



## Wanted: Collaborative intelligence



Susan L. Epstein <sup>a,b,\*</sup>

<sup>a</sup> Department of Computer Science, Hunter College, New York, NY 10065, USA

<sup>b</sup> The Graduate Center of The City University of New York, 695 Park Avenue, New York, NY 10016, USA

### ARTICLE INFO

#### Article history:

Received 5 May 2014

Received in revised form 20 November 2014

Accepted 17 December 2014

Available online 6 January 2015

#### Keywords:

Collaboration

Intelligence

Perception

Dialogue

### ABSTRACT

Although the original vision for artificial intelligence was the simulation of (implicitly human) intelligence, research has gradually shifted to autonomous systems that compete with people. The resultant popular attitude toward artificial intelligence, we argue here, is by turns disdain, grudging acceptance, and fear. That attitude not only limits our work's potential, but also imperils its support. This paper proposes a constructive alternative: the development of collaborative intelligence. As envisioned here, a collaborative intelligence does not require encyclopedic command and need not be limited to a single problem. The necessary components of a collaborative intelligence are nearly at hand, and the key issues readily identified. As a first step, this paper proposes three challenging but accessible problems that would both change the public perception of artificial intelligence and spur substantive research to advance our science.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

This is a position paper, intended both to challenge and to inspire research in artificial intelligence (AI). It considers the path AI research has taken, and seeks to broaden its long-term goals. The vision offered here integrates what AI has learned in a way that would advance our science and support people.

This section restates the core ideas that began AI, describes how we as researchers now see our field, considers the current media portrayal of AI, and notes how our research targets shift. In response, Section 2 proposes collaboration between a person and a machine, and highlights some crucial differences between human and computer collaborators. An extended example of collaborative intelligence in Section 3 identifies key issues inherent in the development of a collaborative intelligence. Section 4 presents three reasonably approachable, exciting problems that would both engage AI researchers and benefit human users.

### 1.1. How AI was first envisioned

In August 1955, in what has come to be known as the Dartmouth manifesto, Claude Shannon, Marvin Minsky, Nathaniel Rochester, and John McCarthy proposed AI as the theme for a conference to be held the following year.<sup>1</sup> Their targets for this new field included precursors of what AI would now term problem solving, natural language processing, artificial neural networks, complexity theory, machine learning, and perception. Their premise was that “every aspect of learning or

\* Corresponding author at: Department of Computer Science, Hunter College, 695 Park Avenue, New York, NY 10065, USA. Tel.: +1 212 7725213.

E-mail address: susan.epstein@hunter.cuny.edu.

<sup>1</sup> <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>, May 1, 2014.

any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.” There is no indication in that document that computers were to work together with people; the intent was to develop an autonomous intelligence.

At the time, Alan Newell and Herbert Simon had already begun work on the Logic Theorist, a program to prove mathematical theorems. Although it could prove 38 of the first 52 theorems in *Principia Mathematica* (a key contribution to the foundations of mathematics [31]), the Logic Theorist was of little interest at the Dartmouth conference [7]. Few scholars then considered computer-constructed proofs of elementary theorems a dramatic achievement. Nonetheless, Newell and Simon, along with J.C. Shaw, continued their work. They were quite clear about the Logic Theorist’s need for knowledge, both about the problem domain and the mechanisms necessary to guide its search. This led to the General Problem Solver (GPS) [20]. GPS tackled problem solving from two perspectives: machine-achievable processes and behavior observed in humans. True to the Dartmouth vision of simulation, college sophomores were asked to think aloud as they solved symbolic logic problems, and GPS was developed to simulate what Newell and Simon observed.

In many ways, in 1958 GPS set the agenda for AI research. It established some of AI’s basic building blocks: goals, objects characterized by feature values, and operators as functions on objects. GPS explicitly recognized the importance of feature identification, problem-dependent heuristics, efficient data structures and algorithms, performance evaluation, and the potential for exponential search spaces. In response, it advocated problem-independent heuristics, including means-ends analysis and planning; a model of problem solving as a sequence of actions; and the careful separation of process from data (i.e., isolation of *what* was being thought about from *how* it was being thought about). We consider next how AI has addressed that agenda.

### 1.2. AI’s trajectory

The path AI research has taken can be readily traced from the first AI conference at Dartmouth through two of the major meetings that followed it. The Dartmouth manifesto postulated general goals: to discover “how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.” There was no mention of interaction with or impact on people, and no targets beyond the things that people can do.

Fourteen years later, the papers presented at the first IJCAI in 1969 reflect the topics AI researchers chose to pursue. The Wordle™ in Fig. 1(a) for the titles of those papers indicates a strong focus on implementation (e.g., the tokens “system,” “program,” “memori,” “method,” “process,” “language,” “robot,” “structure,” and “control”) and theory (e.g., “model,” “theorem,” “base”). There is also a significant interest in “pattern,” a precursor to learning.

Contemporary AI is described by a second Wordle, in Fig. 1(b), this time for the titles of papers presented at the 23rd IJCAI, in 2013. “Learn” and “model” now clearly dominate, with techniques (e.g., “plan,” “gener,” “search,” “constraint,” “logic,” and “network”) as important concerns, along with “game,” “social,” and “data.” This is how we, as researchers, now describe our field to one another. Several subfields, most notably vision and robotics, have by 2013 detached themselves, but a focus on language and structure persists. Although “human” appears in both Wordles, it is not prominent in either of them.

