AI Magazine Volume 13 Number 3 (1992) (© AAAI)

# The Second International Workshop on Human and Machine Cognition

*Eric Dietrich and Stephen Downes*

■ *The Second International Workshop on Human and Machine Cognition was held on 9–11 May 1991. Participation was limited to 40 researchers who are principally involved in computer science, philosophy, and psychology. The workshop focused on the foundational and methodological concerns of those who want to forge a robust and scientifically respectable AI and cognitive science. With the theme of "What do androids know, and when do they know it?" the positions covered a wide range and presented a lot of room for disagreement. The debate between the traditional AI and the situated cognition types and the connnectionists was a focal point for discussion during the workshop.*

The Second International Workshop on Human and Machine Cognition was held at Eden Condominiums on Perdido Key in Florida from 9–11 May 1991. It was supported by the National Science Foundation, the American Association for Artificial Intelligence, the Florida High Technology and Industry Council, and the Institute for the Interdisciplinary Study of Human and Machine Cognition at the University of West Florida. Participation was limited to 40 researchers selected from several disciplines (principally computer science, philosophy, and psychology); although this approach makes for stimulating discussion, it has resulted in a competitive review process (about a 10-percent acceptance rate).

In keeping with the modern trend in U.S. politics, the theme of the Second International Workshop on Human and Machine Cognition was, What do androids know, and when do they know it? The positions covered a wide range: "They can know only what androids can know: Android epistemology is peculiar to androids and is forever different from human epistemology"; "they can know their environments but only when they are situated in them"; "they can know everything we know (which includes imagined environments) but only when they are situated in their own environments"; "they can know everything we know but only when implemented using connectionist architectures"; and the ever-popular "they can know everything we know, provided we find the right symbolic knowledge representation and represent enough knowledge in it." With these positions, there was a lot of room for disagreement, and the participants managed to fill it. There wasn't even complete agreement that androids can know anything (we had a few "searleans" in attendance, members of the "computers-are-only-formal-symbol-manipulators" club).

*…the theme…was, What do androids know, and when do they know it?*

If the first two workshops on human and machine cognition are representative, these meetings will become hotbeds of constructive and much-needed debate. They focus on the foundational and methodological concerns of those who want to forge a robust and scientifically respectable AI and cognitive science. It is just a fact of life that there isn't much agreement about methodology and foundational issues within these two fields.

One feature of the workshop that facilitated and, at times, obstructed fruitful discussion was its highly interdisciplinary nature. The term android epistemology was coined by Clark Glymour (1987). The familiarity of this term in philosophical circles attracted a good many philosophers to the workshop. Of course, many representatives of the AI, computer science, and psychology communities were present as well as a lawyer and a physician. The interdisciplinary makeup allowed for an expansion of the scope of Glymour's original concept. One notable extension was the move from android epistemology to android ethics. The increasing popularity of these workshops with representatives from disparate disciplines bodes well for AI and the cognitive sciences.

In the first talk of the workshop, Margaret Boden presented her work on creativity in humans and computers. Boden is one of the few cognitive scientists explicitly working on creativity; her ideas are original and interesting. One of the several scheduled discussion groups was devoted to various views of creativity (including Boden's and Herb Simon's). Participants discussed whether there are distinctions between creativity in science and creativity in music. Boden proposed a distinction between *P creativity*, creativity relative to a person's own conceptual space, and *H creativity*, creativity relative to all conceptual history, to help focus discussion in both the scientific and the musical cases.

In another talk, Pat Hayes and Ken Ford presented a clean and computationally sophisticated refutation of John Searle's famous Chinese room argument. Clean, sophisticated refutations of Searle's argument are abundant now (although each has its own slant), but it is always good to have a new one—Searle's argument refuses to go away, having more lives than a cat. Hayes and Ford were responding to the debate in *Scientific American* (January 1990) between Searle and the Churchlands about whether a machine could think. Ironically, from the perspective of Hayes and Ford, Searle and the Churchlands are essentially in agreement, diverging only in their advocacy of differing favorite theories about the necessary material basis (biological versus parallel) for intelligence. They both make specific implementation features of brains a necessary condition for intelligence. As might be expected, Paul Churchland objected to this grouping.

