Cambridge University Press
0521386039 - Computation and Human Experience
Philip E. Agre
Excerpt
More information

1 Introduction

Activity

Computational inquiry into human nature originated in the years after World War II. Scientists mobilized into wartime research had developed a series of technologies that lent themselves to anthropomorphic description, and once the war ended these technologies inspired novel forms of psychological theorizing. A servomechanism, for example, could aim a gun by continually sensing the target's location and pushing the gun in the direction needed to intercept it. Technologically sophisticated psychologists such as George Miller observed that this feedback cycle could be described in human-like terms as pursuing a purpose based on awareness of its environment and anticipation of the future.[1] New methods of signal detection could likewise be described as making perceptual discriminations, and the analytical tools of information theory soon provided mathematical ways to talk about communication. In the decades after the war, these technical ideas provided the intellectual license for a counterrevolution against behaviorism and a restoration of scientific status to human mental life. The explanatory power of these ideas lay in a suggestive confluence of metaphor, mathematics, and machinery. Metaphorical attributions of purpose were associated with the mathematics of servocontrol and realized in servomechanisms; metaphorical attributions of discrimination were associated with the mathematics of signal and noise and realized in communications equipment; and metaphorical attributions of communication were associated with the mathematics of information theory and realized in coding devices. The new psychology sought to describe human beings using vocabulary that could be metaphorically associated with technologically realizable mathematics.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521386039 - Computation and Human Experience
Philip E. Agre
Excerpt
More information

The development of the stored-program digital computer put this project into high gear. It is a commonplace that the computer contributed a potent stock of metaphors to modern psychology, but it is important to understand just how these metaphors informed the new research. The outlines of the project were the same as with servocontrol, signal detection, and information theory: a bit of metaphor attached to a bit of mathematics and realized in a machine whose operation could then be narrated using intentional vocabulary.[2] But the digital computer both generalized and circumscribed this project. By writing computer programs, one could physically realize absolutely any bit of finite mathematics one wished. The inside of the computer thus became an imaginative landscape in which programmers could physically realize an enormous variety of ideas about the nature of thought. Fertile as this project was, it was also circumscribed precisely by the boundaries of the computer. The feats of physics and chemistry that supported the digital abstraction operated inside the computer, and not outside.

In this way, a powerful dynamic of mutual reinforcement took hold between the technology of computation and a Cartesian view of human nature, with computational processes inside computers corresponding to thought processes inside minds. But the founders of computational psychology, while mostly avowed Cartesians, actually transformed Descartes's ideas in a complex and original way. They retained the radical experiential inwardness that Descartes, building on a long tradition, had painted as the human condition. And they retained the Cartesian understanding of human bodies and brains as physical objects, extended in space and subject to physical laws. Their innovation lay in a subversive reinterpretation of Descartes's ontological dualism (Gallistel 1980: 6–7). In *The Passions of the Soul,* Descartes had described the mind as an extensionless res cogitans that simultaneously participated in and transcended physical reality. The mind, in other words, interacted causally with the body, but was not itself a causal phenomenon. Sequestered in this nether region with its problematic relationship to the physical world, the mind's privileged object of contemplation was mathematics. The "clear and distinct ideas" that formed the basis of Descartes's epistemology in the *Meditations* were in the first instance *mathematical* ideas (Rorty 1979: 57–62; cf. Heidegger 1961 [1927]: 128–134). Of course, generations of mechanists beginning with Hobbes, and arguably from antiquity, had described human thought in monistic terms as the workings of machinery (Haugeland 1985: 23). But these theorists were always con-

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521386039 - Computation and Human Experience
Philip E. Agre
Excerpt
More information

strained by the primitive ideas about machinery that were available to them. Descartes's physiology suffered in this way, but not his psychology. Although they paid little heed to the prescriptive analysis of thought that Descartes had offered,[3] the founders of computational psychology nonetheless consciously adopted and reworked the broader framework of Descartes's theory, starting with a single brilliant stroke. The mind does not simply contemplate mathematics, they asserted; the mind is *itself* mathematical, and the mathematics of mind is precisely a technical specification for the causally explicable operation of the brain.