The dramatic changes between 1969 (in Fig. 1(a)) and 2013 (Fig. 1(b)) were driven, I believe, by our collective fascination with hard problems. To solve those problems, AI researchers developed a diverse set of representations to model the real world for computers. These include logics, ontologies, semantic nets, and rich graph structures. Then, to harness these representations, AI researchers built inference mechanisms and search algorithms intended to manipulate that knowledge. Empirical work rigorously validates the efficiency and effectiveness of these products.

The combined push to compare and compete has honed performance and driven research. The need for uniform scientific evaluation of empirical work has led to ambitious, shared datasets, including repositories for machine learning, planning, and image labeling. Competition, along with common and exacting evaluation metrics, allows us to see which methods perform best on which data. The earliest instance of this appears to have been the push to prove all the theorems in *Principia Mathematica*, a feat ultimately accomplished independently by both Gilmore and Wang [9,30]. Since then, prize money and targeted research funding have engendered enthusiastic competitions in such areas as speech recognition, game playing, and self-driving cars, where work might have otherwise developed far more slowly.

As a result, AI’s standard for success has become the ability of one system, algorithm, architecture, representation, or approach to outperform another. Clearly, we are in search of the best machine intelligence we can construct, without any regard to what people can do. Meanwhile, this clever problem solving has had some unanticipated results.

### 1.3. AI in the mainstream media

AI’s description in the popular press has done little to enhance its reputation. The media has alternately overpraised AI for techniques that are not new, and over-criticized it for overly optimistic promises.<sup>2</sup> As LeCun notes, “AI ‘died’ about four

<sup>2</sup> Compare, for example, [http://www.dailygalaxy.com/my\\_weblog/2010/01/artificial-intelligence-will-leapfrog-humans-by-2020-says-scifi-great-a-weekend-feature.html](http://www.dailygalaxy.com/my_weblog/2010/01/artificial-intelligence-will-leapfrog-humans-by-2020-says-scifi-great-a-weekend-feature.html) with [http://www.skeptic.com/reading\\_room/artificial-intelligence-gone-awry/](http://www.skeptic.com/reading_room/artificial-intelligence-gone-awry/), January 7, 2015.



“reports” abound, including “Soldiers, Machines, and the Rise of Battlefield Robotics” [5] and “The rise of robot journalists and the world brain: beware of the media industrial complex” [17]. Even highly respected sources trumpet the dangers in their headlines: “Don’t be evil, genius” [2] and “Prof ... predicts next 25 years will bring instant language translation – and a rise in ‘lifelogging’ and cyberwar too” [3].

As a discipline, we have a public opinion problem. For example, a recent British survey provides evidence that AI has begun to frighten and worry the general populace.<sup>4</sup> Among respondents, 39% worried that “there will be a time when robots present a danger to humans,” and “29% believed that machines will pose a danger to the human race.” Particular fears cited include the use of drones and the undermining of “traditional ways of life.” Meanwhile, a recent United States survey reported that people were both optimistic about the quality of their lives in a robot-rich future and concerned about their ability to find employment there [18].

In summary, although AI’s founders did not envision world domination by a supreme cybernetic brain, nor seek to sell small electronic devices, there is now a general perception that AI is, by turns menacing, device-centric, flakey, or a finished parlor trick. It is necessary to respond to this unflattering and inaccurate portrayal, and AI researchers have begun to do so. The 1st International AAAI Workshop on AI and Ethics will be held at AAAI-15. Its call for papers acknowledged both the media torrent and the possibility that it is correct. The workshop’s call solicited papers that address “the ethical questions implicit in such headlines which go to the centre of the quest to build AI systems with potentially super-human intelligence.”<sup>5</sup>

#### 1.4. Shifting targets for AI research

AI research has focused heavily on the targets posed in 1955, and continues to refine those targets when the resultant artifacts do not quite reach the anticipated skill level. We highlight several noteworthy areas here: natural language processing (NLP), the data-driven paradigm, commonsense reasoning, and robotics. Each addresses a shifting target.

The 1955 target “how to make machines use language” was recast by Turing as a program that, when engaged in conversation, could not be distinguished from a person [28]. Recent proclamations that we had reached a “milestone in computing history” because a chatterbot won the Loebner prize [23] have further perpetuated the notion that a computer able to fool people is artificially intelligent. In a 5-minute conversation, a program convinced 30% of its 30 human judges that it was Eugene, a 13-year old Ukrainian boy. The judges excused its poor syntax because English was not its native language, and attributed its cavalier replies to its age. What this program primarily proved, I would argue, is that fooling people is not a sufficient claim to intelligence [21]. NLP research is now considerably more ambitious; it includes, for example, understanding, summarization, and translation.

AI researchers soon recognized that to solve the “kinds of problems now reserved for humans” would require a substantive knowledge base and a host of ways to use it [16]. As engineers produced computers that could store vast quantities of data and process it quickly, the data-driven paradigm arose: what intelligence requires is enough facts. For some purposes, that approach is surprisingly successful (e.g., web search, self-driving cars). The development of such a knowledge base has motivated massive AI efforts from CYC [22] to Watson [26], as well as recent funding initiatives. Too often, however, statistical correlation fails to connect information in meaningful ways. Artifacts like the Jeopardy-playing Watson remind us what remains beyond the reach of a program with a great deal of knowledge but not enough incisive procedures to apply it. For example, Watson carefully followed its statistical algorithm to proffer Toronto as a city in the United States.

