component that triggered similarly strong responses tended to invite enormous backlash, even if the component were a trivially small part of the overall program. Thus in this comment during the discussion, he was trying to convey the potential costs – in expected backlash – of including such a component in any research program on Intelligent Cognitive Assistants.)

After lunch, **Panel 1B** was the second panel on “**Moonshot Proposals**”, or potential usage scenarios for ICAs. Ron Kaplan [O1B_Kaplan] described a **wish-list for a personal assistant** that would have situational awareness and a strong understanding and even anticipation of his likes and dislikes. Observing that “ubiquitous computing comes together with ubiquitous complexity,” Ron pointed out that effective collaboration was essential to make ICAs an improvement and not an additional burden. This will require alignment of consistent goals between man and machine through sufficient back-and-forth discussion, despite the inevitability of misunderstandings/ambiguities and evolving circumstances and beliefs and desires. Ron pointed out that since perfection in ICAs is likely “not possible, expected or required,” it will be critical for the man-machine interaction to be able to “easily repair” the collaboration because the ICA won’t be perfect.

Todd Gureckis [O1B_Gureckis] described his interest in how “**humans learn by tinkering**” with the world around them, using “curious machines” that perform autonomous, self-directed investigation of their environment [FIG.6]. Todd pointed out the difference between the simple classification question (“What is that?”) which drives deep machine learning, and the much more complex questions that children ask (“Do all dogs have tails?” “How do alligators and crocodiles differ?” etc.). Todd made the case that rather than ICAs that just answer questions, we may need ICAs that are capable of asking questions. He described how psychologists model and understand the benefits of self-directed over passive learning (“Fundamental Sampling Dilemma”), in which humans actively seek out “edge cases” (rather than waiting for these rare events) that then speed their learning. Challenges are that asking good questions at the right time requires some amount of knowledge (particularly social information), requiring strong Natural Language Understanding and understanding of context. Short-term progress possible would be if applications that could resolve their uncertainty about goals by asking (A map-directions app could ask “Are you stopping for gas?”), with longer-term progress by asking efficient questions in the context of scientific research.

Misha Pavel [O1B_Pavel] then spoke about “**Model-Based Intelligent Cognitive Assistants**.” He started by describing our human cognitive system as a “legacy system” with imperfect perception of risks and probabilities, limited attention and learning rate, and a host of neuropsychological issues. Misha described two broad categories of potential ICAs: either for augmenting/amplifying human cognitive processes, in stimulus selection, decision suggestion, reminders, and cyber-physical interfaces and robotics; or for improving/training human abilities. Misha stressed the importance of modeling both the world exposed to the human, as well as how his/her brain reacts to those inputs. Later, he affirmed Ron Kaplan’s comments on the need for models even if they are not perfect.

He described challenges in characterizing, representing and inferring the cognitive state, affective state and social influences of humans, and in predicting probable behaviors and resulting cognitive and affective states. He also pointed out the need for minimally-obtrusive sensors capable of measuring the physical, physiological and mental state, for a framework for rigorously characterizing context, for modeling and predicting brain-state dynamics, for stochastic models of dynamic utilities and preferences, and for robust detection of anomalies and incongruences (not just outliers). [FIG.11] Applications he proposed included Intelligent and context-aware advisory & alerting systems, assessing and training creative problem solving abilities (for instance, for training of intelligence analysts), precision healthcare and medical decision making, optimal personal (precision) coaching and teaching systems that would be aware of students’ knowledge, motivation and affective states, interpersonal communication support systems, visualization systems for high-dimensional data presentation, and even artificial humor applications.

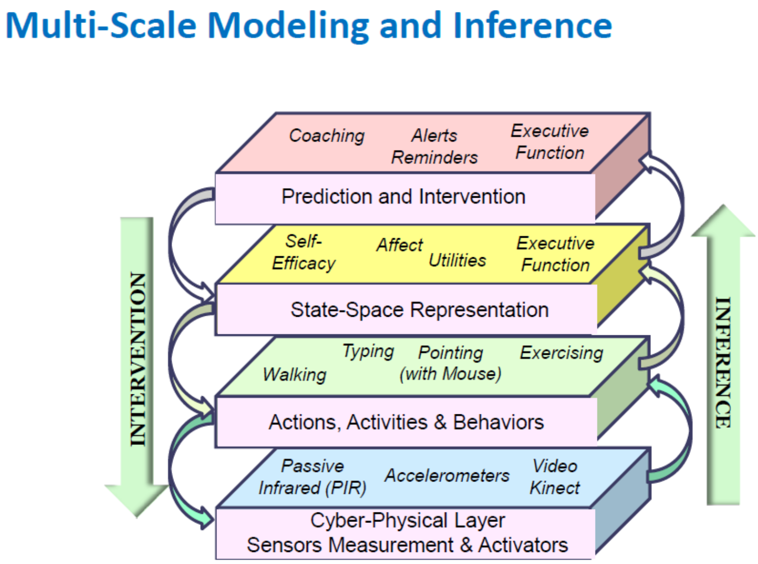

FIG.11 (Misha Pavel)

Jan Rabaey [O1B_Rabaey] proposed a "**human intranet**" – a family of wearable sensors that would introduce different ways of actuating and interacting with the world around us. He felt this would impact humans through "extrospection" (interacting with the world around us), "introspection" (monitoring our own bodies and interacting with them), and "extension/enhancement" (additional or augmented sensors). He described progress in flexible sensing, energy storage and generation, and miniaturization, but pointed out the need for additional improvements in energy efficiency and networking, in handling enormous amounts of data at high bandwidth and low latency without creating power, safety or security issues.

Vijay Saraswat [O1B_Saraswat] described **cognitive assistants in professional settings**. He described a cognitive task landscape that impact humans "at work" [FIG.12]. One example he described was the need for ICAs that could help business and employees comply with regulations, which could help jurists draw upon legal precedent, and that would be able to operate as a professional assistant with the mastery of professionals in the field. One of the challenges he addressed was the long "ladder" that professionals must climb, so that any relevant ICA must somehow accumulate similar expertise. Such ICAs must also be "general purpose," in the sense that one set of hardware and learning algorithms must be able to address different professional domains, in order to "know deeply," "learn continuously," "interact naturally," and "reason with purpose."

