Langton, C. G. (1989*a*), 'Artificial Life', in Langton (1989*b*).
—— (1989*b*), *Artificial Life: Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems* (Santa Fe Institute Studies in the Sciences of Complexity, Proceedings, 6; Redwood City, Calif.: Addison-Wesley).
—— Taylor, C., Farmer, J. D., and Rasmussen, S. (1991) (eds.), *Artificial Life II* (Santa Fe Institute Studies in the Sciences of Complexity, Proceedings, 10; Redwood City, Calif.: Addison-Wesley).
Lindgren, K. (1991), 'Evolutionary Phenomena in Simple Dynamics', in Langton *et al.* (1991), 295–312.
Margolus, L. (1970), *Origin of Eucaryotic Cells* (New Haven: Yale University Press).
Prusinkiewicz, P. (1991), *The Algorithmic Beauty of Plants* (Berlin: Springer-Verlag).
Ray, T. S. (1991), 'An Approach to the Synthesis of Life', in Langton *et al.* (1991), 371–408, and Chapter 3 below.
Reynolds, C. W. (1987), 'Flocks, Herds, and Schools: A Distributed Behavioral Model', Proceedings of SIGGRAPH '87, *Computer Graphics* V 21/4: 25–34.
Toffoli, T. (1984), 'Cellular Automata as an Alternative to (Rather than an Approximation of) Differential Equations in Modeling Physics', in J. D. Farmer, T. Toffoli, and S. Wolfram (eds.), *Cellular Automata: Proceedings of an Interdisciplinary Workshop* (Los Alamos, New Mexico, March 7–11, 1983) = *Physica D* (special issue), 10/1–2.
—— and Margolus, N. (1987), *Cellular Automata Machines* (Cambridge, Mass.: MIT Press).
Ulam, S. (1962), 'On Some Mathematical Problems Connected with Patterns of Growth of Figures', *Proceedings of Symposia in Applied Mathematics*, 14: 215–24. Repr. In Burks (1970).
Von Neumann, J. (1966), *Theory of Self-Reproducing Automata*, ed. and completed by A. W. Burks (Urbana, Ill.: University of Illinois Press).
Wilson, S. W. (1989), 'The Genetic Algorithm and Simulated Evolution', in Langton (1989), 157–65.
Wolfram, S. (1986), 'Cellular Automaton Fluids 1: Basic Theory', *Journal of Statistical Physics*, 45: 471–526.

2

AUTONOMY AND ARTIFICIALITY

MARGARET A. BODEN

1. THE PROBLEM—AND WHY IT MATTERS

When Herbert Simon wrote his seminal book on 'the sciences of the artificial' (Simon 1969), he had in mind artificial intelligence (AI) and cybernetics. Now, the sciences of the artificial include artificial life (A-Life) also. A-Life uses informational concepts and computer-modelling to study the functional principles of life in general (Langton 1989). Simon's use of the word 'sciences' was well chosen. The interests of A-Life, as of AI, are largely scientific, not technological. That is, many researchers in these two fields hope to contribute to theoretical biology and/or psychology.

The relations between A-Life and AI are complex. One might define A-Life as the abstract study of life, and AI as the abstract study of mind. But if one assumes that life prefigures mind, that cognition is—and must be—grounded in self-organizing adaptive systems, then AI may be seen as a sub-class of A-Life. Certainly, A-Life is theoretically close to some recent work in AI (described below). However, A-Life workers often go out of their way to distance their research from AI. In doing so, they usually stress the concept of autonomy—which, they say, applies to A-Life models but not to AI.

Since autonomy of some kind is generally thought to be an important characteristic of both life and mind, A-Life and AI should have implications for our understanding of human autonomy, or freedom. These implications are not purely abstract, to be forgotten when one leaves one's study to play a game of backgammon. What science tells us about human autonomy is practically important, because it affects the way in which ordinary people see themselves—which includes the way in which they believe it is possible to behave.