Commonsense reasoning provides meaning and context to data. Although this target for AI is worthwhile, it is also considerably more difficult. (Indeed, financial gurus were recently reassured that robots will never take everyone’s jobs because so much of our daily work requires commonsense [4].) At the very least, a commonsense reasoner should correctly answer simple questions that require knowledge about properties of ordinary objects. For example, neither Watson nor a web search would be able to answer the question “Can a crocodile run a steeplechase?” [14]. That gap drives the development of COPA, a question bank that will force intelligence-claiming artifacts to answer a broad range of questions about plausible alternatives [24]. Associative techniques based on text corpora do little better than random on the COPA questions [24].

Although robots were not mentioned in the Dartmouth manifesto, as computing machines they quickly attracted researchers’ attention. Although “robot” is relatively prominent in the 1969 Wordle, it did not appear in the title of any paper at AAAI-2013. Today’s commercially deployed robots are engineered primarily for environments where their programs can follow rules, be tele-operated, or rely on fixed structure (e.g., self-driving cars with maps, warehouses with rigid layouts and restricted areas in which to safeguard humans) [1]. Nonetheless, many of the tasks we would like to allocate to robots (e.g., household chores, elder care) would place them in complex, cluttered environments that include a broad variety of objects, people, and perhaps other robots. This requires commonsense knowledge and reasoning, as well as the ability to contend with uncertainty produced by noise in the robots’ sensors and actuators. The focus on robotics at AAAI-15 seeks to reintroduce AI and robotics to address such tasks.

In summary, people have skills that AI research has not yet reproduced in machines. Moreover, there are many tasks we would like to allocate to, or at least share with, computers. This is the impetus for collaborative intelligence.

<sup>4</sup> <http://corporate.uktv.co.uk/news/article/humans-fear-rise-machines/>, July 14, 2014.

<sup>5</sup> <http://www.aaai.org/Workshops/ws15workshops.php#ws01>, October 17, 2014.

## 2. Collaborative intelligence

Historically, people have built machines because they wanted help. If people are to trust an artificial intelligence, we will want to know not only what is true but also why, that is, what is the source of the machine's premises and the pattern of its reasoning. Facts should have provenance (e.g., as in Start [12]) and rationales should be transparent to human users.

As envisioned here, a collaborative intelligence (CI) partners with a person to achieve the person's goals. The assumption is that some subtasks are more reasonably delegated to the person, and others to the computer. A CI is intended not to substitute for a human employee, but to engage in a task with one. For example, on request, a CoBot (a mobile robot) will deliver an item to another floor in its building, but it asks a person to call the elevator and press the button for the correct floor [6]. Moreover, if no person is available, the CoBot signals that it needs help. Inherent here is that people and computers have complementary skills. A robot vacuum cleaner, for example, may collect small dirt particles from the floor more efficiently than a person armed with cleaning tools, but the person knows under what circumstances that robot should be deployed (e.g., not in water or large debris). A robot vacuum cleaner that identified situations where its human user could be of help (e.g., to empty its collection bag or remove a playful pet) and then requested that help would be the basis of an interesting CI. Although interfaces developed in work on human–computer interaction are necessary, a CI is by definition more concerned with an appropriate and supportive division of labor.

To collaborate effectively with a person, a CI must be able to model the human view of the world. This is an area that cognitive science addresses, but one rarely considered in AI. (Notable exceptions here are the work in commonsense reasoning and qualitative reasoning.) For example, in a classic psychology experiment, people regularly reported that Cuba was similar to Russia, but Russia was not similar to Cuba [29]. Nonetheless, most AI research treats object similarity as a symmetric relation. At the very least, a CI should be aware of such human perceptions, particularly when they differ from those of the machine.

A CI must engage in dialogue with its human partner. As an ongoing example, consider a kitchen CI that is asked, "How do you make soup?" A CI with extensive knowledge must filter its possible responses, because its help will only be valued if it is neither too simplistic (e.g., "Just open a can and heat the contents.") nor too detailed (e.g., "...the history of soup is probably as old as the history of cooking..."<sup>6</sup>). Appropriate answers are attuned to the user's demographics and current circumstances. A suggested recipe, for example, would more likely be accepted if it considered the person's time frame, budget, taste profile, and environment. A recipe for a cold soup, for example, is likely to be better received in the summer. A CI should also be able to explain its answers to its human collaborator. For example, if the user questions the incorporation of somewhat limp vegetables, our kitchen CI could explain that they add flavor and can later be puréed to thicken the soup before serving. A CI that provides periodic procedural advice or additional information as a task progresses is particularly useful in response to change. For example, midway in cooking, an ingredient may be found to be unusable or exhausted, so that the recipe requires adjustment.