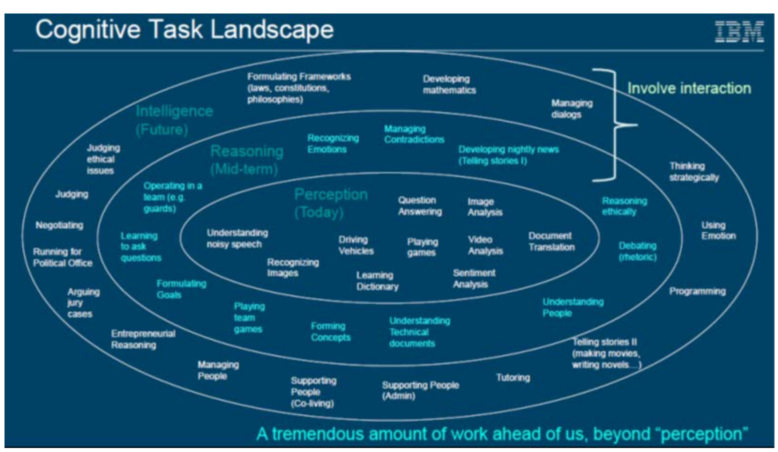

FIG.12 (Vijay Saraswat)

In the first comment to Panel 1B, Ron Kaplan pointed out that professional domains may actually be easier than children’s stories, because of the presence of common-sense reasoning. Gary Marcus pointed out that some professional domains might indeed be easier, while others (legal reasoning) might require both professional concepts and common-sense reasoning. Vijay pointed out that accounting would similarly require similar common-sense reasoning.

A question came to Todd about the importance of intent behind the questions that both humans and machines would ask, and Todd replied that both goals and sub-goals are critical aspects of his and his colleagues’ research. Eric Horvitz asked Todd about the difference between asking questions just for an immediate need vs. questions that might assist an ICA in engaging in long-term, life-long learning. Todd said that his team had observed that humans presented with a relatively abstract task might ask concrete questions that provide the illusion of rapid progress but which don’t actually bring them any closer to actionable information.

After a break, **Panel 2A** moved from ‘potential usage scenarios for ICAs’ to “**Implementation Challenges & Tradeoffs**.” Gary Marcus [O2A_Marcus] started with two premises: ICA are going to need to read and comprehend unstructured text, and are going to need to comprehend visual scenes reliably. He then pointed out that while there has been exponential progress in focused domains (chess, Go), we still have a long ways to go to demonstrate a general-purpose cognitive agent. He felt that one key obstacle was “**The Long Tail Problem**”: that there is often lots of corpus data for a few common examples (that are then easy for many systems), but very little data for less common examples (which are then quite hard for these systems). He then discussed how children learn not by focusing on correlation but on causation (“how” and “why” rather than “how likely X given Y”) [FIG.13a,b]. He then came to his main questions: **are we headed towards a local maximum in performance** (through larger and larger data sets), without making direct progress on systems that are capable of representing the richness of human comprehension? If so, what should we do about it? Could a richer understanding of the mechanisms of human cognitive development and common-sense reasoning help? Gary then suggested that more **research at the intersection of cognitive development and AI** might prove helpful.

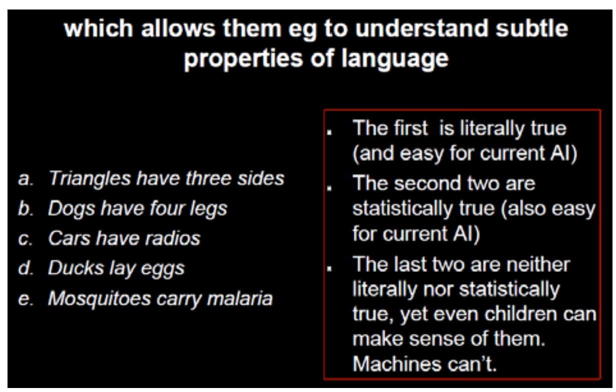

FIG.13a (Gary Marcus) FIG.13b

Bruce Horn [O2A_ Horn] briefly described his work with **Smart Devices**, and described his criteria for ICAs: they must know useful things, learn from experience, manage in a dynamic world, and explain their actions and reasoning [FIG.2]. Such an ICA must handle real-world, online learning with sparse data, yet today’s ICAs are usually demonstrated after a “cold-start” with offline learning and massive data. Knowledge must be acquired and transferred from its repositories, which often means written text, which needs to be accessed by the machine. ICA decisions will need to be explained to human users, including where the underlying information came from. Bruce asked how tacit, implicit and experiential knowledge gets encoded, and about trust and agendas.

Ece Kamar [O2A_ Kamar] talked about the AI challenges inherent in ICAs, in using computation to augment what humans are capable of today. She felt the **biggest challenges were at the interface between humans and**

agents, including modeling of user context, beliefs, intentions and goals; division of tasks and transfer of control between ICA and human, methods for supporting communication between the two (tradeoffs between benefits and costs of interruption), and how the ICA should learn from both human input and from its own mistakes [FIG.3]. She pointed out that these challenges are in fact common to teamwork and collaboration between teams composed entirely of humans. Thus intelligent agents should be designed as team members to their users.

In the same way that we do not expect human assistants to be perfect from day 1, we can clarify for ICAs when they confused, providing them feedback, coaching, and instructions. Ece suggested that if we can develop a trusted, respectful collaboration between ICA and human, this can create a virtuous cycle for the agent to learn from human input to get better. She then pointed out that trust in this partnership is essential, and that not all errors are equal. This can be particularly problem when expectations are too high.

Liz Spelke [O2A_ Spelke] started with a **description of humans as a technological species –** we observe the world around us but we also change it significantly. She suggested that while one might think this is because humans are good general-purpose learners, her experience with infants would seem to indicate otherwise. Infants have good understanding in objects and their motions & interactions, a rudimentary understanding of numbers, their own location, the shapes of objects, the causal impact of their own and others’ actions, and appreciation of the other humans around them as social beings who engage and communicate with each other and with them. However, she pointed out that these infant-brain capabilities are themselves shared in other creatures, and seem to be the result of innate special-purpose systems within the brain, rather than being the product of a single general-purpose learner.

