The Sentient Web

Michael N. Huhns • *University of South Carolina* • *huhns@sc.edu*

"In a startling revelation, a team of university scientists has reported that a network of computers has become conscious and sentient, and is beginning to assume control of online information systems." Thus begins the plot for many recent science fiction novels. In spite of the ominous tone typically chosen for dramatic effect, a sentient Web would be more helpful and much easier for people to use. When people find the Web frustrating, it is usually not because the Web seems too smart, but because it is unintuitive and unintelligent.

This fictional scenario could become reality — the Web has a huge memory capacity and online computational ability and resources are not a limiting factor – but how likely is it? What is the future possibility for some form of sentience, and what is the current sentient capability of the Web and Web technology?

Consciousness, Awareness, and Free Will

Philosophers have debated the concept of consciousness for centuries, and thousands of books and papers have been written about it (David Chalmers has a Web site devoted to referencing these at www.u.arizona.edu/~chalmers/online.html). Four things characterize being conscious:

- knowing,
- having intentions,
- introspecting, and
- experiencing phenomena.

For the first two, it is easy to show that most Web entities possess and demonstrate the use of knowledge, and other entities, including Web services, exhibit intentions. The last two, introspection and phenomenal experience, are facets of awareness and are not as obvious in current Web systems, so I will consider them more thoroughly.

Introspection is often thought to deliver humans' primary knowledge of their mental life. In addition to outwardly perceiving nonmental entities, such as a rock in their environment, people inwardly perceive their own mental entities, as when they "see" visual images using their imagination. Some perceptions arise by involuntary subconscious processes, for example, realizing that your arm itches. Others arise as a result of mental actions, for example, an agent choosing to observe its own decision-making. The latter are the results of introspection and thus constitute *self-consciousness*. Formal approaches to introspection treat it either as always *retrospective* — as a mental state that reflexively represents itself — or as distinctly available for linguistic or rational processing, even if it is not itself perceived or otherwise thought about by the conscious entity. (For more, see Eric Lormand's page on consciousness at www-personal.umich.edu/~lormand/phil/cons/consciousness.htm.)

As an example of human introspection (from John McCarthy's page on robots and consciousness, www-formal.stanford.edu/jmc/consciousness/consciousness.html), suppose I ask you, "Is the president of the United States standing, sitting, or lying down at the moment?" You will probably answer that you don't know. If I then say, "Think harder about it," you will probably reply, "No amount of thinking will help." Introspection is required in order to give this answer, and Web agents will need a similar ability if they are to correctly decide whether to think more about a question or to seek externally the information they require.

An experience or other mental entity is *phenomenally conscious* just in case there is "something it is like" for one to have it, for example, you know what it is like to have a stomach ache. The common kinds are:

- perceptual experiences, such as tasting and seeing;

 Published by the IEEE Computer Society 1089-7801/03/$17.00

- bodily-sensational experiences, such as those of pains, tickles, and itches;
- imaginative experiences, such as those of your own actions or perceptions; and
- streams of thought, as in the experience of thinking in words or in images.

When faced with a choice, an agent with free will has the following experience: first, what the agent ultimately does is what it chooses to do, second, its choice precedes its action, and third, it could have done otherwise (for example, it could have changed its mind at the last minute).[1]

Web Capabilities for Sentience

With approximately 1 billion PCs and 30 billion processors in existence that execute more than an estimated quadrillion CPU cycles per second, and with thousands of petabytes of storage available, the Web has more than enough raw cognitive capacity for human-level sentience. (A human brain has an estimated processing power of 300 million MIPS.[2]) However, the Web has impoverished sensing and effecting capabilities. With relatively few physical sensors, it lacks a sense of space, place, and time. Moreover, the Web has even fewer physical devices that can perform actions, make changes, or conduct experiments in the real world.

A Thought Experiment

Consider a typical thermostat with a single temperature sensor and a controller for turning a heater on and off. Now imagine the thermostat had an unbounded cognitive capability for remembering and reasoning: what could the thermostat come to know?

It might first deduce a cause-and-effect relationship between turning on the heater and a rise in temperature. It might also notice that sometimes the cause-and-effect relationship does not hold, as when the heater is turned off for repair, the weather changes suddenly, or a door is left open — none of which can the thermostat sense directly.

Given enough time, it might induce the existence of daily and yearly temperature cycles. Of course, that would require a means for sensing time, which might be achieved by giving the thermostat another sensor, such as a clock, or better, the ability to count its own cycles — a rudimentary form of self-awareness or introspection.

Given free will, the thermostat could decide not to turn on the heater when the temperature dropped below a threshold. Computationally, this means the thermostat could take one of two possible actions: turning the heater on (or off) and deciding to turn the heater on (or off). Representing and then remembering its decisions provides the thermostat with a first level of introspection.