Productive dialogue requires not only a shared vocabulary but also a shared context in which to construct good answers. A CI should elicit that context by asking questions. Many human endeavors have multiple goal criteria and require multiple decisions, each of which impacts several goals. A CI that elicits its user's goals can ask how they should be prioritized, so that it can model both the user's knowledge and the user's expectations. Our kitchen CI, for example, could ask whether the soup ingredients are to be purchased or limited to those on hand. Some important questions address ambiguity or motivation. It is far easier to disambiguate with a question, such as "Did you want me to prepare a soup or would you just like a recipe?" Questions about the authority and significance of knowledge sources also help a CI filter information to reduce it to a manageable quantity: "Do you prefer the approach of a particular chef?" Questions can help a CI contend with misunderstandings and unanticipated occurrences. Questions also relieve the CI of an obligation to understand perfectly.

Dialogue also permits the CI to signal its internal state to the user in several ways. Questions indicate that the CI has too many or too few options, and therefore needs guidance. A good collaborator monitors its own declarative knowledge ("I am unfamiliar with 'thickening'"), procedural skill ("I've rarely made stock before"), and progress on the task ("This appears to be taking longer than I expected"). Such self-awareness permits the CI to evaluate its current strategy and even to change it: "The soup is still not very thick. Should we add some cream or some cooked rice?" This kind of transparency also gives the CI the opportunity to learn from its users. Nonetheless, the construction of a CI must contend with the fundamental differences in the way that any machine and any person function.

A CI must recognize that its human collaborator is very different from a machine, so that the CI can interpret what it experiences and how it should communicate. Assuredly, people perform some calculations more slowly and remember facts less quickly and accurately. Nonetheless, people are embodied in a complex and dynamic environment from which they continually receive vast quantities of perceptual data. People are engineered to exploit multiple sensory modalities (hearing, vision, smell, touch, proprioception), and they seamlessly integrate that input with prior knowledge. In contrast, a computer must rely on artificial sensory devices (e.g., cameras, lidar, infrared, sonar, and wireless) whose percepts are fundamentally different and require deliberate integration with the computer's knowledge store.

People use percepts from a single modality in more than one way at the same time. For example, when a person looks at a visual stimulus, the rods and cones in her eyes do not send a pixelated snapshot of what she sees to her brain. Rather,

<sup>6</sup> A search on "soup" found this quickly at <http://www.foodtimeline.org/foodsoups.html>, May 1, 2014.

- H: Hey, Joe. (1)  
 CI: Good morning. How can I help? (2)  
 H: When can I have lunch with Alice next week? (3)  
 CI: Tuesday or Friday would work. (4)  
 H: Just a minute. (*Pause while H speaks to Alice.*) Okay, Tuesday at 1. (5)  
 CI: That would give you an hour. You have a meeting uptown at 2. (6)  
 H: Right. (*Another pause while H speaks to Alice.*) 12:30 then. (7)  
 CI: Would you like me to make a reservation? (8)  
 H: Sure. Some place around here. (9)  
 CI: Last time you and Ms. Sentara went to Jimmy's, but Jimmy's is closed on Tuesdays. (10)  
 H: Pepe's then. (11)  
 CI: I have confirmed with her CI and reserved a table. Done. (12)

**Fig. 2.** A sample dialogue between a human H and a collaborative intelligence CI.

they transmit a pattern of neural impulses that describe the location of the stimulus and such properties as its luminance, color, shape, and motion. Once past her primary visual cortex, that data continues to be transformed and streamed by two distinct pathways [10]. In the parietal region, which also has a map of the person's body and its position in space, what she has just seen controls her movements while, at the same time, her temporal lobe constructs the objects and people she has seen. Thus, vision is an intrinsic part not only of recognizing what we see but also of deciding what can be done with it and how to do so, and these processes occur in parallel.

People also clearly apply their perceptions for multiple purposes. People use sound to communicate, to alert, and to convey emotion, all of which other people detect by hearing. People use vision to identify objects, situations, and appropriate responses, and to detect emotion in other people. By comparison, camera-equipped computers, even when supported by other kinds of sensors, are severely impoverished.

Furthermore, within their (ordinary) perceptual flood, people routinely focus their own attention and that of others. For example, master Go players, despite hundreds of possible moves, look only at the three or four that are most promising, and detect those few in a fraction of a second [25]. In addition, human gestures are fast, expressive, and cogent. Pointing, for example, is a way to focus the attention of another agent on an object or situation of interest. In contrast, the inability of a program confronted by large quantities of data to focus its attention appropriately is a well-known concern in AI. Finally, people engage with multiple agents, reference multiple data sources, and address multiple goals. To do so, they often employ multiple reasoning methods and multiple heuristics.

In summary, there are dramatic differences between our proposed pair of collaborators: a person and a machine. They perceive the world differently, they compute differently, and they communicate differently. If a computer is to collaborate with a person, there are a number of key issues to address, described next.

### 3. An extended example of a CI

As envisioned here, a CI is not a general intelligence. Each CI would target problem areas in which it could assist people, and provide representations and procedures to support particular human activities. Many of the key issues for a CI, however, have already received preliminary research attention. We take now as a running example, the sample dialogue with an appointment secretary in Fig. 2. Line 1 is an alert to the system by the human H; line 2 is a response from the collaborative intelligence CI.