This remarkable proposal set off what is justly called a "revolution" in philosophy and psychology as well as in technology. Technology is in large measure a cultural phenomenon, and never has it been more plainly so than in the 1950s. Computational studies in that decade were studies of faculties of *intelligence* and processes of *thought,* as part of a kind of cult of cognition whose icons were the rocket scientist, the symbolism of mathematics, and the computer itself.[4] The images now strike us as dated and even camp, but we are still affected by the technical practice and the interpretation of human experience around which artificial intelligence, or AI, was first organized.

I wish to investigate this confluence of technology and human experience. The philosophical underside of technology has been deeply bound up with larger cultural movements, yet technical practitioners have generally understood themselves as responding to discrete instrumental "problems" and producing technologies that have "effects" upon the world. In this book I would like to contribute to a *critical technical practice* in which rigorous reflection upon technical ideas and practices becomes an integral part of day-to-day technical work itself.

I will proceed through a study in the intellectual history of research in AI. The point is not to exhaust the territory but to focus on certain chapters of AI's history that help illuminate the internal logic of its development as a technical practice.[5] Although it will be necessary to examine a broad range of ideas about thought, perception, knowledge, and their physical realization in digital circuitry, I will focus centrally on computational theories of action. This choice is strategic, inasmuch as action has been a structurally marginal and problematic topic in AI; the recurring difficulties in this computational research on action, carefully interpreted, motivate critiques that strike to the heart of the field as it has historically been constituted. I aim to reorient research in AI away from *cognition* – abstract processes in the head – and toward *activity* – concrete

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521386039 - Computation and Human Experience
Philip E. Agre
Excerpt
More information

undertakings in the world. This is not a different subject, but a different approach to the same subject: different metaphors, methods, technologies, prototypes, and criteria of evaluation. Effecting such a reorientation will require technical innovation, but it will also require an awareness of the structure of ideas in AI and how these ideas are bound up with the language, the methodology, and the value systems of the field.

Roughly speaking, computational research into activity seeks technical ideas about action and representation that are well suited to the special requirements of *situated, embodied agents* living in the physical world. The "agents" could be robots we would like to build or creatures we would like to understand. The word *agent,* though common in AI, does not appeal to everyone. Its advantage is its ambiguity – robots, insects, cats, and people are all agents.[6] Such vocabulary tacitly promises, of course, that computation provides useful ways of talking about robots, insects, cats, and people at the same time without reducing all of them to a bloodless technical order. In any event, I will have little to say about insects and cats. To say that an agent is *situated* is to emphasize that its actions make little sense outside of the particular situation in which it finds itself in the physical and social world; it is always provided with particular materials and involved with particular other agents. To say that an agent is *embodied* is simply to say that it has a body. Even better, following Merleau-Ponty (1962 [1945]), it *is* a body or *exists as* a body. As a physical being, it has a definite location, limited experience, and finite abilities. It is *in* the world, *among* the world's materials, and *with* other agents. The claim is not simply that these things are true (hardly anybody would deny them), but also that taking them seriously requires an overhaul of basic ideas about both computation and activity.

My project is both critical and constructive. By painting computational ideas in a larger philosophical context, I wish to ease critical dialogue between technology and the humanities and social sciences (Bolter 1984; Güzeldere and Franchi 1995). The field of AI could certainly benefit from a more sophisticated understanding of itself as a form of inquiry into human nature. In exchange, it offers a powerful mode of investigation into the practicalities and consequences of physical realization.