Paul Churchland presented a controversial paper on neural nets and stereo vision entitled "A Feedforward Network for Fast Stereo Vision with a Movable Fusion Plane." Churchland's

paper provoked a debate, especially among the traditional AI researchers. Some tension was generated during discussions of his paper and Hayes and Ford's earlier one, but nothing was resolved. Much of the talk generated by these papers had to await attempted resolution at the final session, but it was apparent that the dispute between proponents of orthodox AI and proponents of connectionism was going to take the foreground.

The first morning ended with a thought-provoking paper by Mark Bickhard, a psychologist-philosopher whose sympathies are with those studying situated cognition. He argued for the impossibility of traditional approaches to representation and for a perspective that he calls "interactivism." In particular, Bickhard posited that interactivism can account for how representation could emerge from nonrepresentational phenomena—and, therefore, how representation could exist at all.

After the afternoon discussion periods, we assembled for a cookout in an area overlooking the Gulf of Mexico. After the cookout and a swim, many of the participants went sailing, while the landlubbers made their way to McGuire's Irish Pub and Brewery (the workshop provided a designated driver). There, Maggie Boden and Pat Hayes took the stage, tunefully holding forth in Irish and English song.

Eight sessions were held on the second day (four in the morning and four in the afternoon); they were entitled Reality, Representation, and Situated Cognition; Approximate Reasoning: Probabilistic, Plausible, and Otherwise; Reality, Constructivism, and Android Epistemology; Toward Computational Ethics and Reflection; Rationality and Robot-Android Epistemology; Abduction, Inference, and Epistemology; Epistemology: Android and Naturalized; and Representation and Content. The following paragraphs give some highlights from this day.

As we observed earlier, the debate between the traditional AI and the situated cognition types on the one hand and the connectionists on the other began early in the workshop. To give the reader a flavor of this debate, we offer Lynn Stein's paper "Imagination and Situated Cognition":

Stein argued that "cognition is imagined interaction." She has taken Toto, a goal-directed, mobile robot whose ability to follow walls emerges from its collection of low-level behaviors, such as avoiding obstacles, and added a level of imagination to it. The basic Toto represents landmarks in its environment with sonar configurations, compass readings, and the like. Basic Toto adds to its store of knowledge only by visiting locations; that is, Toto is sort of a mini–Bishop Berkley: If it hasn't perceived something, the thing doesn't exist for it. Stein added to Toto the ability to imagine what a never-before-encountered landmark would be like in terms of sonar configurations, compass readings, and so on. Toto's imagination capacity is implemented using the same mechanisms Toto uses to experience the world. Thus, Stein's claim is that cognition is, first and foremost, imagined sensation and action.

*…Stein's claim is that cognition is…imagined sensation and action.*

Stein's paper drew an excited response from Zenon Pylyshyn: He strongly disagreed with her interpretation of her results and then with the entire situated cognition paradigm. In the traditional view of cognitive robots, cognition sits on top of robotic capabilities. Thus, in the traditional view, robotics and cognition can be tackled separately. One advantage of the traditional view is that it licenses a two-pronged attack on the problem of an intelligent robot: One group can tackle the robot part, and another can tackle the cognition part. The central claim of the situated types is that this happy arrangement is a methodological recipe for failure.

Other interesting papers addressing situatedness and the related notions of intentionality and symbol grounding included Judea Pearl's paper, "Situated Androids in Search of External Reality"; Michael Miller's paper, "Reasoning about Appearance and Reality"; Bob Wielinga and Jacobijn Sandberg's paper, "How Android Can Situated Automata Be?"; Ronald Chrisley's paper, "Taking Embodiment Seriously: Non-Conceptual Content and Robotics"; Donald Perlis's paper, "Comparing Minds vis-a-vis Self-Concepts"; and Selmer Bringsjord's well-done paper, "Could, How Could We Tell If, and Why Should Robots Have Inner Lives?"

In another session, Henry Kyburg presented his stimulating paper entitled "How to Win an Argument." He showed argumentation and justification to be two sides of the same coin—they are both concerned with either the acceptance of a statement or the assignment of a rational degree of belief to it. Kyburg elaborated on a framework based on evidential probability and compared it to Bayesian approaches, nonmonotonic formalisms, and defeasible reasoning systems.