First, a CI needs a clear statement of the shared task. There are two ways this might proceed: either the person poses the problem in some formal language that is transparent to the computer, or the computer queries the person for a description that it can manipulate. Since assistance is our motivation here, the communicative burden is more appropriately placed on the machine. Given an ontology of tasks and their components, the computer would assume that any new conversation is a request for help, and seek to elicit from the person some clarification of the task at hand, including the person's goals and relevant constraints. Line 3 could be parsed as a reference to a known task with clear goals and constraints, taken from their shared ontology.

Once the task is well defined, the CI needs to assemble the necessary facts. Its knowledge base would provide an appointment calendar. Its scheduling algorithm would ensure that appointments do not overlap, allow for travel time and typical duration, and possibly consider similar previous or future appointments. These are reflected in lines 4 and 6.

The selection of relevant facts is to some extent task-dependent. Enough time to travel to a destination (the restaurant and then the meeting) and enough time to complete a task (here, eating) are fundamentals whose range of values should be part of a commonsense knowledge base. Note that, in line 10, CI checked to see if the restaurant it considered most probable was open on Tuesday *before* it offered it to H. This is an example of a proactive CI, whose queries to its knowledge sources are aware of the context in which they are made.

A CI should also tailor itself to the needs of the person it serves. For example, in line 10, the CI might have checked where H often eats instead of where H has eaten with Alice, or whether H is already scheduled to eat at Jimmy's soon before or soon after lunch on Tuesday. When it perceives an ambiguity, or early in its experience with a particular person, a CI should ask more questions, such as "Do you mean Alice Sentara?" or "Do you want to eat there two days in a row?" As it learns the person's preferences, the CI should eventually ask fewer questions. The learning envisioned here is not complex – probability distributions over the person's history would likely suffice. For example, a commonsense knowledge

base may know “lunch is an activity that occurs in an eating place for between 3 and 180 minutes between 11 AM and 3 PM.” More specific ranges of values and Alice’s full name are easy to learn; restaurant features preferred by a particular person could be represented with a confidence level or a probability distribution, so that the CI can offer alternatives when appropriate. The degree to which the CI is decisive (e.g., when to suggest alternatives) is also learnable. Good collaborators do not dictate.

Periodically, a CI should consolidate and consider its own knowledge store. It should analyze its collaboration history offline (i.e., when not engaged in a task) to detect flaws, gaps, or speedups. If, for example, it notes that a particular question is often met with an unanticipated response, or that it must be told exactly what to do in a given situation, the task knowledge may require correction. (Recognition of negative human emotion would also be a good indicator.) Such gradual development would support human expansion of the ontology. If the CI notes that some value is perfectly consistent over an extended period (e.g., H always eats at Jimmy’s) it may ask the person whether it may safely make that assumption. Such adaptations would supportively tailor its behavior without overstepping its bounds.

Although a core facet of collaboration is communication, this paper does not envision a general dialogue agent. Dialogue with a CI is intentionally formulaic and controlled. Within the communication it receives, a CI will detect only agreement (e.g., “Sure”), disagreement (e.g., “nope”), questions (e.g., by intonation), and context-dependent instruction (e.g., “have lunch” will trigger a scheduling algorithm). Moreover, a CI’s replies will be instantiable but formulaic.

At every stage in its development, a CI requires evaluation by human subjects who are, after all, the ultimate authority on its success. Note that, in line 12, H’s CI contacts Alice’s CI to confirm and then notifies H that the task is complete. Although the scenario in Fig. 2 could have become a negotiation between two calendar agents, that is not how collaborative intelligence is envisioned here. Final decisions remain the purview of the people involved.

A CI is an active, not a passive, collaborator. Consider, for example, the differences between the proactive CI in Fig. 2 and its ancestor, the Calendar APprentice (CAP), which also scheduled appointments [19]. Offline, CAP learned thousands of user-customized, user-understandable rules from five years of data on each user. Each rule explained some feature of user-scheduled meetings (e.g., day of week or time of day) given others (e.g., participants). CAP was expected to combine and refine those rules over time, and to become increasingly autonomous, so that eventually it could schedule a meeting the way the user would. Without a computer–user dialogue, however, CAP’s scheduling often proved unsatisfactory. This is because a user’s motivation for a meeting to some extent drives that meeting’s features. Without sufficient knowledge, accurate generalizations cannot be formulated. CAP had no access to the user’s motivation and could not ask about it.

Although it is unreasonable to expect a CI to detect the subtleties that dictate how people make decisions, a CI that detected inconsistencies in observed patterns could certainly ask questions to improve its performance. For example, as electronic calendars become the norm, it is increasingly likely that requests for “personal” and “business” meetings will be directed to the same CI. If a user meets someone to plan a project one day and meets the same person to play tennis the next, a CI that asks for clarification (e.g., “isn’t Saturday an unusual meeting time?”) has the opportunity to learn a new feature value and new associations for it.

We are not far from much of Fig. 2. In response to a spoken query, my smartphone can tell me where I am, what I have scheduled for today and for next week, and what time my dinner appointment is tonight. When told “Make a reservation for me for dinner tomorrow,” it replied, “These restaurants have tables for two around 6 pm. Tap the one you’d like to reserve.” The same smartphone, however, made 15 restaurant suggestions when I asked it *when* I could have lunch today, and returned the same 15 when I asked if I was *free* for lunch today. (Presumably, it was more concerned about when the restaurants were open.) It can address times and locations, but it needs much more commonsense knowledge.