My recommendation of a shift of focus from cognition to activity converges with a number of other intellectual trends, each of which is also founded in a critique of Cartesianism. These include the otherwise

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521386039 - Computation and Human Experience
Philip E. Agre
Excerpt
More information

disparate traditions that descend from Heidegger's phenomenological analysis of routine activity, Vygotsky's theory of human development, and Garfinkel's studies of the interactional construction of social reality.[7] Each of these schools of thought has attempted to replace the philosophical opposition between a self-contained perceiving subject and an independent external object by describing our relationships to things as fundamentally bound up with their role in our ongoing projects, which in turn are defined by our cultures, located in forms of embodied activity, and acquired through socialization into a system of cultural practices. As AI reorients itself toward the study of activity, it will be able to engage in mutually beneficial dialogue with these traditions of research. This process begins with computational ways of thinking about routine activity.

Planning

Although the AI tradition has placed its principal emphasis on processes it conceives of as occurring entirely within the mind, there does exist a more or less conventional computational account of action. The early formulation of this account that had the most pervasive influence was George Miller, Eugene Galanter, and Karl Pribram's book, *Plans and the Structure of Behavior* (1960).[8] These authors rejected the extreme behaviorist view that the organized nature of activity results from isolated responses to isolated stimuli. Instead, they adopted the opposite extreme view that the organization of human activity results from the execution of mental structures they called Plans. Plans were *hierarchical* in the sense that a typical Plan consisted of a series of smaller sub-Plans, each of which consisted of yet smaller sub-Plans, and so forth, down to the primitive Plan steps, which one imagines to correspond to individual muscle movements. (Miller, Galanter, and Pribram capitalized the word "Plan" to distinguish their special use of it, especially in regard to the hierarchical nature of Plans, from vernacular usage. Subsequent authors have not followed this convention. I will follow it when I mean to refer specifically to Miller, Galanter, and Pribram's concept.)

What is a Plan? "A Plan is any hierarchical process in the organism that can control the order in which a sequence of operations is to be performed" (Miller et al. 1960: 16). They state, as a "scientific hypothesis" about which they are "reasonably confident," that a Plan is "essentially

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521386039 - Computation and Human Experience
Philip E. Agre
Excerpt
More information

the same as a program for a computer," a connotation the term has carried to the present day. Shortly thereafter, though, they state that "we shall also use the term 'Plan' to designate a rough sketch of some course of action, just the major topic headings in the outline, as well as the completely detailed specification of every detailed operation" (Miller et al. 1960: 17). Thus a new Plan's hierarchical structure need not initially reach down to the most primitive actions, though the hierarchy must be constructed in full detail by the time any given step of it is executed. They define execution by saying that "a creature is executing a particular Plan when in fact that Plan is controlling the sequence of operations he is carrying out" (Miller et al. 1960: 17).

Miller, Galanter, and Pribram applied the term "Plan" as broadly as they could. In considering various aspects of everyday life, they focused everywhere on elements of intentionality, regularity, and goal-directedness and interpreted each one as the manifestation of a Plan. As with the servos, radars, and codes that first inspired Miller and his contemporaries in the 1940s, the concept of a Plan combined the rhetoric of structured behavior with the formalisms of programming and proposed that the latter serve as models of biological systems. A great difficulty in evaluating this proposal is the imprecise way in which Miller, Galanter, and Pribram used words like "Plan." They demonstrated that one can find aspects of apparent planfulness in absolutely any phenomenon of human life. But in order to carry out this policy of systematic assimilation, important aspects of activity had to be consigned to peripheral vision. These marginalized aspects of activity were exactly those which the language of Plans and their execution tends to deemphasize.

These ideas had an enormous influence on AI, but with some differences of emphasis. Although they occasionally employ the term "planning," Miller, Galanter, and Pribram provide no detailed theory of the construction of new Plans. The AI tradition, by contrast, has conducted extensive research on plan construction but has generally assumed that execution is little more than a simple matter of running a computer program. What has remained is a definite view of human activity that has continued, whether implicitly or explicitly, to suffuse the rhetoric and technology of computational theories of action. In place of this view, I would like to substitute another, one that follows the anthropologically motivated theoretical orientations of Suchman (1987) and Lave (1988) in

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521386039 - Computation and Human Experience
Philip E. Agre
Excerpt
More information

emphasizing the situated nature of human action. Let me contrast the old view and the new point by point:

- Why does activity appear to be organized?

