Agents on Stage: Advancing the State of the Art of AI

Extended Abstract for an Invited Talk

Barbara Hayes-Roth
Computer Science Department
Stanford University
701 Welch Road, Bldg C
Palo Alto, CA 94304
U.S.A.

1. Introduction

Intelligent computer agents are both the original goal and the ultimate goal of artificial intelligence research. In striving toward that goal, our community has followed a practical research strategy of "divide-and-conquer," with different sub-communities attacking important component functions of intelligence, such as planning, search, knowledge representation, vision, and natural language. This strategy has been almost too successful, yielding both challenging theoretical problems that have come to dominate the inquiries of many researchers and a spate of practical techniques that have enabled other researchers to move into profitable commercial enterprises. While each of these pursuits is worthy in its own right, together they have had the unfortunate side-effect of fragmenting the field and almost completely diverting it from efforts to build intelligent agents.

I believe that the AI community should return its attention to the science and technology of intelligent agents. Taking the title of my talk *metaphorically*, I suggest that: (a) the renewal of interest already has begun, with agent research and applications popping up everywhere; and (b) this is the best way for us to advance the state of the technical art of AI. Taking my title *literally*, I suggest further that: (a) we expand our purview to include a promising new class of intelligent agents—characters and actors whose performances are designed to create particular kinds of impressions on their human audiences [Bates, 1994; Bates, etal, 1994a; 1995]; and (b) the effective design and application of such agents will require us to expand our requirements specification for intelligent agents and to include artistic techniques among our design strategies. The following sections explain these metaphorical and literal meanings of my title and outline the remainder of my talk.

2. The Metaphorical Interpretation

The last several years have witnessed a renewed interest in intelligent agents. In his 1990 book, *Unified Theories of Cognition*, Newell declared that we have accumulated enough component solutions to component problems and demanded that we work on integrating those solutions in a smoothly functioning whole. Reinforcing Newell's position in my review of his book [Hayes-Roth, 1993b], I suggested that divide-and-conquer is almost sure to fail as a strategy for creating a science and technology of intelligent agents. Software engineers (and other engineers) have known for decades that it is difficult, expensive, and often impossible to integrate independently-developed components in a smoothly functioning whole. Independently-developed components carry too many implicit and incompatible assumptions and too little shared infrastructure to support effective interoperation. Why should it be any different for AI components? Thus, if our goal is a science and technology of intelligent agents, *our research strategy must emphasize development of comprehensive agent architectures and principles for component integration and interoperation.*

Complementing this holistic rationale for a comprehensive approach to intelligent agents, during the last few years, we have seen a ground swell of academic research on a variety of intelligent agents (Table 1), closely followed by the appearance of "agents" in high-profile commercial software products. References to agents have begun to appear in the popular press

and agents are becoming a recognizable concept to the lay public.

Table 1. Intelligent Agents

- Problem-solving agents
  Hayes-Roth, 1995; Newell, 1990
- Personal assistants
  Etzioni & Weld, 1994; Kautz etal, 1994; Maes, 194; Mitchell etal, 1994
- Robotic agents
  Albus, 1981; Arkin, 1985; Brooks, 1986; Firby, 1992; Gat, 1990; McDermott, 1992 Hayes-Roth, 1995; Laird et al, 1989; Vere & Bickmore, 1990
- Agents for toy problems
  Agre & Chapman, 1987; Bresina & Drummond, 1990; Pollack, 1991
- Monitoring agents
  Hayes-Roth, 1989