If science tells us that we lack freedom, we may be less likely to try to exercise it. An observable decline of personal autonomy was reported many

 An emended version of a paper called 'Artificial Intelligence and Human Dignity', in J. Cornwell (ed.), *Nature's Imagination: The Frontiers of Scientific Vision* (Oxford: Oxford University Press, 1995), 148–60. Reprinted by permission of Oxford University Press.

years ago by the psychotherapist Rollo May (1961). His experiences in the consulting-room led him to complain of 'the dehumanizing dangers in our tendency in modern science to make man over into the image of the machine, into the image of the techniques by which we study him' and of 'the undermining of [modern man's] experience of himself as responsible, the sapping of his willing and decision' (May 1961: 20). May's mention of 'the machine' referred not to A-Life or AI (which had barely begun), but to the mechanistic implications of the natural sciences in general and behaviourist psychology in particular. Indeed, the relevant implications of behaviourism were soon to be made explicit by B. F. Skinner, in a spirited attack on the concepts of freedom and dignity (Skinner 1971).

'Freedom', Skinner argued, is an illusion, grounded in our ignorance of the multiple environmental pressures determining our behaviour. As for 'dignity', this is a matter of giving people credit, of admiring them for their self-generated achievements. But his behaviourist principles implied that 'the environment', not 'autonomous man', is really in control (Skinner 1971: 21). No credit, then, to *us*, if we exercise some skill—whether bodily, mental, or moral. And no surprise, if people experience unhappiness and apathy at this downgrading of their humanity. The cure for this unhappiness, said Skinner, is to drop our sentimental, unscientific picture of autonomous man. Only then shall we be in a position (with the help of his scientific psychology) to solve our life-problems.

Behaviourism, then, questions received notions of human worth. According to its high priest, Skinner, it has no room for human 'autonomy' or 'freedom'. But it is at least concerned with life. Animals are living things, and *Rattus Norvegicus* a moderately merry mammal. Some small shred of our self-respect can perhaps be retained, if we are classed with rats, or even pigeons. But what of the artificial sciences? Surely, all they can offer is dead, automatic tin-cannery? Is A-Life, for all its stress on self-organization and autonomy, really any better in this respect than AI?

To many people, the notion that computer models could help us to an adequate account of humanity—above all, of freedom, creativity, and morals—seems quite absurd. To the contrary, the suspicion is that the concepts and explanations of A-Life and/or AI must be incompatible with the notion of human freedom. If that suspicion is correct, then the artificial sciences—A-Life included—are the enemy of true autonomy: if they prosper, we can expect to hear more complaints like May's in the future.

In the next section, I explain how A-Life (and some recent AI) addresses the phenomenon of autonomy, and how it differs from traditional AI in this respect. In Section 3, I discuss the concept of autonomy, and claim that crucial aspects of the strongest case (human free action) are not captured by the A-Life approach.

Indeed, these aspects—self-reflection, deliberation, and reasoned prioritizing—are better modelled by what John Haugeland (1985) has called GOFAI: Good Old-Fashioned AI. My conclusion is not that the artificial sciences deny, or even downgrade, our freedom. Rather, they help us to see how autonomous behaviour (of various kinds) is possible, and to appreciate the awesome complexity of much human choice.

2. AI, A-LIFE, AND ANTS

At first sight, it may seem that the explanations offered by any 'artificial' science must be incompatible with human freedom. For it is not only behaviourists who see conditions in the external environment as the real causes of apparently autonomous behaviour. Only a few years after May's complaint quoted above, Simon—in defining the sciences of the artificial—took much the same view (Simon 1969).