Liz then asked what kinds of Intelligent Cognitive Assistants will allow humans (as flexible learners building upon a common suite of special-purpose cognitive tools) to flourish? She proposed that any ICA that works reasonably well could potentially work for a wide variety of humans, given the commonality in underlying cognitive capabilities across humans in many different settings. Human cognition proved flexible enough to have adapted effectively to past advances in technology such as books, movies, remote controls, telephones, and toys. However, Liz pointed out that those advances occurred slowly, and were likely adapted to our intrinsic capacities rather than the reverse. She wondered whether we could be equally confident about our abilities to adapt to future advances in technology.

She then posed several challenges for **how ICAs might adversely affect humans**. The first challenge she stated as “What are the consequences of off-loading aspects of our basic cognitive tasks like navigating or remembering things?”, Then she restated it as “How can we find the right kinds of ways of addressing human cognitive limitations, so that we can enhance those capacities as opposed to encouraging people to turn off those capacities and allow the machine to do those things for them?”

She then asked “What are the consequences of getting information from different places at once?” in terms of affecting our human “sense of place” when information and sensory data is simultaneously arriving from multiple locations. “What are the consequences of interacting with entities that share some of the properties of ordinary objects, agents, and social beings, without being any of these things?” [FIG.14]. She wondered how kids who grow up interacting with ICAs are going to be affected by these experiences.

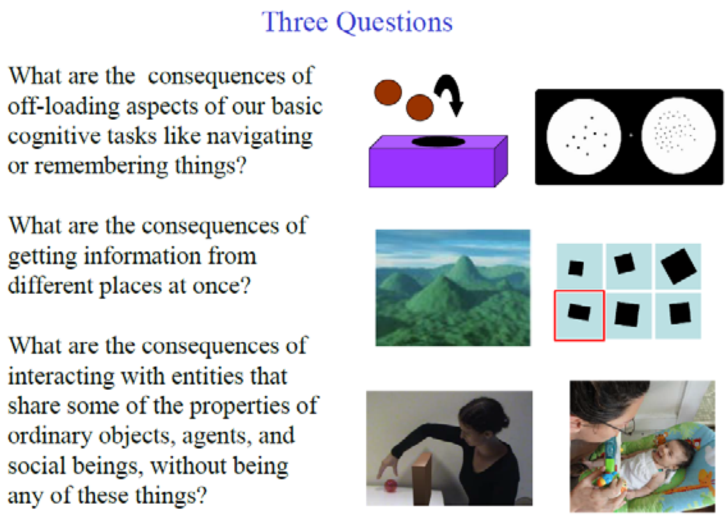

FIG.14 (Liz Spelke)

Brian Scassellati [O2A_ Scassellati] started by showing a number of examples from his work at the intersections of **AI, Robotics and Psychology**, using these to make the case that ICAs could affect much more than a scheduling assistant, affecting everything from manufacturing to education of adults to teaching sign-language to deaf 6-month-old infants. Brian felt that the essential challenge was going to be understanding the social environment in which the ICA will have to function.

In his work teaching kids with robots, he has found that humans engage robots differently than they engage with human teachers, including lack of social anxiety (about making mistakes while learning) and complying with requests without resentment. Since we can shape the role of the robot to our advantage, we can generate controlled situations, ranging from peer-to-peer interactions to demonstrations of mastery. And since a teaching robot as a physical agent becomes a cognitive artifact to the human, embodiment can impact learning substantially. Brian listed some challenges, including the need to model the dynamics of social interaction, to create algorithms for personalization (an ICA-human relationship that grows over time), and to plan for long-term performance and achieve a long-term goal. [FIG.7]

In the first comment, Jaron Lanier suggested that researchers may need to be more careful about raising expectations and misunderstandings in the way AI research is communicated to the public. Greg asked Liz whether humans are cognition-ready systems at birth ("nature"), or whether we learn these capabilities along the way ("nurture"), to which Liz responded "The answer to your either-or question is Yes." She then pointed out that babies are born with some cognitive tools, but then use these tools to learn from a rush of data and build better cognitive tools.

A relevant follow-up comment concerned the negative impact on the quality of questions that children visiting a museum would ask after an iPad application was updated to summarize for them what they had learned – rather than improving their ability to ask good questions, this additional summarization made the questions worse.

The next question concerned **the role of emotion in ICAs** – one could add it to increase effectiveness, but at the risk of it being perceived as "fake." Brian's response was that humans will attribute emotions to an embodied ICA, whether the designer intended to convey emotion or not. Eric then asked about stages of plasticity in neuroscientific development. Gary responded that while textbook summaries of the literature seem to suggest different plateaus of neuroplasticity, the actual papers are much less clear – and thus adult's brains can be just as plastic as children's, save only that the changes might need to occur in smaller steps. Liz responded that while after a certain age, a cat's brain might not be able to fuse binocular vision – but that cats reared in the dark would have the tools for binocular vision. She remarked that while it might seem that putting a chip in someone's brain is "affecting their brain," any human that interacts extensively with an ICA for education or training is also "affecting their brain" as well.

In the **third keynote talk**, Blaise Aguera Y Arcas [K3_AguerayArcas] began by tracing the **origins of "computing and neuroscience**", before these origins diverged. He pointed out that Turing wrote both about neural networks and about the Turing machine (Turing tape) in his 1948 paper. Blaise described Artificial Intelligence systems "based on symbolic approaches" as "useful but not intelligent" and contrasted these systems to the perceptual intelligence and motor-control of mammals such as chimpanzees.

He observed that logic and language capabilities in humans are performed "steampunk-style, using this machinery that was definitely not evolved to do it." He then returned to discuss neural networks, rapidly tracing progress in hierarchical feedforward networks, recurrent networks, and reinforcement learning. While he made a comparison between the emergence of Gabor-like filters and specialized "Jennifer Aniston neurons" in large feedforward neural networks, he also pointed out that there must be more going on in the human brain, since these networks "cannot reason, cannot generalize, and they take way too long to train." He pointed out that since these systems require lots of training, they require lots of data, which gives a decided advantage to companies that have access to large datasets.