Table 2. Research Issues for Intelligent Agents

- Agent architecture
  Brooks, 1986; Firby, 1992; Gat, 1990; Hayes-Roth, 1985; 1990; 1993; 1994; 1995; Laird, etal, 1986; Maes, 1990; Nilsson, 1989; Rosenschein & Kaelbling, 1986; Schoppers, 1986
- Integration of perception, thought, and action
  Chrisman & Simmons, 1991; Hayes-Roth, 1993c; Washington & Hayes-Roth, 1989
- Reconfigurable components of expertise
  Hayes-Roth, etal, 1995a, b
- Coordination of multiple tasks and goals
  Hayes-Roth, 1989; 1994
- Following instructions
  Gans etal, 1990; Hayes-Roth, 1986; 1994; Huffman, 1994; Webber etal, 1993
- Conversation
  Grosz & Sidner, 1988; Moore, 1994
- Multi-agent cooperation
  Charib-draa, 1992; Durfee, 1991; Gasser & Huhns, 1989; Lesser & Corkill, 1987; Sycara, 1989

Thus, in a metaphorical sense, intelligent agents are now very much "on stage"—in the research community, in the commercial software industry, and in the popular imagination. What does this mean for the AI community? Depending on one's perspective, we face either a daunting challenge or a great opportunity to design and deliver intelligent agents that fill the bill. Personally, I perceive a welcome opportunity, replete with new forms and sources of resources to fuel our progress and with stimulating and gratifying applications to be realized with our contributions.

More important, I view renewed work on intelligent agents as a salutary refocusing of the community's attention on a combination of old and new research issues (Table 2) that will take us where we ought to go as a science. Continuing the metaphorical sense of my title, I believe that keeping our "eye on the prize" [Nilsson, 1995] is the best way to "advance the state of the art"—the state of the technical art of AI.

3. The Literal Interpretation

During the last few years, there also has been an explosion of interest in the nature and applications of human-computer interaction (HCI). In my opinion, advanced HCI applications beg for the services of intelligent agents. We need agents to serve behind the scenes, orchestrating successful interactions for users. But we especially need them to appear front and center—"agents on stage" as characters or actors whose performances are designed to create desired intellectual, psychological, emotional, social, esthetic or other responses in their human audiences.

Table 3. Expanded Functional Requirements

- Perceive, think, act
  Greater frequency, variety, interleaving
- Follow directions
  More interpretation, initiative, autonomy, variability
- Opportunism
  More prevalent, intrinsically desirable
- Intelligence
  Broader, shallower, common sense, naive psychology, social convention
- Life-like qualities
  Variability, idiosyncrasy, motivation, emotion, personality
- Shared control
  Unqualified commitment to multi-agent collaboration in achieving performance objectives
- Expertise
  Process (not goal) oriented, successful (not optimal) performance

In giving agents these new roles to play, we substantially expand the set of functional

requirements they must satisfy (Table 3). We reconnect AI with its sister disciplines, including biology, cognitive science, social psychology, communications, and organization theory. And we introduce a new dimension to our work, aiming literally to "advance the state of the art of AI"—the state of the artistic techniques we bring to bear in AI agents.

4. The Remainder of the Talk

The remainder of my talk is organized as follows. First, I will discuss the conceptual commonalties and differences among "Agents, Characters, and Actors" and describe potential applications for each. After that, I will focus on how we might advance the state of the art of AI characters and actors. First I will discuss "Advancing the State of the Technical Art" and then "Advancing the State of the Artistic Technique." For each of these major topics, I will highlight a few key problem areas and present some approaches currently under investigation in the community. Finally, I will identify major technical challenges and promising opportunities for incremental progress in our efforts to develop a science and technology of intelligent agents.

5. References

[Agre and Chapman, 1987] Agre, P.E., and Chapman, D.. Pengi: An implementation of a theory of activity. In Proceedings of the National Conference on Artificial Intelligence, San Mateo, Ca: Morgan Kaufmann, 1987.

[Albus, 1981] Albus, J.S. Brains, behavior, and robotics. Peterborough, N.H. : BYTE Books, 1981.

[Bates, 1994] Bates, J. The role of emotion in believable agents. Communications of the ACM, 1994, 37, 122-125.