Simon described the erratic path of the ant, as it avoided the obstacles on its way to nest or food, as the result of a series of simple and immediate reactions to the local details of the terrain. He did not stop with ants, but tackled humans too. For over twenty years, Simon has argued that rational thought and skilled behaviour are largely triggered by specific environmental cues. Unlike Skinner, he allows that many internal, 'mentalistic', cues are important also; but these produce their effects in just as direct a way as the external conditions do. The extensive psychological experiments and computer-modelling on which Simon's argument is based were concerned with chess, arithmetic, and typing (Newell and Simon 1972; Card *et al.* 1983). But he would say the same of every case of skilled activity or intelligent thought.

Simon's ant was not taken as a model by most of his AI colleagues. Instead, they were inspired by his earliest, and significantly different, work on the computer simulation of problem-solving (Newell and Simon 1961; Newell *et al.* 1963). This ground-breaking theoretical research paid no attention to environmental factors, but conceived of human thought purely in terms of internal mental/computational processes, such as hierarchical means–end planning and goal-representations.

Driven by this 'internalist' view, the young AI community designed—and in some cases built—robots guided top-down by increasingly sophisticated internal planning and representation (Boden 1987, ch. 12). Plans were worked out ahead of time. In the most flexible cases, certain contingencies could be foreseen, and the detailed movements, and even the sub-plans, could be decided on at the time of execution. But even though they were placed in the physical world, these robots were not real-world, real-time creatures. Their environments were

simple, highly predictable 'toy-worlds'. They typically involved a flat ground-plane, polyhedral and/or pre-modelled shapes, white surfaces, shadowless lighting, and—by human standards—painfully slow movements. Moreover, they were easily called to a halt, or trapped into fruitless perseverative behaviour, by unforeseen environmental details.

Recently, however, the AI pendulum has swung towards the ant. What is sometimes called '*nouvelle* AI' sees behaviour as being controlled by an ongoing interaction between relatively low-level mechanisms in the system (robot or organism) and the constantly changing details of the environment.

For example, current research in *situated robotics* sees no need for the symbolic representations and detailed anticipatory planning typical of earlier AI robotics. Indeed, the earlier strategy is seen as not just unnecessary, but ineffective. Traditional robotics suffers from the brittleness of classical AI programs in general: unexpected input can cause the system to do something highly inappropriate, and there is no way in which the problem-environment can help guide it back on to the right track. Accepting that the environment cannot be anticipated in detail, workers in situated robotics have resurrected the insight—often voiced within classical AI, but also often forgotten—that the best source of information about the real world is the real world itself.

Accordingly, the 'intelligence' of these very recent robots is in the hardware, not the software (Braitenberg 1984; Brooks 1991). There is no high-level program doing detailed anticipatory planning. Instead, the creature is engineered in such a way that, within limits, it naturally does the right (adaptive) thing at the right time. Behaviour apparently guided by goals and hierarchical planning can, nevertheless, occur (Maes 1991).

Situated robotics is closely related to two other recent forms of computer-modelling, likewise engaged in studying 'emergent' behaviours. These are genetic algorithms (GAs) and A-Life.

GA systems are self-modifying programs, which continually come up with new rules (new structures) (Holland 1975; Holland *et al.* 1986). They use rule-changing algorithms modelled on genetic processes such as mutation and cross-over, and algorithms for identifying and selecting the relatively successful rules. Mutation makes a change in a single rule; cross-over brings about a mix of two, so that (for instance) the left-hand portion of one rule is combined with the right-hand portion of the other. Together, these algorithms (working in parallel) generate a new system better adapted to the task in hand.

One example of a GA system is a computer-graphics program written by Karl Sims (1991). This program uses genetic algorithms to generate new images, or patterns, from pre-existing images. Unlike most GA systems, the selection of the 'fittest' examples is not automatic, but is done by the programmer—or by someone fortunate enough to be visiting his office while the program is

being run. That is, the human being selects the images which are aesthetically pleasing, or otherwise interesting, and these are used to 'breed' the next generation. (Sims could provide automatic selection rules, but has not yet done so—not only because of the difficulty of defining aesthetic criteria, but also because he aims to provide an *interactive* graphics environment, in which human and computer can cooperate in generating otherwise unimaginable images.)