As we noted, the workshop extended talk of the purely epistemological into the ethical, and several papers were presented on computers and their possible ethical claims. Anatol Rapoport chaired this session and kicked it off with a well-received short paper called "The Vitalist's Last Stand." This paper was followed by James Gip's "Toward the Ethical Robot" in which he provided a brief map of ethical theories with an eye to automating ethical reasoning. An interesting computational model of ethical reasoning was presented by Jack Adams-Webber and Ford in "A Conscience for PINOCCHIO: A Computational Model of Ethical Cognition." It is based on Vladimir Lefebvre's mathematical theory for representing the formal structure of human reflexive (for example, ethical) processes. It seems that the empirical predictions of the model have proven surprisingly accurate over a wide range of situations. Ford and Adams-Webber have developed a computer program, PINOCCHIO, that implements some of these models in conducting dialogues with human respondents about ethical issues in their own lives. The models are relatively simple, but PINOCCHIO is remarkably robust. (Was PINOCCHIO situated? We couldn't decide.) The last paper presented in this session was "The Ethics of Autonomous Learning Systems" by Umar Khan.

Worthy of note as the most interdisciplinary endeavor of the workshop, apart from the workshop itself, was Michael McMillan and Donald Walter's paper "Hebbot Epistemology: Intentionality, Rationality, and Realism." They work in the Applied Epistemology Research Group at Arkansas Children's Hospital. They started out on a project that appeared

to be a straightforward expert system development task, but their work in developing a medical expert system led them into new terrain. Their connectionist system embodies elements borrowed from the philosopher Ruth Milikan, the psychologist Donald Hebb, insights from medical practice, connectionist computer architecture, a little evolutionary biology, and even some Alfred North Whitehead.

The second day concluded with a gala party (hosted by the provost of the University of West Florida) that went into the wee hours. McCarthy opined that someone ought to write a definitive history of AI now, while the memories of what really happened in those halcyon days of yore still exist.

In a final-day discussion session, Ronald Chrisley and Lynn Stein defended both connectionism and situated automata theory against attacks made by traditional AI researchers. Pylyshyn, defending what might be called the classical position in cognitive science, harked back to his paper, co-written with Jerry Fodor, on the failings of connectionism.

In a separate discussion session, areas included the philosophy of science and the potential for connectionist systems. Two philosophy of science issues were what should a scientific explanation look like in AI and what is the use of theory in building expert systems (some workers in expert systems argued that they need not engage in theoretical work, claiming that expert system research was more akin to engineering than any theoretical pursuit; we suggested this dichotomy was false). This "no need for theory in expert systems" view was strongly criticized by Ford and Adams-Webber.

Relative to connectionism, we discussed whether, as some connectionists claim, connectionist networks might one day falsify the Church-Turing thesis by being able to compute functions not computable by Turing machines. Bringsjord provided a quick but convincing argument that this task was impossible: Neural nets are computationally equivalent to cellular automata, which, in turn, are computationally equivalent to K-tape Turing machines, from which it follows that the Church-Turing thesis remains true even for neural net research. Hence, no claims about different types of computable functions for different types of machines need be entertained. Churchland then pointed out that in "Multi-Layer Feedforward Networks Are Universal Approximators," Hornik, Stinchcombe, and White (1989) presented a slow, rigorous proof that standard feedforward networks with as little as one hidden layer are capable of approximating any (Borel) measurable function. Churchland was a bit surprised that most of us were ignorant of this paper, thus demonstrating again the importance of interdisciplinary conferences such as this one to cognitive science and AI.