[Bates, etal, 1994] Bates, J., Hayes-Roth, B., Laurel, B., and Nilsson, N. Working Notes of the AAAI Symposium Menlo Park, CA: AAAI, 1994.

[Bates, etal, 1995] Bates, J., Hayes-Roth, B., and Maes, P. Interactive Story Systems. Working Notes of the AAAI Symposium, Menlo Park CA, 1994.

[Bresina and Drummond, 1990] Bresina, J., and Drummond, M. Integrating planning and reaction: A preliminary report. In J. Hendler (ed.), SRC TR 90-45, Systems Research Center, University of Maryland, 1990.

[Brooks, 1986] Brooks, R. A. A robust layered control system for a mobile robot. IEEE Journal of Robotics and Automation, 2, 14-23, 1986.

[Charib-draa, etal, 1992] Charib-draa, B., Mandiau, R., and Millot, P. Distributed artificial intelligence: An annotated bibliography. SigArt Bulletin, 3, 1992, 20-37.

[Chrisman and Simmons, 1991] Chrisman, L., and Simmons, R.. Sensible planning: Focusing perceptual attention. In Proceedings of the National Conference on Artificial Intelligence, San Mateo, Ca: Morgan Kaufmann, 1991.

[Durfee, 1991] Durfee, E. The DAI Melting Pot. IEEE Transactions on Systems, Man, and Cybernetics, 1991.

[Etzioni and Weld, 1994] Etzioni, O., and Weld, D. A softbot-based interface to the internet. Comm. of the ACM, 1994, 37, 72-79.

[Firby, 1992] Firby, J. Building symbolic primitives with continuous control routines. In Proceedings of the First International Conference on AI Planning Systems, College Park, MA, June 1992, p.62-69.

[Gans, etal, 1990] Gans, A., Confrey, T., Harvey, J., and Hayes-Roth, B. How WATCH works. Stanford University, Knowledge Systems Laboratory, Report KSL 90-03, 1990.

[Gasser and Huhns, 1987] Gasser, L., and Huhns, M. Readings in Distributed Artificial Intelligence. San Mateo, CA: Morgan Kaufmann, 1987.

[Gat, 1992] Gat, E. Integrating planning and reacting in a heterogeneous asynchronous architecture for controlling a real-world mobile robot. Proceedings of the National Conference on Artificial Intelligence, San Mateo, Ca: Morgan Kaufmann, 1992.

[Grosz and Sidner, 1988] Grosz, B. J., and Sidner, C. L. Plans for discourse. In P.R. Cohen, J. Morgan, and M. E. Pollack (eds.), Intentions in Communication, 1988.

[Hayes-Roth, 1985] Hayes-Roth, B. A blackboard architecture for control. Artificial Intelligence, 1985, 26, 251-321.

[Hayes-Roth, 1995] Hayes-Roth, B. An architecture for adaptive intelligent systems. Artificial Intelligence, in press, 1995.

[Hayes-Roth, 1990] Hayes-Roth, B. Architectural foundations for real-time performance in intelligent agents. Real-Time Systems, 2, 99-125, 1990.

[Hayes-Roth, 1993a] Hayes-Roth, B. Intelligent control. *Artificial Intelligence*, Special Retrospective Issue on Most Frequently Cited AI Journal Articles, 59, 1993, 211-218.

[Hayes-Roth, 1993b] Hayes-Roth, B. On building integrated cognitive agents: A review of Newell's Unified Theories of Cognition. *Artificial Intelligence*, 59, 1993, 329-341. Reprinted in W. Clancey, S. Smoliar, M. Stefik (Eds.), Contemplating Mind, MIT Press, 1994.

[Hayes-Roth, 1993c] Hayes-Roth, B. Opportunistic control of action. IEEE Trans. on Systems, Man, and Cybernetics, 12, 1575-1587 1993.