Blaise described an "embedding approach" for recognizing photos of landmarks based on a convolutional network of moderate size (millions of parameters) that is able to out-perform much larger models by optimizing on "triplet loss" (e.g., "these two photos are both of the Eiffel Tower and this other photo is not"). He described the dimensionality reduction shown by this kind of network as critical to building future intelligent systems. Blaise also pointed out that when he and his colleagues are able to train networks that can identify the year in which a photo was taken, it is not clear to them exactly what information from the photos these networks use to achieve this result. One way his team has attempted to improve their understanding is by **asking trained networks to "dream**". Given a network, trained by optimizing the weights *w* to generate the target *y* from input image *x*, instead they lock the weights *w* and ask the network to optimize for an input *x* given a target y. He then showed a number of results using such "dreaming" capabilities, using networks optimized for faces, for ImageNet, and with twentieth-century poetry. Blaise then discussed machine translation, evolving from structural to statistical to neural techniques.

He then addressed privacy issues that he viewed with ICAs, which he characterized as "X-men vs. Borg." He felt that one approach that would respect privacy issues was "federated optimization," in which learning on individual devices is fed back to the cloud. He discussed the importance of latency, energy, connectivity, and privacy issues, and felt that such a distributed learning approach would help address biases (such as racial, gender, and socioeconomic biases) in training datasets.

He talked about the difference in structural trust for software programs (e.g., programs such as Microsoft Word have historically always been trusted) versus the trust and structural guarantees concerning the usage of data (as stored on corporate servers either explicitly or implicitly by transmission from a smartphone). Blaise drew a distinction between ICAs that are simply commanded with voice, between personal assistants (back and forth discussion with a 'you' and a 'me') to augmentation "where language is just one of the modalities in which we converse with our technology."

Blaise waxed philosophical in the last section of his talk, pointing out the drawbacks of a paranoid and acquisitive mindset for humans and our place in the world, in terms of fear of technology, and over-consumption of limited resources by over-population. He called for more self-awareness about these drawbacks. Drawing upon an earlier presentation by Rich Sutton, he broke all of history into the "Age of Physics," the "Age of Replication," and the "Age of Design," and asked "When we don't have to struggle for survival, then what? What do we want to be when we grow up?"

In the first comment, Gary Marcus asked about the significant remaining gap between the results Blaise showed (in terms of "producing" Shakespeare) and what Gary felt was the lack of higher-level representations that would be needed to enable true common-sense reasoning. Blaise agreed, saying that he did not believe that "more of the same" was going to bridge this gap. He observed that the emergence of a feature hierarchy and other similarities to what we observe in the brain was interpreted by him as proof that we have in hand at least some part of what would prove to be the overall system needed for such capabilities, the "first breakthrough in what will need to be a long chain of such breakthroughs."

The next commenter affirmed that Blaise's talk had not really addressed "the AI problem," to which Blaise responded that he felt that such problems and the perception work he had showed were not in fact distinct separate problems. He felt that the emergence of semantic maps within translation systems was one such example of this. He went on to say that he felt one of the mistakes of earlier AI approaches was to attempt to separate perception from binding to concepts.

Vijay Srinivasan asked about the possibility that ICAs could create problems by moving humanity from an era where each human is quite similar to each other, to one where each human is quite different from his/her neighbor. Blaise responded that he was encouraged by "the expanding circle of empathy" as technology has been empowering humans and moving them out of the need to survive solely by self-replication, but worried by the rise of "radical polarization of all sorts," in terms of economic inequality and culture wars. He agreed that ICAs could become sources of even further differentiation, but could not predict where this might lead, saying "this is a choice we have to think about pretty hard."

Jaron Lanier pointed out that a common aspect of anti-modernity, where humans view technological progress as fearful, are afraid of losing memory or continuity from their past experiences. Blaise agreed, saying that he was "a humanist, not a post-humanist."

Ken Forbus asked whether the visual relational representations in Blaise's systems would be suitable and sufficient for reasoning, not just classification. Blaise responded that the combination of his perceptual systems with fairly straightforward additional models should be capable of implementing at least simple reasoning tasks, although probably not the most difficult such tasks.

Janet Vertesi commented that although sci-fi movie themes about "robot overlords" may be overblown, many humans are in fact losing their jobs to computer automation. She pointed out that the trend of decreasing form-factor and thus increasing computation (and thus increasing personal data) in the cloud, posed a series of security risks and privacy concerns. She also pointed to "the rise of the personal data economy" as the source of the large datasets that power modern deep learning. She asked if there were other approaches that could reduce the need for accumulation of so much personal data.

On the topic of ICAs for job-augmentation rather than job-replacement, Blaise drew on studies of the disruptions introduced during by Industrial Revolution. He worried that there wouldn't be enough service jobs to compensate for the replacement of information jobs, and felt that the only viable long-term choice was some sort of "guaranteed minimal income," despite the political difficulties in putting such a system into place.

On the topic of **data privacy**, he felt that the federated learning approach would help reduce the need to centralize personal data, but that this still required the need to aggregate statistics while maintaining security and privacy. Any hub-and-spokes model might lead to too much power at the hub, but any distributed approach would have to be able to show the same or better performance as a centralized approach.

In the **final panel (2B)** of the first day, continuing with the topic of "**Implementation Challenges & Tradeoffs**", started with Alberto Martinez [O2B_Martinez]. Alberto described how the reduction of power-per-compute has impacted computing, describing the interplay between increased device density (which then provide more sources of data), natural language interfaces providing the need for increased computation, and the increased importance of data security and data privacy. He pointed out the need to rebalance away from the bias towards commercialization, citing Alexa as an application that is always attempting the sell him something. He predicted that "App Stores" would evolve to "**Skill Stores**" where ICAs would drive a business model for "harnessing the development of capabilities for profit." He felt that this opportunity would be accompanied by challenges, such as the need to balance economic interests against the increased societal pressures. He proposed an '**Open Data model**', with ImageNet as an example where non-corporate-hosting of data-sets helped drive innovation. [FIG.15]

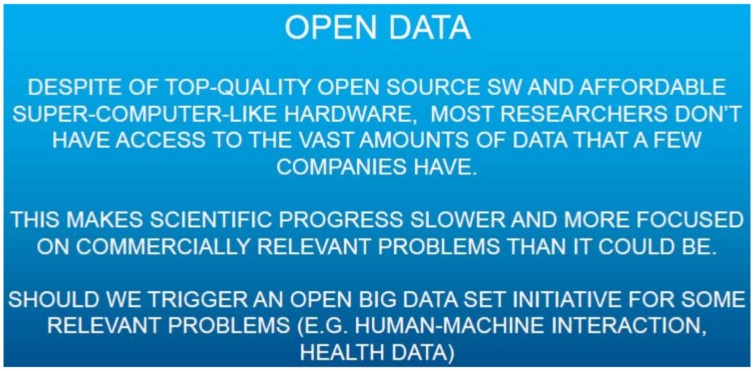

FIG.15 (Alberto Martinez)

