AI and Philosophy: How Can You Know the Dancer from the Dance?

Linda World

Aaron Sloman was teaching philosophy at the University of Sussex in 1969, when he met Max Clowes. Clowes was a charismatic teacher, who had done pioneering work in computer image interpretation. Now, he was haranguing Sloman to drop the way he learned to do philosophy at Oxford and to start studying artificial intelligence instead.

Two years later, Sloman presented his first paper at the International Joint Conference on Artificial Intelligence (IJCAI 1971). This paper, "Interactions between Philosophy and Artificial Intelligence: The Role of Intuition and Non-logical Reasoning in Intelligence," in some ways extended his DPhil thesis, exploring the philosopher Immanuel Kant's ideas about the limitations of logical deduction and the non-analytic nature of some mathematical knowledge. But the paper primarily presented a philosophical critique of an influential AI paper written two years earlier—namely, "Some Philosophical Problems from the Standpoint of AI," by John McCarthy and Patrick C. Hayes.

Looking back, Sloman explains, "I tried to show how reasoning processes could be generalized to include manipulations of analogical representations that use relations to represent relations, instead of using only explicit symbols to represent relations, as in logic and algebra."

Coming of age in AI

Sloman's philosophic investigation of analogical representations and reasoning continues to this day, but his conversion to AI was complete after a year's study on a research grant at the University of Edinburgh in 1972.

"I learned more in that year in Edinburgh than in any year since about the age of four," Sloman says.

He returned to the University of Sussex in 1973, where he helped found the AI department. Five years later, he published *The Computer Revolution in Philosophy* (Harvester, 1978) in which he extolled AI's power to extend our ability to think. He also predicted that, within a few years, philosophers who weren't familiar with the main developments in AI could be "fairly accused of professional incompetence."

A design stance

Throughout a long, productive academic and research career, Sloman has continued his philosophical work to free AI from the restrictions of particular formalisms or mechanisms and to promote instead a broad, flexible approach that might match its domain.

"After I learned about AI," Sloman says, "it soon became clear to me that an intelligent system would need an architecture combining multiple interacting components performing different tasks."

He wrote the first draft of an architectural specification during the year he spent at Edinburgh and subsequently expanded it as part of his book, *The Computer Revolution in Philosophy* (www.cs.bham.ac.uk/research/cogaff/crp). In the first of the book's two parts, Sloman argues for AI as a tool for exploring mental mechanism and architecture designs and for extending "our understanding of what is possible, rather than just our understanding of what happens when."

Sloman adopts what the philosopher Daniel Dennett calls the "design stance" toward the study of mind, whereas Dennett favors what he calls the "intentional stance," advanced in his book *Brainstorms* (Penguin, 1978). According to Sloman, the intentional stance makes an unnecessary "presumption of rationality." By contrast, he says, "the design stance considers functionality, which is possible without rationality, as insects and microbes demonstrate well. So do people in much of their everyday life. Most of what they do is neither rational nor irrational, they just do it—breathing, maintaining balance while walking, playing with toys or gadgets, and so on."

The second part of *The Computer Revolution in Philosophy* begins his lifelong work, with his graduate students—at the University of Sussex until 1991, and since then at the University of Birmingham—to describe a theoretical architecture that a computing system could implement to simulate general features of the human mind.

Varieties of architectures

"It was clear to me that there was not just one 'right' kind of architecture," Sloman says, "but a wide variety of archi-

1541-1672/05/$20.00 © 2005 IEEE
Published by the IEEE Computer Society

tectures, including many produced at different times and in different environments by evolution."

He developed the concept "design space" to stress the importance of all possible architectures. He combined it with the concept "niche space," borrowed from biology, to address different sets of design requirements.

Sloman says the work split into two streams fairly early on. These are documented now in the CogAff schema and the H-CogAff architecture in the University of Birmingham Computer Science Department's Cognition and Affect Project (www.cs.bham.ac.uk/ ~axs/cogaff).

The CogAff schema provides a methodological framework for developing an ontology that can help the research community compare rival proposals for mentality. (The CogAff site includes the SimAgent toolkit for exploring alternative architectures, which Sloman designed and built with the help of its users. He also manages the open-source version of the Poplog AI development system, which originated during his tenure at Sussex and existed for several years as a commercial product.)

The CogAff schema makes a threefold distinction between mechanisms for reactive, deliberative, and metamanagement semantic processes and an orthogonal threefold distinction between perceptual, central-processing, and action mechanisms. Although Sloman describes the nine-component schema as a "crude and inadequate ontology," it nevertheless achieves an important functionality by allowing perceptual and action subsystems to evolve in parallel with central processing.

The H-CogAff architecture is a specific elaboration of the schema, which tries to explain many aspects of the human mind.

"One consequence of the design-based approach to the study of mind," Sloman says, "is that within the framework of an architecture we can generate precisely defined concepts describing many kinds of states and processes that can occur at the information-processing level." He compares this to the development of a theory of the architecture of matter, which generated new concepts about the "stuff" of matter.

Aaron Sloman. A philosophic software developer and juggler.

VMs and causation

Apart from layers of functionality, Sloman defines a layering concept related to the analytic philosophy notion of *supervenience*, which allows for mental characteristics that depend on physical characteristics without being reducible to them. Sloman relates this idea to virtual machines, which are implemented in both lower-level VMs and physical machines. He sees the significance of VMs for design-based AI in their power to cause events.

"Software engineers understand these matters and use them in their work every day," Sloman says. "But this is craft knowledge, not articulated explicitly in a manner that clarifies the philosophical issues. As a philosophical software engineer, I have tried to explain things in a way that will, I hope, clarify some debates in philosophy, AI, cognitive science, psychology, neuroscience, and biology."

Instead of a single atomic state kind of functionalism, a VM can include many subsystems, each having its own state and state transitions with some providing inputs to other subsystems.

Sloman thinks VMs are the right tool for exploring possible information-processing systems, whether artificial or natural, intelligent or dumb. Such systems can support the complexity from which all manner of mental and affective behaviors might arise.

He prefers to consider "information" as a theoretical notion (in its commonsense meaning, not the Shannon statistical sense). Like "energy," information is a notion that can't be explicitly defined in terms of unproblematic pretheoretical concepts.

"Such concepts are partly defined by a web of relationships to other concepts in a theory or collection of related theories," he explains, "and as the theories change, the concepts change."

Ontological blindness

Like many others, Sloman is concerned about fragmentation in the AI field. He and his colleague Ron Chrisley coined the term "ontological blindness" to describe the threat this can pose to research. The term recalls the six blind men who mistake the whole elephant for the part each one is touching.

"Ontological blindness keeps AI research from identifying the various subfunctions that intelligent systems need to model," Sloman says.

As a way to maintain an overall focus on intelligent systems research, he has recently proposed a scenario-based methodology for defining AI goals. In this way, he says, researchers could agree on long-term behaviors while disagreeing on how to achieve them.

Sloman sees "a deep continuity" between AI and very old problems in philosophy. Philosophy needs AI to progress in its study of difficult questions about the nature of mind. AI needs philosophy to clarify its requirements analyses.

Echoing George Santayana's famous quotation about history, he says, "Those who are ignorant of philosophy are doomed to reinvent it—badly."