[Hayes-Roth, etal, 1989] Hayes-Roth, B., Washington, R., Hewett, R., Hewett, M., and Seiver, A. Intelligent monitoring and control. *Proc. of the InteJoint Conf on Artificial Intelligence*, Detroit, Mi., 1989.

[Hayes-Roth, etal, 1994a] Hayes-Roth, B., Brownston, L., and van Gent, R. Multiagent collaboration in directed improvisation. Stanford University Report KSL-94-69, 1994.

[Hayes-Roth and Larsson, 1995] Hayes-Roth, B., and Larsson, J.E. A domain-specific software architecture for a class of intelligent patient monitoring agents. Journal of Theoretical and Experimental AI, in press, 1995.

[Hayes-Roth, etal, 1995] Hayes-Roth, B., Pfleger, K., Lalanda, P., Morignot, P., and Balabanovic, M. A domain-specific software architecture for adaptive intelligent systems. IEEE Transactions on Software Engineering, in press, 1995.

[Hayes-Roth, etal, 1994b] Hayes-Roth, B., Sincoff, E., Brownston, L., Huard, R., and Lent, B. Directed improvisation. Stanford University Report KSL-94-61, 1994.

[Huffman, 1994] Huffman, S. Instructable autonomous agents. U. of Mich., PhD Thesis, 1994.

[Kautz, etal, 1994] Kautz, H.A., Selman, B., and Coen, M. Bottom-up design of software agents. Comm. of the ACM, 1994, 37, 143-147.

[Lesser and Corkill, 1987] Lesser, V., and Corkill, D.

Maes, P. Agents that reduce work and information overload. Comm. of the ACM, 1994, 37, 30-40.

[Mitchell, etal, 1994] Mitchell, T., Caruana, R., Freitag, D., McDermott, J., and Zabowski, D. Experience with a learning personal assistant. Comm. of ACM, 1994, 37, 80-91.

[Maes, 1994] Maes, P. Situated agents can have goals. Robotics and Autonomous Systems, 6, 49-70, 1990.

[McDermott, 1992] McDermott, D. Robot planning. AI Magazine, Summer, 1992.

[Moore, 1994] Moore, J. Invited Lecture, National Conference of the AAAI, 1994.

[Newell, 1990] Newell, A. Unified Theories of Cognition. Cambridge, MA: Harvard University Press, 1990.

[Nilsson, 1989] Nilsson, N. Action networks. In Proceedings of the Rochester planning workshop: From formal systems to practical systems. Rochester, N.Y.: University of Rochester, 1989.

[Nilsson, 1995] Nilsson, N. Eye on the Prize. Stanford University Technical Report, 1995.

[Pollack, 1991] Pollack, M. Computers and Thought Address, International Joint Conference of Artificial Intelligence, 1991.

[Rosenschein and Kaelbling, 1986] Rosenschein, S.J., and Kaelbling, L.P. The synthesis of digital machines with provable epistemic properties. In J. Halpern (ed.), Proceedings of the conference on theoretical aspects of reasoning about knowledge, Monterey, Ca, 83-98, 1986.

[Schoppers, 1987] Schoppers, M. Universal plans for reactive robots in unpredictable environments. In Proceedings of the International Joint Conference on Artificial Intelligence, San Matao, Ca: Morgan Kaufmann, 1987.

[Sycara, 1989] Sycara, K.

[Washington and Hayes-Roth, 1989] Washington, R., and Hayes-Roth, B. Input Data Management in Real-Time AI Systems. *Proc. of the International Joint Conference on Artificial Intelligence*, Detroit, MI, 1989.

[Webber, etal, 1993] Webber, B., Badler, N., Di Eugenio, B., Geib, C., Levison, L., and Moore, M. Instructions, intentions, and expectations. U. of Pennsylvania: Report MS-CIS-93-61, 1993.

[Vere and Bickmore, 1990] Vere, S., and Bickmore, T. A basic agent. Computational Intelligence, 6, 1990, 41-60.