Murray Campbell [O2B_Campbell] shared some observations from the continued evolution of computer chess since DeepBlue ("19 years ago today.") He showed a graph of chess ("Elo") rating for humans, for computers, and for **human+computer teams**. All three curves have increased over the past 35 years, with computers catching humans in the late 1990s, plateauing for five years and then increasing well beyond human capabilities. However, he pointed out that while humans+computer teams continue to outperform computers, in the context of tournament chess ("only 3 minutes to make a move"), computer-only implementations are getting to the point where the human doesn't add anything to the capabilities of the computer by itself. [FIG.16]

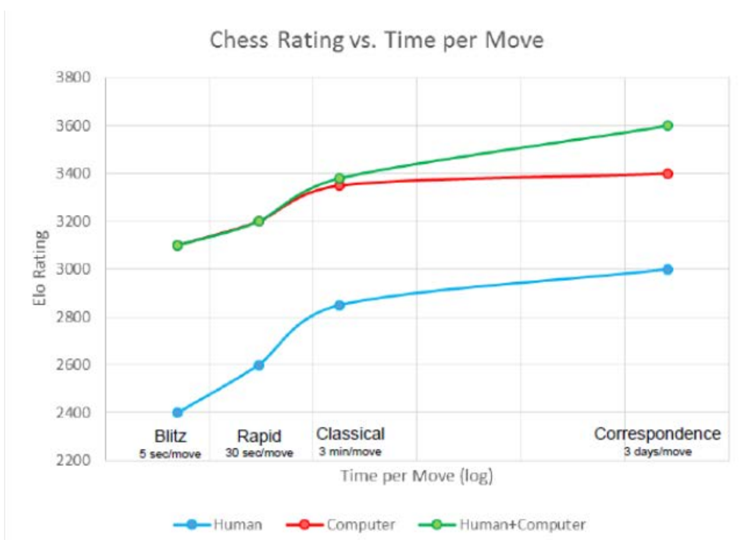

FIG.16 (Murray Campbell)

That said, Murray pointed out that the presence of chess-playing computers has been augmenting and improving human performance for the past 30 years. He also pointed out that even in chess, there are aspects at which humans continue to out-perform computers – one example was recognizing "fortress" formations. He also pointed out that humans continue to "provide value" in chess playing when the time between moves provides enough time for reasoning. In this context, human players use the computer as a tool to consider their options carefully. Thus, for Murray, he felt that there would be an important distinction between "real time" decisions (where humans may not be able to provide value to computers) to longer-term decisions where human reasoning could provide significant value.

He also pointed out that humans who trusted the chess-computer's recommendations too much (who "turned off their critical thinking skills") ended up missing opportunities for better chess moves. Murray also pointed out the critical importance of the computer being able to explain its recommendations, in terms of the human players needing to be able to "deal with" the situations that result from following the computer's recommendations.

In closing, he observed that the most "superhuman" aspect of chess computers is their ability to "escape" when placed in positions that a human would consider to be inescapable.

Ping Wang [O2B_Wang] discussed implementation challenges for ICAs from the viewpoint of entrepreneurs. He pointed out that Venture Capital-funded startups represent an important source of innovation beyond academia and large corporate research efforts. One challenge that Ping saw was the need to increase the number of researchers that possess the skills and perspectives to participate in entrepreneurial innovation, and to navigate the "IP walls" between academic and industrial research groups. Ping proposed that **Entrepreneurial Accelerators** could contribute to innovation in ICAs, and that researchers in ICAs (even if squarely in academia or industrial research) should maintain a strong focus on mentorship, in order to produce other talented researchers, whether for academia, industrial research or the startup world. He encouraged both academic and industrial researchers to not fear startups, but to take the opportunity to guide and balance the commercial incentives inherent in startups. He felt that university incubators needed to be retuned away from revenue generation for the universities back to an interdisciplinary and supportive role, and stressed the role of open innovation in a healthy innovation ecosystem involving academia, large industry and startups.

In the first comment, Eric Horvitz asked Murray whether the observations that humans help augment "correspondence chess," where the time between turns is long, might depend on the way in which the chess computers were optimized for faster turns, and whether this gap might disappear if chess computers were to be optimized for longer turns. Murray observed that in situations when computers have had plenty of time to prepare, for instance in the overnight computation of opening moves, they have not provided much advantage, so that existing algorithmic approaches would be unlikely to close this gap (where human reasoning provides advantage for non-real-time decisions).

Blaise observed that innovations in chess computers that made them capable of explaining their reasoning would presumably also improve their ability to reason ("to detect fortresses") as well. Murray observed that while there is a style of chess described as "computer chess" – a move that no human would think of, he pointed out that younger human players who have learned to play chess by working with chess-computers are in fact more capable of coming up with these kinds of moves.

Gary Marcus asked whether Murray's plot of Elo-rating has a theoretical ceiling ("the best chess possible") or not, to which Murray responded that human+computer performance has continued to improve, so it is hard to be sure.

Gary then asked Ping why universities should not fear startups as a drain of talent away from academia, to which Ping responded that this effect would be temporary, followed by increased supply of talent into academia driven by interest in the field, providing an increased supply of talented researchers.

As a final comment, Eric Horvitz asked whether there are additional opportunities for embedding special capabilities into the human+computer loop for computer-chess aimed at supporting and extending human abilities in chess versus as a separate black box for playing chess. Murray responded that there were – he described a series of "freestyle" chess tournaments a few years ago, that were won not by the best computer program, nor by the best humans, but by a team of average human players who had worked out the optimal way to interact with a fairly modest set of chess-computers. He extended this observation to ICAs: that refining the best way of computer-human interaction is a very powerful way to optimize overall performance.

This closed the discussions for Day One.