In a typical run of the program, the first image is generated at random (but Sims can feed in a real image, such as a picture of a face, if he wishes). Then the program makes nineteen independent changes (mutations) in the initial image-generating rule, so as to cover the VDU-screen with twenty images: the first, plus its nineteen ('asexually' reproduced) offspring. At this point, the human uses the computer mouse to choose either *one* image to be mutated, or *two* images to be 'mated' (through cross-over). The result is another screenful of twenty images, of which all but one (or two) are newly generated by random mutations or cross-overs. The process is then repeated, for as many generations as one wants.

(The details of this GA system need not concern us. However, so as to distinguish it from magic, a few remarks may be helpful. It starts with a list of twenty very simple LISP functions. A 'function' is not an actual instruction, but an instruction-schema: more like 'X + Y' than '2 + 3'. Some of these functions can alter parameters in pre-existing functions: for example, they can divide or multiply numbers, transform vectors, or define the sines or cosines of angles. Some can combine two pre-existing functions, or nest one function inside another (so multiply-nested hierarchies can eventually result). A few are basic image-generating functions, capable (for example) of generating an image consisting of vertical stripes. Others can process a pre-existing image, for instance by altering the light-contrasts so as to make 'lines' or 'surface-edges' more or less visible. When the program chooses a function at random, it also randomly chooses any missing parts. So if it decides to *add* something to an existing number (such as a numerical parameter inside an image-generating function), and the 'something' has not been specified, it randomly chooses the amount to be added. Similarly, if it decides to *combine* the pre-existing function with some other function, it may choose that function at random.)

As for A-Life, this uses computer-modelling to study processes that start with relatively simple, locally interacting units, and generate complex individual and/or group behaviours. Examples of such behaviours include self-organization, reproduction, adaptation, purposiveness, and evolution.

Self-organization is shown, for instance, in the flocking behaviour of flocks of birds, herds of cattle, and schools of fish. The entire group of animals seems to behave as one unit. It maintains its coherence despite changes in direction, the (temporary) separation of stragglers, and the occurrence of obstacles—which

the flock either avoids or 'flows around'. Yet there is no overall director working out the plan, no sergeant-major yelling instructions to all the individual animals, and no reason to think that any one animal is aware of the group as a whole. The question arises, then, how this sort of behaviour is possible.

Ethologists argue that communal behaviour of large groups of animals must depend on local communications between neighbouring individuals, who have no conception of the group-behaviour as such. But just what are these 'local communications'?

Flocking has been modelled within A-Life, in terms of a collection of very simple units, called Boids (Reynolds 1987). Each Boid follows three rules: (1) keep a minimum distance from other objects, including other Boids; (2) match velocity to the average velocity of the Boids in the immediate neighbourhood; and (3) move towards the perceived centre of mass of the Boids in the neighbourhood. These rules, depending as they do only on very limited, local, information, result in the holistic flocking behaviour just described. It does not follow, of course, that real birds follow just those rules: that must be tested by ethological studies. But this research shows that it is at least *possible* for group-behaviour of this kind to depend on very simple, strictly local, rules.

Situated robotics, GAs, and A-Life can be combined, for they share an emphasis on bottom-up, self-adaptive, parallel processing. At present, most situated robots are hand-crafted. But some are 'designed' by evolutionary algorithms from the GA/A-Life stable: fully simulated robots have already been evolved, and real robots are now being constructed with the help of simulated evolution. The automatic evolution of real physical robots *without any recourse to simulation* is more difficult (Brooks 1992), but progress is being made in this area too.