  Planning view: If someone's activity has a certain organization, that is because the person has constructed and executed a representation of that activity, namely a plan.

  Alternative: Everyday life has an orderliness, a coherence, and patterns of change that are emergent attributes of people's interactions with their worlds. Forms of activity might be influenced by representations but are by no means mechanically determined by them.

- How do people engage in activity?

  Planning view: Activity is fundamentally planned; contingency is a marginal phenomenon. People conduct their activity by constructing and executing plans.

  Alternative: Activity is fundamentally improvised; contingency is the central phenomenon. People conduct their activity by continually redeciding what to do.

- How does the world influence activity?

  Planning view: The world is fundamentally hostile, in the sense that rational action requires extensive, even exhaustive, attempts to anticipate difficulties. Life is difficult and complicated, a series of problems to be solved.

  Alternative: The world is fundamentally benign, in the sense that our cultural environment and personal experiences provide sufficient support for our cognition that, as long as we keep our eyes open, we need not take account of potential difficulties without specific grounds for concern. Life is almost wholly routine, a fabric of familiar activities.

The alternative view of human activity that I have sketched here contains a seeming tension: how can activity be both improvised and routine? The answer is that the routine of everyday life is not a matter of performing precisely the same actions every day, as if one were a clockwork device executing a plan. Instead, the routine of everyday life is an emergent

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521386039 - Computation and Human Experience
Philip E. Agre
Excerpt
More information

phenomenon of moment-to-moment interactions that work out in much the same way from day to day because of the relative stability of our relationships with our environments.

My sketched alternative also denies a central role to the use of plans. People certainly use plans. But real plans are nothing like computer programs. Sensibly organized goal-directed activity need not result from the use of a plan. And plans never serve as direct specifications of action. Instead, a plan is merely one resource among many that someone might use in deciding what to do (Suchman 1987). Before and beneath any use of plans is a continual process of moment-to-moment improvisation. "Improvisation," as I will employ the term, might involve ideas about the future and it might employ plans, but it is always a matter of deciding what to do *now*. Indeed, the use of plans is a relatively peripheral phenomenon and not a principal focus here.[9]

To speak of a "planning view" is misleading in one respect: few people are aware of having committed themselves to such a view. Future chapters will explain more precisely the sense in which the planning view has governed research in AI. For the time being, it will be helpful to consider Heidegger's (1961 [1927]) account of why the emergence of something like the planning view is nearly inevitable. Most of us, Heidegger observes, spend our days immersed in practical concerns. We are concerned with the traffic, the paperwork, the dust, the celery – with the objects that we encounter as we pursue our goals and enact our identities. We find it natural, therefore, to see the world as a constellation of objects. Moreover, the occasions on which particular objects really come to our attention are not representative of activity as a whole. Sometimes we momentarily detach ourselves from our daily concerns to contemplate an object in a special way – as, for example, a work of art. And sometimes an object simply becomes obstinate; perhaps it is broken, or missing, or not the right size. In these situations, we confront the object as a stranger – as something very much separate from us. It is *problems* that attract our attention, and problems play a wildly disproportionate role in the stories we tell about our lives. We hardly notice the vast background of ordinary, routine, unproblematic activities from which our lives are largely made. Even when a problem does arise, the detection and resolution of the problem both consist of concrete activities that are mostly routine. Because this unproblematic background of routine activity goes largely unnoticed, we can succumb to the illusion that life is basically a series of

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521386039 - Computation and Human Experience
Philip E. Agre
Excerpt
More information

problems, that situations in life typically require thinking and planning, and that our normal way of relating to objects involves detached contemplation. This illusion does not simply arise in individual experience; it is also handed down through metaphors, narrative conventions, philosophical systems, and other cultural constructs. It is this illusory view of life – the planning view – that first crystallized in its modern form in Descartes and that originally defined the tacit agenda for research in AI. Yet, I will argue, the planning view is inadequate both as an account of human life and as an approach to computational modeling.