In summary, a CI requires procedures for knowledge elicitation, knowledge extraction, evidence combination, learning, and metareasoning, as well as a substantial knowledge base with an ontology. Such a knowledge base would admittedly be extensive, but its gradual development should generalize well at appropriate levels of granularity. Eventually, for example, the approach that created the appointments secretary could serve to arrange more complex meetings among more than two people. The next section proposes three reasonably approachable but difficult CI tasks.

#### 4. Challenge problems

CI has the potential to make productive and welcome contributions in many human endeavors. The RoboCup@Home<sup>7</sup> and RoboCup@Work<sup>8</sup> challenges have highlighted the need for systems that collaborate with people, not replace them, to address the noise and uncertainty of the real world. A CI, however, does not require an embodied (i.e., robotic) presence. This section describes three CIs that would provide positive assistance to people: a virtual emergency room, a collaborative organizer, and a collaborative designer. Each of them requires the integration of sensory modalities and dialogue with reasoning and problem solving. Moreover, and most importantly, each of them would support a better outcome for the human user.

<sup>7</sup> <http://www.robocupathome.org/>, January 7, 2015.

<sup>8</sup> <http://www.robocupatwork.org/>, January 7, 2015.

#### 4.1. The virtual emergency room

A *virtual emergency room* would provide remote first aid. Such a CI is intended to support people who lack the time, resources, or inclination to consult a medical professional. (As a side benefit, it could reduce the burden on healthcare systems.) This CI would have a sensing station that could detect, for example, whether the color, scent, and/or temperature of the skin surrounding a wound indicated a likely infection. It would be sensitive to emotion as expressed in the person's voice and face. It would also address its knowledge about the individual: age, frequency of recourse to the CI, and medical history.

The virtual emergency room is not intended as a substitute medical professional, merely as a screener and information resource. It would address “Is there something I should do given this condition?” followed by “Is there something I can do myself?” It would be able to track data and offer advice over a period of time (e.g., monitor changes in a wound), ask diagnostic questions (e.g., “is there numbness?”), and present its rationale for a course of action (e.g., “Because of the color and pain, I think this is infected. Your alternatives are...”). As it deemed necessary, the CI could contact a medical professional for guidance or even summon help.

Note that the CI would not make decisions for the person. Instead it would offer information from its knowledge base, given what it perceives. Initially, such a CI could be placed in convenient locations, or even in emergency rooms themselves, where it could take preliminary medical information from the prospective patient. Eventually, we might all have a virtual emergency room in our workplace, school, or home, to which we turned for quick, inexpensive advice. Success would be measured by the same standards to which we hold doctors and hospitals, including accuracy of diagnosis, cure rate, recurrence, and complications.

#### 4.2. The collaborative organizer

A *collaborative organizer* would be a greatly enhanced version of the appointment secretary in Fig. 2. A collaborative organizer is intended to support arrangements that bring people together, for example, for a board meeting or a scientific conference. It would solicit information from the person about the crucial features of a meeting (e.g., location, time, attendees).

This CI would work at multiple levels of granularity. First, it would identify the person's priorities among what are likely to be multiple goals (e.g., conflicts, convenience, budget). As decisions are made, the CI could ask for further refinements (e.g., from “in 2016” to “in summer, 2016” to “about a week after most European universities' spring terms end”). It would remind the person about ancillary services (e.g., tech support, catering, photography). The CI would also have data on similar past meetings, including those arranged by the current user and by other users with their own CIs. It would understand concurrent events, ordering, and the inability of a person or an object to be in two places or serve two purposes simultaneously. Moreover, it would understand the need to revisit prior decisions that are impacted by current ones. For example, funds should be reallocated based on an overall budget; collocated, preliminary, and follow-on events must be addressed when dates or locations change. An interesting mathematical challenge here is appropriate pricing, which must address multiple, likely conflicting goals (e.g., attendance and profit). To help the person envision the meeting as it takes shape, the CI would provide a variety of templates, checklists, and interactive displays for budgets, schedules, assignments, and advertisements.

A collaborative organizer would accept instructions and offer suggestions but, once again, a person would guide the process and make the final decisions. Success would be measured by questionnaires that gauge the user's satisfaction not only with respect to the task's stated goals, but also with respect to the quality of the CI as a collaborator, including its communicative clarity and ease of use. This requires thoughtful empirical design, supported by insights from psychologists with expertise in how to elicit data from people.

#### 4.3. The collaborative designer

A *collaborative designer* would be a greatly enhanced version of the soup assistant in Section 3. The construction of an object, be it a building, a park, or a bridge, effectively instantiates a template with values. Construction of a complex object requires raw material, a variety of processes, a partial-order planner, and the ability to address such diverse (and likely conflicting) evaluation criteria as safety, economy, and novelty. This CI would be expected to solicit feature values from the person (e.g., height, cost) and to have prior knowledge of fundamental processes for the object's construction (e.g., permits, excavation, foundation). Together, the collaborative designer and the person would either evaluate the object after it is assembled or evaluate the design in simulation.