If it were endowed with curiosity, the thermostat might discover the maximum temperature that can be reached when the heater is left on indefinitely, or the minimum temperature when the heater is left off indefinitely, or what happens if the heater is cycled on and off rapidly. (Such experimenting is essentially what Doug Lenat's Automated Mathematician (AM) program and its successor, Eurisko, did in the domains of number theory and the learning of heuristics in rule-based knowledge, respectively. [3])

To perform such experiments, the thermostat would need the ability to alter its own algorithm for how it responds to temperature changes. This might be done by giving it two copies of the algorithm — an original for reference and a second one that it could alter — and then allowing it a third possible action: altering the algorithm. Such curiosity is not essential for consciousness, but it would imply that the thermostat had a level of awareness about its temperature-controlling functionality.

The thermostat would soon run out of experiments it could try, however — not because of a cognitive limit, but because its sensing and acting domains are so limited. Many more interesting things would happen if we now connected this thermostat to the rest of the Web, but the limitation would be the Web's relatively few physical actions and experiments it could perform.

Building Sentience

The building blocks from which we can describe consciousness are actions, sensory information, memory, attention (the ability to decide about actions and the decision-making process), and time (both real and virtual). That is, consciousness involves inputs, memories, ongoing internal processes, and output intentions.

Similar to the cognitive thermostat, each Web entity would require the five blocks, along with a representation of its information and actions in a form that it can use to communicate with other entities.

There is an obvious intersection between these building blocks and the basic definition of a software agent: an agent is an active, persistent software component that perceives, reasons, and acts, and whose actions include communication. Agents inherently take intentional actions based on sensory information and memories of past actions. (See the "Seven Years of Agents on the Web" sidebar for a description of how previous columns have examined similar aspects of agents.) All agents have the necessary communication ability, but they do not necessarily possess introspective capabilities or awareness of time.

As fundamental as time is to an agent situated in the Web or the real world, it has largely been neglected in agent research to date. For example, time is not considered in the Foundation for Intelligent Physical Agents (FIPA) specification for an agent-communication language. Similarly, agents do not typically maintain models of their own reasoning, which they need for introspection.

Seven Years of Agents on the Web

This column, Agents on the Web, has appeared 38 times since the inaugural issue of *IEEE Internet Computing* in 1997. During that time, my various coauthors and I have addressed numerous aspects of Internet-situated agents and their applications. We have also addressed agent characteristics, such as their sensing, reasoning, and communicating abilities. This final column, which will be replaced by an Agents track that will feature a more in-depth article in each issue in 2004, provides an opportunity for some introspection.

We have shown how agents can make and execute commitments based on their beliefs and intentions; how they can be embedded into environments; how they can potentially be endowed with philosophies that guide their actions; how agents can negotiate with each other and with people; and how they can reach a mutual understanding by reconciling their individual ontologies and even learning a language. The focus of research in this area has now shifted

- from how agents can cooperate with each other to how much they should cooperate;
- from what language agents can use to communicate to from what and when agents should communicate;
- from how agents can learn to what should be learned;
- from how we can describe and reason about agent commitments to how commitments can be monitored and enforced; and
- from how we can implement agents to how we can use them to implement other systems.

During the course of these seven years, we have made progress on all agent characteristics, to the point where system developers now routinely apply and rely on agents. Agents are used to fight spam, monitor firewalls for intrusions, manage network connections, automate online auctions and supply chains, implement knowledge portals, and aid individuals searching for information and shopping. I have enjoyed writing about agents because they have been shown to be useful and because I believe they will become the basis for most programming and computing in the near future.

With the support of efforts such as DARPA's Agent Markup Language (DAML) and Control of Agent-Based Systems (CoABS), the agent paradigm has been instrumental in moving the Web from a resource for people to the Semantic Web — a resource for computers. Soon it will enable the "pragmatic Web" — a resource for making distributed decisions and getting work done. In the end, it just might bring about a sentient Web.

Future Prospects

Work on developing human-like sentience and consciousness for the Web and other networked systems is just beginning. There is also heated debate about whether a computational system can ever attain such consciousness, or is merely limited to mimicking it.[4] Whether it attains or mimics consciousness, the Web will continue to become a richer and more interesting place for people and machines to use, and for agents to reside.

Acknowledgment

The US National Science Foundation supported this work under grant no. IIS-0083362.

References

1. S. Harnad, "Consciousness: An Afterthought," *Cognition and Brain Theory*, vol. 5, no. 1, 1982, pp. 29–47.
2. H. Moravec, "Robots, After All," *Comm.ACM*, vol. 46, no. 10, October 2003, pp. 91–97.
3. D.B. Lenat and J.S. Brown, "Why AM and EURISKO Appear to Work," *Artificial Intelligence*, vol. 23, no. 3, 1984, pp. 269–294.
4. D.C. Dennett, "Caveat Emptor," *Consciousness and Cognition*, vol. 2, no. 1, 1993, pp. 48–57.

Michael N. Huhns is a professor of computer science and engineering at the University of South Carolina, where he also directs the Center for Information Technology.