It is hard to know, of course, how to evaluate Heidegger's argument. Perhaps we should treat it as a just-so story; Heidegger, in any case, presented it as a phenomenological description and a reconstruction of the history of philosophy, not a logical deduction from premises or a scientific inference from evidence. For our purposes here, though, that is enough. Heidegger's story is useful in several ways. It confers an overall sense on the more detailed analyses of Descartes and other theorists of mechanism. It also directs our attention heuristically to technical difficulties that might otherwise have gone undiagnosed or misunderstood. Above all, it helps us cultivate an awareness of our own experience as human beings. Heidegger's crucial insight is that philosophical ideas tend to formalize the ways we experience our lives; if we experience our lives in superficial ways then our philosophies will be correspondingly superficial. The same reasoning applies to computational models, which (whether the model-builders realize it or not) are derived from philosophical theories and guided in their development by experiences of everyday life. Better descriptions of everyday life do not disprove technical ideas, but they do motivate different intuitions, and they also help evaluate the appeals to everyday intuition that are found throughout AI research. AI's pervasive focus on problems, for example, aligns with the unreflective emphasis on problems that Heidegger finds in the modern experience of everyday life. By failing to place problems in the context of an unproblematic background, AI may fall prey to a mistaken conception of them and an excessive emphasis on attempts to solve them. The point in each case is not to debunk AI or technology in general, but to gain what Heidegger would call a "free relation" to it, so that technological modes of thinking do not colonize our awareness of our lives (Dreyfus 1995). Let us turn to the methodological issues that arise as AI research is rethought in this way.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
0521386039 - Computation and Human Experience
Philip E. Agre
Excerpt
More information

Why build things?

Every discipline has its distinctive ways of knowing, which it identifies with the activities it regards as its own: anthropologists do fieldwork, architects design buildings, monks meditate, and carpenters make things out of wood. Each discipline wears its defining activity as a badge of pride in a craftworker's embodied competence. It will be said, "You can read books all your life, but you don't really know about it until you do it." Disciplinary boundaries are often defined in such ways – you are not an anthropologist unless you have spent a couple years in the field; you are not an architect unless you have built a building; and so forth – and neighboring disciplines may be treated with condescension or contempt for their inferior methods. Each discipline's practitioners carry on what Schön (1983: 78) would call "reflective conversations" with their customary materials, and all of their professional interactions with one another presuppose this shared background of sustained practical engagement with a more or less standard set of tools, sites, and hassles. Defining a discipline through its own special activity carries risks. If a disciplinary community cultivates invidious contrasts between its own methods and those of other fields, it will surely become inbred and insular, emptied by hubris and intolerance. If it is guided by critical reflection on its practices and presuppositions, however, it has at least a chance of continually deepening its self-awareness, renewing itself through interdisciplinary dialogue without losing its distinctive advantages. The culture of any particular discipline will presumably be found somewhere between these extremes.

The discipline in question here is computational modeling, and specifically AI. Although I will criticize certain computational ideas and practices at great length, my enterprise is computational nonetheless. AI's distinctive activity is building things, specifically computers and computer programs. Building things, like fieldwork and meditation and design, is a way of knowing that cannot be reduced to the reading and writing of books (Chapman 1991: 216–217). To the contrary, it is an enterprise grounded in a routine daily practice. Sitting in the lab and working on gadgets or circuits or programs, it is an inescapable fact that some things can be built and other things cannot. Likewise, some techniques scale up to large tasks and others do not; and some devices operate robustly as environmental conditions fluctuate, whereas others break down. The AI

© Cambridge University Press www.cambridge.org