Rather than invent an entirely new object, this CI would likely present prototypes to the person, and accept and suggest changes to them (e.g., model a bridge after an existing one). This requires a large knowledge base of objects accessible through multiple indices (e.g., heavily-trafficked, short length, cable-stayed bridges) and the ability to examine three-dimensional representations. Success would be measured by comparison of the finished product to its original specifications, and again by user satisfaction along multiple dimensions, including degree of responsiveness to specified goals and how easy the CI is to use.

#### 4.4. Discussion

Regardless of the task, a crucial evaluation metric for any collaborative intelligence, in my opinion, must be the extent to which a person believes that her work experience or product has been facilitated or improved by the collaboration. A successful CI will be welcome and used; it will be refined based on user requests and hypotheses it formulates for the person to confirm. It will be a model of shared understanding that allows two very different kinds of problem solvers to collaborate on interesting tasks.

The challenge problems above are increasingly difficult. A virtual emergency room is fundamentally a reactive, diagnostic tool. It requires only relatively constrained dialogue, but introduces a variety of perceptive modalities, communication requirements, and a knowledge base that includes commonsense as well as historical and scientific knowledge. The collaborative organizer introduces multiple goals and requires multiple skills at various levels of granularity. It also requires a more elaborate knowledge base of precedents, and must contend with constraints and prolonged interaction with the user over days or months of planning. The collaborative designer presents the greatest challenge. It requires a more nuanced representation of time, a still larger knowledge base, multiple ways to represent the object to the user, and explicit specification of a broad range of evaluation criteria, including government regulations and aesthetic criteria.

The order of these three problems deliberately provides a developmental spectrum to support method transfer. Their dialogue requirements, user models, and knowledge requirements are increasingly more complex. Thus approaches from the simplest one, such as appropriate responses to emotion and detection of changed goals, will remain useful for subsequent CIs.

Each of these CIs has obvious simpler versions from which to begin. For the virtual emergency room, one might build an extensible set of medical cases and learn to classify them, beginning perhaps with an expert-labeled set of wound data, that include multiple images and heat maps of the same wound. For the collaborative organizer, one might begin with the appointment secretary in Fig. 2, and extend it to a meeting among more than two members of the same organization who require several support services. For the collaborative designer, one might begin with a simple artifact (e.g., a storage shed or a swimming pool) and then consider how the required processes and features can be generalized to identify the essential components of structural development.

Finally, at the core of collaboration is dialogue. The dialogue required of a CI, however, is not a full-blown system that purports to “understand.” To serve, a CI need only respond appropriately and ask constructive questions. The person’s intonation and facial expression should provide important guidance.

#### 5. Conclusion

Although tales of human–computer collaboration are rampant in science fiction, few artifacts seek to combine the best talents of a person and a computer. There are at least two plausible reasons for this gap. First, it is easier not to include people in empirical work. Because people are non-uniform, costly, slow, error-prone, and sometimes irrational, properly designed empirical investigations with them are considerably more complex. Nonetheless, the ultimate judge of a CI’s success must be the person it is intended to assist. Second, the original vision for AI foresaw an autonomous machine. We have argued here, however, that a machine that shares a task with a person requires all the behaviors the Dartmouth proposal targeted, plus one more – the ability to collaborate on a common goal.

At this point in AI’s development, it is constructive to re-evaluate the significance of difficult tasks and human simulation. Just because an artifact can do something very difficult does not make it useful to people. Although it may be enjoyable to watch an AI program clash wits with a person, collaborative artifacts would be both considerably more useful and better able to attract societal support. Moreover, just because a computer can be engineered to do something as well as a person may not be a good enough reason to have the computer do it. A CoBot, for example, could be engineered to press elevator buttons, but it would then require arm-like appendages that might complicate its ability to maneuver past people and other CoBots in narrow passageways. Rather than re-design our world for computers or submit to their decisions, we should begin to share our tasks with them. Indeed, recent evidence suggests that access to technology like the Web and Wikipedia actually contributes positively to human decision making [27].

“We think,” read the Dartmouth manifesto, “that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.” It will assuredly take us longer to produce CIs that respond in real time and learn to support human endeavors. AI has already spawned subfields that address key CI issues. Now CI offers us the opportunity to reunite these productive but isolated research communities. CI integrates multiple sensory modalities with what we know about intelligent software agents, human–computer interfaces, knowledge and evidence, learning, and metareasoning.

Successful CIs could establish a synergy between people and computers to accomplish human goals. Computers would do what they do best (or what people would prefer not to do at all), while people would reserve to themselves the ability to set priorities and to deal with the plethora of unforeseen situations that arise in a shared, dynamic world. A new focus on artificial intelligence that collaborates with people would incorporate and ultimately strengthen the advances AI has already made, improve AI’s public image, and provide AI researchers with a host of interesting and productive challenges.

## Acknowledgements

Thanks go to Eric Osisek, who carefully preprocessed the text for the Wordles. Manuela Veloso, colleagues at the recent AI Summit, an anonymous reviewer, and the students in Hunter College's Machine Learning Lab offered many constructive suggestions. This material is based upon work supported by the National Science Foundation under IIS-1117000.