Workshop website (agenda, recordings, presentations, speaker bios, etc.):
https://www.src.org/calendar/e006057/

References for APPENDIX C:

[K1_Horvitz] Eric Horvitz, "People, Computing, Cognition: Toward Effective, Valued Personal Assistants" (Day 1, 12:40) **https://www.src.org/calendar/e006057/eric-horvitz.pdf**

[K2_Breazeal] Cynthia Breazeal, "Living and Working with Intelligent Cognitive Assistants: Societal Challenges and Opportunities for Impact" (Day 1, 1:17:45) **https://www.src.org/calendar/e006057/cynthia-breazeal.pdf**

[O1A_Perlmutter] Saul Perlmutter, "Intelligent-Cognitive-Assisted Deliberation Tools" (Day 1, 2:03:30) **https://www.src.org/calendar/e006057/saul-perlmutter.pdf**

[O1A_Forbus] Kenneth Forbus, "Vision: AI Assistants for Learning Science" (Day 1, 2:12:35) **https://www.src.org/calendar/e006057/kenneth-forbus.pdf**

[O1A_Olshausen] Bruno Olshausen, <<Towards the human visual system>> (Day 1, 2:20:50) **https://www.src.org/calendar/e006057/bruno-olshausen.pdf**

[O1A_Vertesi] Janet Vertesi, "Humans do not work alone" (Day 1, 2:26:10) **https://www.src.org/calendar/e006057/janet-vertesi.pdf**

[O1A_Laird] John Laird, "Neural Intelligent Cognitive Orthotic" (Day 1, 2:37:55) **https://www.src.org/calendar/e006057/john-laird.pdf**

[O1B_Kaplan] Ron Kaplan, <<Observations on ICAs>> (Day 1, 3:04:15) **https://www.src.org/calendar/e006057/ron-kaplan.pdf**

[O1B_Gureckis] Todd Gureckis, "Computational curiosity" (Day 1, 3:14:18) **https://www.src.org/calendar/e006057/todd-gureckis.pdf**

[O1B_Pavel] Misha Pavel, "Model-Based Intelligent Cognitive Assistants" (Day 1, 3:28:45) **https://www.src.org/calendar/e006057/misha-pavel.pdf**

[O1B_Rabaey] Jan Rabaey, "Cognitive Personal Assistants – Anno 2040" (Day 1, 3:40:25) **https://www.src.org/calendar/e006057/jan-rabaey.pdf**

[O1B_Saraswat] Vijay Saraswat, "Upton: Achieving Professional Facility in Deep Domains Realizing Unified Theories of Cognition" (Day 1, 3:48:30) **https://www.src.org/calendar/e006057/vijay-saraswat.pdf**

[O2A_Marcus] Gary Marcus, "Four questions for AI and cognitive science" (Day 1, 4:08:45) **https://www.src.org/calendar/e006057/gary-marcus.pdf**

[O2A_ Horn] Bruce Horn, "NSF Interactive Cognitive Agents: Implementation Challenges & Tradeoffs" (Day 1, 4:16:55) **https://www.src.org/calendar/e006057/bruce-horn.pdf**

[O2A_ Kamar] Ece Kamar, "People and AI: Together not Alone" (Day 1, 4:21:52) **https://www.src.org/calendar/e006057/ece-kamar.pdf**

[O2A_ Spelke] Liz Spelke, "Core Cognition" (Day 1, 4:30:52) **https://www.src.org/calendar/e006057/liz-spelke.pdf**

[O2A_ Scassellati] Brian Scassellati, "Providing Cognitive and Social Support with Intelligent Robots" (Day 1, 4:46:00) **https://www.src.org/calendar/e006057/brian-scassellati.pdf**

[K3_AguerayArcas] Blaise Aguera y Arcas, "deep learning, big data, and personal devices" (Day 1, 5:05:05) **https://www.src.org/calendar/e006057/blaise-aguera-y-arcas.pdf**

[O2B_Martinez] Alberto Martinez, "The era of intelligent ambient computing" (Day 1, 6:23:35) **https://www.src.org/calendar/e006057/alberto-martinez.pdf**

[O2B_Campbell] Murray Campbell, "Collaboration with Intelligent Cognitive Agents: Some Lessons from Computer Chess" (Day 1, 6:36:47) **https://www.src.org/calendar/e006057/murray-campbell.pdf**

[O2B_Wang] Ping Wang, "Implementation Challenges & Tradeoffs via Entrepreneurship" (Day 1, 6:49:00) **https://www.src.org/calendar/e006057/ping-wang.pdf**

[FirstDaySummary] "key concepts questions_day 1" (was prepared, but was not shown during the workshop)

[BreakoutSummary1] "NSF Workshop Break Out Session J-3"

[BreakoutSummary2] "NSF-slides"

[BreakoutSummary2A] "Team 2 summary notes_JonC"

[BreakoutSummary3] "AuditoriumNotes_V2"

APPENDIX D

Recommended Reading/Viewing

Final slides from DARPA ISAT Study on Foundations of Augmented Cognition: **http://research.microsoft.com/~horvitz/augmented_cognition_August_2001_study_briefing.ppt**

Workshop on "Foundations of Augmented Cognition, DARPA ISAT, National Academy of Sciences, August 2001. **http://research.microsoft.com/en-us/um/people/horvitz/acog.htm**

E. Horvitz, C. M. Kadie, T. Paek, D. Hovel. 'Models of Attention in Computing and Communications: From Principles to Applications', *Communications of the ACM* 46(3):52-59, March 2003.

E. Horvitz. 'Reflections on Challenges and Promises of Mixed-Initiative Interaction', *AAAI Magazine* 28, Special Issue on Mixed-Initiative Assistants (2007).

E. Horvitz. 'Principles of Mixed-Initiative User Interfaces', Proceedings of CHI '99, ACM SIGCHI Conference on Human Factors in Computing Systems, Pittsburgh, PA, May 1999.

E. Horvitz, A. Jacobs, D. Hovel. 'Attention-Sensitive Alerting', Proceedings of UAI '99, Conference on Uncertainty and Artificial Intelligence, July 1999, Morgan Kaufmann Publishers: San Francisco. pp. 305-313.

E. Kamar, S. Hacker, E. Horvitz. 'Combining Human and Machine Intelligence in Large-scale Crowdsourcing', *AAMAS 2012*, Valencia, Spain, June 2012.

D. Bohus and E. Horvitz. 'Decisions about Turns in Multiparty Conversation: From Perception to Action', *ICMI 2011: International Conference on Multimodal Interaction*, Alicante, Spain, November, 2011.