The final, plenary panel session was a lively one. In this panel, McCarthy reavowed his commitment to a formalist approach in AI, whereas Churchland avidly promoted connectionist approaches. (By this point in the workshop, situated automata appeared to have dropped out; it was to be a two-horse race: connectionism versus orthodox AI.) Pylyshyn was the last of the panelists to have a say, and he reaffirmed the conclusions of his and Fodor's paper (1988). Both he and McCarthy staged a sustained attack on connectionism with McCarthy claiming that the fact that we read sentences serially told against connectionism and that connectionists were locked in a 1950s paradigm. Churchland defended connectionism by arguing for a clear distinction between appearances and the actual mechanisms behind them. He drew parallels between the way conceptual change occurs in science and the way it occurs in connectionist systems to promote connectionist systems as the more correct representation of the mechanism behind psychological conceptual change. Pylyshyn argued that connectionist machines show no internal organization and charged that connectionist systems have a real problem with narrative memory. Pylyshyn relied heavily on Newell's hierarchy of weak methods in presenting his case. Finally, he claimed that whenever a connectionist machine successfully carries out a computational task, it does so by virtue of being an instantiation of a classical architecture. This claim is interesting; an apparent counterexample is William Bechtel's connectionist program that can solve simple deductive logic problems without learning any of the rules of deductive logic. It was noted that this program was given a grade of B on a freshman logic test. This fact left Kyburg cold. He wanted to know why we should be impressed by a program—connectionist or otherwise—that got B's when symbolic systems got A's.

*…the workshop extended talk of the purely epistemological into the ethical…*

We think that throughout the workshop some hints were raised about what a compromise position might look like, but such a position wouldn't produce anything like a straight connectionist or orthodox system. This debate didn't shed any more light on what an epistemology for androids might look like because the question of how the appropriate androids should be constituted proved such a great obstacle. (Apparently, therefore, epistemology and structure or constitution are related—an unusual result from a philosophical perspective.) The lack of progress on the central problem of epistemology should not deter future endeavors of this sort because most really deep philosophical problems have been around for some time; turning them into interdisciplinary projects might help focus debate, eventually leading to a resolution, but not in three days.

Resolving philosophical problems has never been the hallmark of success for workshops. This one succeeded in bringing together a group of participants who maintained intense discussion throughout in both the conference room and the local oyster bars. If the first two workshops are anything to go on, the next workshop should be one to look forward to. The May 1993 conference is tentatively entitled Knowledge, Expertise, and AI. The expanded versions of papers presented at the second workshop will be published in two volumes, one edited by Ken Ford and Clark Glymour and the other edited by Ken Ford and Pat Hayes.

## References

Fodor, J., and Pylyshyn, Z. 1988. Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition* 28(1): 3–71.

Glymour, C. 1987. Android Epistemology and the Frame Problem. In *The Robot's*

# AAAI-93 Robot Exhibition

Preliminary Call for Participation

Following the highly successful robotics exhibition at AAAI-92, AAAI is planning to hold a robot competition at the national conference in Washington D.C. in July of 1993. The purpose of this Call is to advise potential participants of the event, and to solicit input on the format of the exhibition and rules of the competition.Last year's competition was a three-stage event in which mobile robots demonstrated skills of reactivity, exploration, and directed search (a detailed description is in the Summer 1992 issue of *AI Magazine*).

Mobile robotics is an area where much of the research in diverse AI areas can be effectively and creatively combined to give interesting results. At AAAI-93, we would like to extend the competition to highlight as wide a range of robotic research as possible, and to stress the "intelligent" aspects of their behavior. In addition to mobile robots, we are also considering having a competition among robotic manipulators, either stationary or attached to mobile platforms.

If you are interested in participating, and would like to receive more detailed information about

the competition, please contact one of the following:

Kurt Konolige, Artificial Intelligence Center
SRI International, 333 Ravenswood Avenue
Menlo Park, CA 94025, (konolige@ai.sri.com)

Reid Simmons, School of Computer Science
Carnegie Mellon University, 5000 Forbes Avenue
Pittsburgh, PA 15213, (reids@cs.cmu.edu)

*Dilemma*, ed. Z. Pylyshyn, 65–75. Norwood, N.J.: Ablex.

Hornik, K.; Stinchcombe, M.; and White, H. 1989. Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks* 2:359–366.

---

**Eric Dietrich** is an assistant professor of philosophy at the State University of New York at Binghamton. His research interests are analogical reasoning and intentionality bashing. He is the editor of the *Journal of Experimental and Theoretical AI* and the forthcoming book *Thinking Computers and Virtual Persons* (Academic Press, 1993).

---

**Stephen Downes** is an assistant professor of philosophy at the University of Utah. He is currently on leave at Northwestern University on a postdoctoral fellowship. His research interests are the sociology of knowledge and the cognitive science of science.