Recent work in evolutionary robotics (Cliff *et al.* 1993) has simulated insect-like robots, with simple 'brains' controlling their behaviour. The (simulated) neural net controlling the (simulated) visuomotor system of the robot gradually adapts to its specific (simulated) task-environment. This automatic adaptation can result in some surprises. For instance, if—in the given task-environment—the creature does not actually need its (simulated) in-built whiskers as well as its eyes, the initial network-links to the whiskers may eventually be lost, and the relevant neural units may be taken over by the eyes. *Eyes* can even give way to *eye*: if the task is so simple that only one eye is needed, one of them may eventually lose its links with the creature's network-brain.

Actual (physical) robots of this type can be generated by combining simulated evolution with hardware-construction (Cliff *et al.* 1993). The detailed physical connections to, and within, the 'brain' of the robot-hardware are adjusted every n generations (where n may be 100, or 1,000, or . . .), mirroring the current blueprint evolved within the simulation. This acts as a cross-check: the

real robot should behave as the simulated robot does. Moreover, the resulting embodied robot can roam around an actual physical environment, its real-world task-failures and successes being fed into the background simulation so as to influence its future evolution. The brain is not the only organ whose anatomy can be evolved in this way: the placement and visual angle of the creatures' eyes can be optimized, too. (The same research-team has begun work on the evolution of physical robots without any simulation. This takes much longer, because every single evaluation of every individual in the population has to be done using the real hardware.)

A-Life (some of which uses GAs) and situated robotics have strong links with biology: with neuroscience, ethology, genetics, and the theory of evolution. As a result, animals are becoming theoretically assimilated to *animats* (Meyer and Wilson 1991). The behaviour of swarms of bees, and of ant-colonies, is hotly discussed at A-Life conferences, and entomologists are constantly cited in the A-Life and situated-robotics literatures (Lestel 1992). Environmentally situated (and formally defined) accounts of apparently goal-seeking behaviour in various animals, including birds and mammals, are given by (some) ethologists (McFarland 1989). And details of invertebrate psychology, such as visual tracking in the hoverfly, are modelled by research in connectionist AI (Cliff 1990; 1992).

In short, Simon's ant is now sharing the limelight on the AI stage. Some current AI is more concerned with artificial insects than with artificial human minds. But—what is of particular interest to us here—this form of AI sees itself as designing 'autonomous agents', as A-Life in general seeks to design 'autonomous systems'.

3. AUTONOMOUS AGENCY

Autonomy is ascribed to these artificial insects because it is their intrinsic physical structure, adapted as it is to the sorts of environmental problem they are likely to meet, which enables them to act appropriately. Unlike traditional robots, their behaviour is not directed by complex software written for a general-purpose machine, imposed on their bodies by some alien (human) hand. Rather, they are specifically constructed to adapt to the particular environment they inhabit.

We are faced, then, with two opposing intuitions concerning autonomy. Our (and Skinner's) original intuition was that response determined by the external environment lessens one's autonomy. But the intuition of *nouvelle* AI, and of A-Life, is that to be in thrall to an internal plan is to be a mere puppet. (Notice that one can no longer say 'a mere robot'.) How can these contrasting intuitions be reconciled?

Autonomy is not an all-or-nothing property. It has several dimensions, and many gradations. Three aspects of behaviour—or rather, of its control—are crucial. First, the extent to which response to the environment is direct (determined only by the present state in the external world) or indirect (mediated by inner mechanisms partly dependent on the creature's previous history). Second, the extent to which the controlling mechanisms were self-generated rather than externally imposed. And third, the extent to which inner directing mechanisms can be reflected upon, and/or selectively modified in the light of general interests or the particularities of the current problem in its environmental context. An individual's autonomy is the greater, the more its behaviour is directed by self-generated (and idiosyncratic) inner mechanisms, nicely responsive to the specific problem-situation, yet reflexively modifiable by wider concerns.