E. Kamar and E. Horvitz, 'Jogger: Investigation of Principles of Context-Sensitive Reminding', *Proceedings of International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tapei, May 2011.

E. Horvitz, J. Apacible, and P. Koch. 'BusyBody: Creating and Fielding Personalized Models of the Cost of Interruption', *Proceedings of CSCW, Conference on Computer Supported Cooperative Work*, ACM Press, November 2004.

E. Horvitz and J. Apacible. 'Learning and Reasoning about Interruption'. *Proceedings of the Fifth ACM International Conference on Multimodal Interfaces*, November 2003, Vancouver, BC, Canada.

E. Horvitz, S. Dumais, P. Koch. 'Learning Predictive Models of Memory Landmarks', *Cognitive Science 2004: 26th Annual Meeting of the Cognitive Science Society*, Chicago, August 2004.

S. T. Iqbal and E. Horvitz. 'Disruption and Recovery of Computing Tasks: Field Study, Analysis, and Directions', *Proceedings of CHI 2007*, San Jose, California, April 2007.

E. Horvitz, J. Apacible, R. Sarin, and L. Liao. 'Prediction, Expectation, and Surprise: Methods, Designs, and Study of a Deployed Traffic Forecasting Service', *Proceedings of the Conference on Uncertainty and Artificial Intelligence 2005*,, AUAI Press, July 2005.

M. Ringel, E. Cutrell, S. Dumais, E. Horvitz. 'Milestones in Time: The Value of Landmarks in Retrieving Information from Personal Stores'. *Proceedings of Interact 2003: Ninth International Conference on Human-Computer Interaction*, September 2003, Zürich, Switzerland.

Czerwinski, M. and Horvitz, E. 'An Investigation of Memory for Daily Computing Events'. *Proceedings of HCI 2002: Sixteenth British HCI Group Annual Conference*, London, England, September 2002.

M. Czerwinski, M., E. Horvitz, and S. Wilhite. 'A Diary Study of Task Switching and Interruptions', *Proceedings of CHI 2004, Human Factors in Computing Systems*. Vienna, April 2004.

D. Graus, P.N. Bennett, R.W. White, E. Horvitz. 'Analyzing and Predicting Task Reminders', 24th Annual Conference on User Modeling, Adaptation and Personalization (UMAP 2016), July 2016.

A. Kapoor, B. Lee, D. Tan, E. Horvitz. 'Performance and Preferences: Interactive Refinement of Machine Learning Procedures', *AAAI 2012*, Toronto, July 2012

A. Kapoor, B. Lee, D. Tan, E. Horvitz. 'Interactive Optimization for Steering Classification', *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, Atlanta, GA, April 2010.

S.T. Iqbal, Y.C. Ju, E. Horvitz. 'Cars, Calls, and Cognition: Investigating Driving and Divided Attention', *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, Atlanta, GA, April 2010.

S.T. Iqbal and E. Horvitz. 'Notifications and Awareness: A Field Study of Alert Usage and Preferences', *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW)*, Savannah, GA, February 2010.

A. Kapoor, D. Tan, P. Shenoy, E. Horvitz. 'Complementary Computing for Visual Tasks: Meshing Computer Vision with Human Visual Processing', *IEEE International Conference on Automatic Face and Gesture Recognition*.

E. Horvitz, J. Apacible, M. Subramani. 'Balancing Awareness and Interruption: Investigation of Notification Deferral Policies' *Proceedings of the Tenth Conference on User Modeling* (UM 2005), July 2005, Edinburgh, Scotland.

E. Horvitz, P. Koch, C.M. Kadie, and A. Jacobs. 'Coordinate: Probabilistic Forecasting of Presence and Availability' *Proceedings of UAI '02, Proceedings of the Eighteenth Conference on Uncertainty and Artificial Intelligence*, Edmonton, Canada, July 2002, pp. 224-233.

'Art in the Age of Machine Intelligence' (**https://medium.com/artists-and-machine-intelligence/what-is-ami-ccd936394a83#.d4ucypppk**)

'Federated Learning of Deep Networks Using Model Averaging' (**http://arxiv.org/abs/1602.05629**)

'What Difference Does a Robot Make? The Material Enactment of Distributed Coordination' (**http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2438709**)

'Artificial intelligence in business: The state of play and future prospects' (**http://www.zdnet.com/article/artificial-intelligence-in-business-the-state-of-play-and-future-prospects/**)

'Google CEO Pichai says devices will fade away - but launches new hardware division' (**http://www.zdnet.com/article/google-ceo-pichai-says-devices-will-fade-away-but-launches-new-hardware-division/?tag=nl.e539&s_cid=e539&ttag=e539&ftag=TRE17cfd61**)

'A million spiking-neuron integrated circuit with a scalable communication network and interface' (**http://science.sciencemag.org/content/345/6197/668.full?cm_mc_uid=46042019264314537463586&cm_mc_sid_50200000=1461295013**)

'Convolutional Networks for Fast, Energy-Efficient Neuromorphic Computing' (**http://arxiv.org/abs/1603.08270**)

'Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors'; Kanerva, P. (2009); Cognitive Computation, 1(2), 139-159. (**http://redwood.berkeley.edu/vs265/kanerva09-hyperdimensional.pdf**)

'Scene analysis in the natural environment'; Lewicki MS, Olshausen BA, Surlykke A, Moss CF (2014); Frontiers in Psychology, 5, article 199. (**http://redwood.berkeley.edu/bruno/papers/scene-analysis.pdf**)

'A Neurobiological Model of Visual Attention and Invariant Pattern Recognition Based on Dynamic Routing of Information'; Olshausen BA, Anderson CH, Van Essen DC (1993); The Journal of Neuroscience, 13(11), 4700-4719. (**http://redwood.berkeley.edu/bruno/papers/jneurosci93.pdf**)

'20 years of learning about vision: Questions answered, Questions unanswered, and Questions not yet asked'; Olshausen BA (2012); in: '20 Years of Computational Neuroscience'; J. Bower, Ed. Springer. (**http://redwood.berkeley.edu/bruno/papers/CNS2010-chapter.pdf**)

'The Human Intranet – Where Swarms and Humans Meet'; J. Rabaey; IEEE Pervasive Computing, Vol 14 Issue 1, January 2015.