## References

- [1] R. Alterovitz, S. Koenig, M. Likhachev, Robot planning in the real world: research challenges and opportunities, NSF/NRI-sponsored Workshop Report, National Science Foundation, 2014.
- [2] Anonymous, Don't be evil, genius, *Economist* (2014) 57.
- [3] C. Arthur, A 'Babelfish' could be the web's next big thing, says AI expert, *Guardian* (2014).
- [4] D.H. Autor, Polanyi's paradox and the shape of employment growth, in: The Federal Reserve Bank of Kansas, Jackson Hole Conference, 2014.
- [5] K. Chayka, Soldiers, machines and the rise of battlefield robotics, *Syd. Morning Her.* (2014).
- [6] B. Coltin, M. Veloso, Optimizing for transfers in a multi-vehicle collection and delivery problem, in: Proceedings of the 11th International Symposium on Distributed Autonomous Robotic Systems, DARS'12, Baltimore, USA, 2012.
- [7] D. Crevier, *AI: The Tumultuous Search for Artificial Intelligence*, Basic Books, New York, NY, 1993.
- [8] K. Drum, Welcome Robot Overlords. Please don't fire us? Smart machines probably won't kill us all – but they'll definitely take our jobs, and sooner than you think, *Mother Jones* May/June (2013).
- [9] P.C. Gilmore, A proof method for quantification theory: its justification and realization, *IBM J. Res. Dev.* 4 (1) (1960) 28–35.
- [10] M.A. Goodale, J.P. Meenan, H.H. Bülthoff, D.A. Nicolle, K.J. Murphy, C.I. Racicot, Separate neural pathways for the visual analysis of object shape in perception and prehension, *Curr. Biol.* 4 (7) (1994) 604–661.
- [11] J. Heggstuen, One in every 5 people in the world own a smartphone, one in every 17 own a tablet, cited 2014, available from: <http://www.businessinsider.com/smartphone-and-tablet-penetration-2013-10>, 2013.
- [12] B. Katz, G. Borchardt, S. Felshin, Natural language annotations for question answering, in: Proceedings of the 19th International FLAIRS Conference, FLAIRS 2006, Melbourne Beach, FL, 2006.
- [13] Y. Lejacq, Anki drive review: taking artificial intelligence for a test drive, available from: <http://www.nbcnews.com/technology/anki-drive-review-taking-artificial-intelligence-test-drive-2D11758886>, 2013.
- [14] H. Levesque, On our best behavior, *Artif. Intell.* 212 (2014) 27–35.
- [15] G. Marcus, Hying artificial intelligence, yet again, in elements: science, technology, and the things that make up our world, *New Yorker* (2014).
- [16] J. McCarthy, *The advice taker*, in: *Semantic Information Processing*, MIT Press, Cambridge, MA, 1968.
- [17] P. McGuire, The rise of robot journalists and the world brain: beware of the media industrial complex, available from: <http://www.newswithviews.com/McGuire/paul204.htm>, 2014.
- [18] C.C. Miller, Will you lose your job to a robot? Silicon Valley is split, *N.Y. Times* (2014).
- [19] T.M. Mitchell, R. Caruana, D. Freitag, J. McDermott, D. Zabowski, Experience with a learning personal assistant, *Commun. ACM* (1994) 81–91.
- [20] A. Newell, J.C. Shaw, H.A. Simon, Report on a general problem-solving program, Rand Cooperation, 1958.
- [21] G. Nunberg, Do feelings compute? If not, the Turing test doesn't mean much, available from: <http://www.npr.org/blogs/alltechconsidered/2014/07/01/323984864/do-feelings-compute-if-not-the-turing-test-doesnt-mean-much>, 2014.
- [22] K. Panton, C. Matuszek, D. Lenat, D. Schneider, M. Witbrock, N. Siegel, B. Shepard, Common sense reasoning – from CYC to intelligent assistant, in: Y. Cal, J. Abascal (Eds.), *Ambient Intelligence in Everyday Life*, 2006, pp. 1–31.
- [23] Turing test success marks milestone in computing history, University of Reading, 2014, available from: <http://www.reading.ac.uk/news-and-events/releases/PR583836.aspx>.
- [24] M. Roemmele, C.A. Bejan, A.S. Grdon, Choice of plausible alternatives: an evaluation of commonsense causal reasoning, in: Proceedings of the AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning, AAAI: Stanford University, 2011.
- [25] Y. Saito, A. Yoshikawa, Go as a testbed for cognitive science studies, in: Proceedings of IJCAI-97 Workshop on Using Games as an Experimental Testbed for AI Research, 1997.
- [26] G. Tesauro, D.C. Gondek, J. Lenchner, J. Fan, J.M. Prager, Analysis of Watson's strategies for playing Jeopardy!, *J. Artif. Intell. Res.* (2013) 205–251.
- [27] C. Thompson, *Smarter than You Think: How Technology Is Changing Our Minds for the Better*, The Penguin Press, 2013.
- [28] A. Turing, Computing machinery and intelligence, in: E.A. Feigenbaum (Ed.), *Computers and Thought*, McGraw-Hill, New York, 1963.
- [29] A. Tversky, Features of similarity, *Psychol. Rev.* 84 (4) (1977) 327–352.
- [30] H. Wang, Toward mechanical mathematics, *IBM J. Res. Dev.* 4 (1) (1960) 2–22.
- [31] A.N. Whitehead, B. Russell, *Principia Mathematica*, Cambridge University Press, Cambridge, 1910.