The first aspect of autonomy involves behaviour mediated, in part, by inner mechanisms shaped by the creature's past experience. These mechanisms may, but need not, include explicit representations of current or future states. It is controversial, in ethology as in philosophy, whether animals have explicit internal representations of goals (Montefiore and Noble 1989). And, as we have seen, AI includes strong research-programmes on both sides of this methodological fence. But this controversy is irrelevant here. The important distinction is between a response wholly dependent on the current environmental state (given the original, 'innate', bodily mechanisms), and one largely influenced by the creature's experience. The more a creature's past experience differs from that of other creatures, the more 'individual' its behaviour will appear.

The second aspect of autonomy, the extent to which the controlling mechanisms were self-generated rather than externally imposed, may seem to be the same as the first. After all, a mechanism shaped by experience is sensitive to the past of that particular individual—which may be very different from that of other, initially comparable, individuals. But the distinction, here, is between behaviour which 'emerges' as a result of self-organizing processes, and behaviour which was deliberately prefigured in the design of the experiencing creature.

In computer-simulation studies within A-Life, and within situated robotics also, holistic behaviour—often of an unexpected sort—may emerge. It results, of course, from the initial list of simple rules concerning locally interacting units. But it was neither specifically mentioned in those rules, nor (often) foreseen when they were written.

A flock, for example, is a holistic phenomenon. A bird-watcher sees a flock of birds as a unit, in the sense that it shows behaviour that can be described only at the level of the flock itself. For instance, when it comes to an obstacle, such as a tall building, the flock divides and 'flows' smoothly around it, reorganizing itself into a single unit on the far side. But no individual bird is

divided in half by the building. And no bird has any notion of the flock as a whole, still less any goal of reconstituting it after its division.

Clearly, flocking behaviour must be described on its own level, even though it can be explained by (reduced to) processes on a lower level. This point is especially important if 'emergence-hierarchies' evolve as a result of new forms of perception, capable of detecting the emergent phenomena *as such*. Once a holistic behaviour has emerged it, or its effects, may be detected (perceived) by some creature or other—including, sometimes, the 'unit-creatures' making it up.

(This implies that a creature's perceptual capacities cannot be fully itemized for all time. In Gibsonian terms, one might say that evolution does not know what all the relevant affordances will turn out to be, so cannot know how they will be detected. The current methodology of AI and A-Life does not allow for 'latent' perceptual powers, actualized only by newly emerged environmental features. This is one of the ways in which today's computer-modelling is biologically unrealistic (Kugler 1992).)

If the emergent phenomenon can be detected, it can feature in rules governing the perceiver's behaviour. Holistic phenomena on a higher level may then result . . . and so on. Ethologists, A-Life workers, and situated roboticists all assume that increasingly complex hierarchical behaviour can arise in this sort of way. The more levels in the hierarchy, the less direct the influence of environmental stimuli—and the greater the behavioural autonomy.

Even if we can *explain* a case of emergence, however, we cannot necessarily *understand* it. One might speak of intelligible vs. unintelligible emergence.

Flocking gives us an example of the former. Once we know the three rules governing the behaviour of each individual Boid, we can see lucidly how it is that holistic flocking results.

Sims's computer-generated images give us an example of the latter. One may not be able to say just why *this* image resulted from *that* LISP expression. Sims himself cannot always explain the changes he sees appearing on the screen before him, even though he can access the mini-program responsible for any image he cares to investigate, and for its parent(s) too. Often, he cannot even 'genetically engineer' the underlying LISP expression so as to get a particular visual effect. To be sure, this is partly because his system makes several changes simultaneously, with every new generation. If he were to restrict it to making only one change, and studied the results systematically, he could work out just what was happening. But when several changes are made in parallel, it is often impossible to understand the generation of the image *even though* the 'explanation' is available.