- **Books**:

'The Second Machine Age'; E. Brynjolfsson, A. McAfee

'The Master Algorithm'; Pedro Domingos

'Goedel, Escher, Bach'; Doug Hofstadter

'Surfaces and Essences'; Doug Hofstadter

'In Search of Memory'; Eric Kandel

'You are Not a Gadget'; Jaron Lanier

'Who Owns the Future'; Jaron Lanier

'Brains: How They Seem to Work'; Dale Purves

'Incognito: the Secret Lives of the Brain'; David Eagleman

'Pandora's Star'; Peter Hamilton

'The Last Firewall'; William Hertling

'Seeing like a Rover: How robots, teams and images craft knowledge of Mars'; Janet Vertesi

'Superintelligence'; Nick Bostrom

'How to Create a Mind'; Ray Kurzweil

'The Turing Option'; Marvin Minsky and Harry Harrison

'2001, A Space Odyssey'; Arthur C. Clarke

'I, Robot'; Isaac Asimov

'Do Androids Dream of Electric Sheep?'; Philip Dick

- **Movies**:

'Robot and Frank'

'Her'

'2001, A Space Odyssey'

'Ex Machina'

- **TED/YouTube Videos:**

https://www.ted.com/talks/cynthia_breazeal_the_rise_of_personal_robots

https://www.youtube.com/watch?v=dpoVh9xwdD4 (E. Horvitz, TED presentation on People and Cognition: February 2013)

http://tedxtalks.ted.com/video/Making-Friends-With-Artificia-2 (Eric Horwitz)

https://www.youtube.com/watch?v=PtHicmmblMk ("Can AI Become More Human?"; Gary Marcus, et al)

https://www.ted.com/talks/rodney_brooks_why_we_will_rely_on_robots

https://www.ted.com/talks/blaise_aguera_y_arcas_demos_photosynth

https://www.ted.com/talks/blaise_aguera_y_arcas_how_computers_are_learning_to_be_creative

https://www.ted.com/talks/anthony_goldbloom_the_jobs_we_ll_lose_to_machines_and_the_ones_we_won_t

https://www.ted.com/talks/chieko_asakawa_how_new_technology_helps_blind_people_explore_the_world

https://www.ted.com/talks/fei_fei_li_how_we_re_teaching_computers_to_understand_pictures

https://www.ted.com/talks/jeremy_howard_the_wonderful_and_terrifying_implications_of_computers_that_can_learn

https://www.ted.com/talks/guy_hoffman_robots_with_soul

https://www.ted.com/talks/shyam_sankar_the_rise_of_human_computer_cooperation

https://www.ted.com/talks/hod_lipson_builds_self_aware_robots

APPENDIX E

ACKNOWLEDGEMENTS

ICA Workshop Organizing Committee & Report Authors:

Cynthia Breazeal

Geoff Burr

Eric Horvitz

Gary Marcus

Terry Sejnowski

Jon Candelaria

Janet Vertesi (report co-author)

Kenneth Forbus (report edits)

Oversight Committee:

Mike Roco

Taffy Kingscott

Celia Merzbacher

Dustin Todd

APPENDIX F

WORKSHOP AGENDA

Thursday, May 12, 2016		
7:30 - 8:30 am	Registration & Continental Breakfast	
8:30 - 9:00 am	Welcome, Workshop Goals, Day 1 Agenda	Jon Candelaria (SRC) Jeff Welser (IBM) Mihail Roco (NSF)
9:00 - 9:45 am	**Keynote Talk I** "**People, Computing, and Cognition: Toward Effective and Valued Personal Assistants**"	Eric Horvitz (Microsoft)
9:45 - 10:05 am	Break	
10:05 - 10:45 am	**Keynote Talk II** "**Living and Working with Intelligent Cognitive Assistants: Societal Challenges and Opportunities for Impact**"	Cynthia Breazeal (MIT)
10:45 - 11:30 am	**Open Panel Session IA** "Moonshot Application Concepts"	Saul Perlmutter (UC/Berkeley) Kenneth Forbus (Northwestern) Bruno Olshausen (UC/Berkeley) Janet Vertesi (Princeton) John Laird (Univ. of Michigan)
11:30 - 12:45 pm	Lunch	
12:45 - 1:30 pm	**Open Panel Session IB** "Moonshot Application Concepts"	Ron Kaplan (Amazon) Todd Gureckis (NYU) Misha Pavel (Northeastern) Jan Rabaey (UC/Berkeley)
1:30 - 1:45 pm	Break	
1:45 - 2:30 pm	**Open Panel Session IIA** "Implementation Challenges & Tradeoffs"	Gary Marcus (NYU) Bruce Horn (Intel) Ece Kamar (Microsoft) Liz Spelke (Harvard) Brian Scassellati (Yale)
2:30 - 2:45 pm	Break	
2:45 - 3:30 pm	**Keynote Talk III** "**Privacy Challenges for Big Data**"	Blaise Aguera y Arcas (Google)
3:30 - 4:15 pm	**Open Panel Session IIB** "Implementation Challenges & Tradeoffs"	Blaise Aguera y Arcas (Google) Alberto Martinez (Intel) Murray Campbell (IBM) Ping Wang (Ansir) Pamela Hinds (Stanford)
4:15 - 4:30 pm	Break	
4:30 - 5:15 pm	Day 1 Panel Summaries	

“Intelligent Cognitive Assistants”

Friday, May 13, 2016		
7:30 - 8:30 am	Continental Breakfast	
8:30 - 8:45 am	Day 1 Summary, Day 2 Goals	Geoffrey Burr (IBM)
8:45 - 10:00 am	**(3) Parallel Breakouts:** “Research Priorities & Roadmap Goals/Milestones”	
10:00 - 10:20 am	Break	
10:20 - 11:30 am	**(3) Parallel Breakouts:** “Research Priorities & Roadmap Goals/Milestones”	
11:30 - 12:30 pm	Lunch	
12:30 - 1:45 pm	Day 2 Breakout Summaries	
1:45 - 2:00 pm	Wrap-Up	Jon Candelaria (SRC)
2:00 pm	Adjourn	

WORKSHOP WEBSITE (slides, participant list & bios, videos): https://www.src.org/calendar/e006057/