Where real creatures are concerned, of course, we have multiple interacting changes, and no explanation at our fingertips. At the genetic level, these multiple

changes and simultaneous influences arise from mutations and cross-over. At the psychological level, they arise from the plethora of ideas within the mind. Think of the many different thoughts which arise in your consciousness, more or less fleetingly, when you face a difficult choice or moral dilemma. Consider the likelihood that many more conceptual associations are being activated unconsciously in your memory, influencing your conscious musings accordingly. Even if we had a listing of all these 'explanatory' influences, we might be in much the same position as Sims, staring in wonder at one of his nth-generation images and unable to say why *this* LISP expression gave rise to it. In fact, we cannot hope to know about more than a fraction of the ideas aroused in human minds (one's own, or someone else's) when such choices are faced.

The third criterion of autonomy listed above was the extent to which a system's inner directing mechanisms can be reflected upon, and/or selectively modified, by the individual concerned. One way in which a system can adapt its own processes, selecting the most fruitful modifications, is to use an 'evolutionary' strategy such as the genetic algorithms mentioned above. It may be that something broadly similar goes on in human minds. But the mutations and selections carried out by GAs are modelled on biological evolution, not conscious reflection and self-modification. And it is conscious deliberation which many people assume to be the crux of human autonomy.

For the sake of argument, let us accept this assumption at face value. Let us ignore the mounting evidence (e.g. Nisbett and Ross 1980) that our conscious thoughts are less relevant than we like to think. Let us ignore neuroscientists' doubts about whether our conscious intentions actually direct our behaviour (as the folk-psychology of 'action' assumes) (Libet 1987). Let us even ignore the fact that *unthinking spontaneity*—the opposite of conscious reflection—is often taken as a sign of individual freedom. (Spontaneity may be based in the sort of multiple constraint satisfaction modelled by connectionist AI, where many of the constraints are drawn from the person's idiosyncratic experience.) What do the sciences of the artificial, and AI-influenced psychology, have to say about conscious thinking and deliberate self-control?

Surprisingly, perhaps, neither A-Life nor the most biologically realistic (more accurately: the least biologically unrealistic) forms of AI can help us here. Ants, and artificial ants, are irrelevant. Nor can connectionism help. It is widely agreed, even by connectionists, that conscious thought requires a sequential 'virtual machine', more like a von Neumann computer than a parallel-processing neural net. As yet, we have only very sketchy ideas about how the types of problem-solving best suited to conscious deliberation might be implemented in connectionist systems.

The most helpful 'artificial' approach so far, where conscious deliberation is involved, is classical AI, or GOFAI—much of which was inspired by human introspection. Consciousness involves reflection on one level of processes going on at a lower level. Work in classical AI, such as the work on planning mentioned above, has studied multi-level problem-solving. Computationally informed work in developmental psychology has suggested that flexible self-control, and eventually consciousness, result from a series of 'representational redescriptions' of lower-level skills (Clark and Karmiloff-Smith 1993).

Representational redescriptions, many-levelled maps of the mind, are crucial to creativity (Boden 1990, esp. ch. 4). Creativity is an aspect of human autonomy, and workers whose working-conditions allow them no room for creative ingenuity (or even for choice) may describe themselves as 'monkeys', 'robots', 'machines', or even 'objects' (Terkel 1974, p. xi). Their discontent is understandable, given that our ability to think new thoughts in new ways is one of our most salient, and most valued, characteristics.

This ability often involves someone's doing something which they not only *did not* do before, but which they *could not* have done before. To do this, they must either explore a formerly unrecognized area of some pre-existing 'conceptual space', or transform some dimension of that generative space. Transforming the space allows novel mental structures to arise which simply could not have been generated from the initial set of constraints. The nature of the creative novelties depends on which feature has been transformed, and how. Conceptual spaces, and procedures for transforming them, can be clarified by thinking of them in computational terms. But this does not mean that creativity is predictable, or even fully explicable *post hoc*: for various reasons (including those mentioned above), it is neither (Boden 1990, ch. 9).

Autonomy in general is commonly associated with unpredictability. Many people feel AI to be a threat to their self-esteem because they assume that it involves a deterministic predictability. But they are mistaken. Some connectionist AI systems include non-deterministic (stochastic) processes, and are more efficient as a result.

Moreover, determinism does not always imply predictability. Workers in A-Life, for instance, sometimes justify their use of computer-simulation by citing chaos theory, according to which a fully deterministic dynamic process may be theoretically unpredictable (Langton 1989). If there is no analytic solution to the differential equations describing the changes concerned, the process must simply be 'run', and observed, to know what its implications are. The same is true of many human choices. We cannot always predict what a person will do. Moreover, predicting *one's own* choices is not always possible. One may have to 'run one's own equations' to find out what one will do, since the outcome cannot be known until the choice is actually made.

4. CONCLUSION

One of the pioneers of A-Life has said:

> The field of Artificial Life is unabashedly mechanistic and reductionist. However, this *new mechanism*—based as it is on multiplicities of machines and on recent results in the fields of nonlinear dynamics, chaos theory, and the formal theory of computation—is vastly different from the mechanism of the last century. (Langton 1989: 6; italics in original)

Our discussion of A-Life and *nouvelle* AI has suggested just how vast this difference is. Similarly, the potentialities of classical AI systems go far beyond what most people think of as 'machines'. If this is reductionism, it is very different from the sort of reductionism which insists that the only scientifically respectable concepts lie at the most basic ontological level (neurones and biochemical processes, or even electrons, mesons, and quarks).

In sum, the sciences of the artificial can model autonomy of various kinds. A-Life (like *nouvelle* AI) explicitly highlights autonomy, as a characteristic of living things. A-Life can teach us a great deal about how increasing complexity arises from self-organization on successive levels, and how a creature can negotiate its environment by constant interaction with it. However, the kind of autonomy that we call free choice is better illuminated by the theoretical approach of classical AI. If this is generally accepted the result need not be insidiously dehumanizing, as the acceptance of behaviourism sometimes was. Properly understood, AI does not reduce our respect for human minds. If anything, it increases it. Far from denying human autonomy, it helps us to understand how it is possible.

REFERENCES

Boden, M. A. (1987), *Artificial Intelligence and Natural Man* (2nd edn.; London: MIT Press).

—— (1990), *The Creative Mind: Myths and Mechanisms* (London: Weidenfeld and Nicolson).

Braitenberg, V. (1984), *Vehicles: Essays in Synthetic Psychology* (Cambridge, Mass.: MIT Press).

Brooks, R. A. (1991), 'Intelligence Without Representation', *Artificial Intelligence*, 47: 139–59.

—— (1992), 'Artificial Life and Real Robots', in F. J. Varela and P. Bourgine (eds.), *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life* (Cambridge, Mass.: MIT Press), 3–10.

Card, S. K., Moran, T. P., and Newell, A. (1983), *The Psychology of Human–Computer Interaction* (Hillsdale, NJ: Erlbaum).

Clark, A., and Karmiloff-Smith, A. (1993), 'The Cognizer's Innards: A Psychological and Philosophical Perspective on the Development of Thought', *Mind and Language*, 8: 487–568.

Cliff, D. (1990), 'The Computational Hoverfly: A Study in Computational Neuroethology', in J.-A. Meyer and S. W. Wilson (eds.), *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior* (Cambridge, Mass.: MIT Press), 87–96.

—— (1992), 'Neural Networks for Visual Tracking in an Artificial Fly', in F. J. Varela and P. Bourgine (eds.), *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life* (Cambridge, Mass.: MIT Press), 78–87.

—— Harvey, I., and Husbands, P. (1993), 'Explorations in Evolutionary Robotics', *Adaptive Behavior*, 2/1: 73–110.

Haugeland, J. (1985), *Artificial Intelligence: The Very Idea* (Cambridge, Mass.: MIT Press).

Holland, J. H. (1975), *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence* (Ann Arbor: University of Michigan Press; reissued MIT Press, 1991).

—— Holyoak, K. J., Nisbet, R. E., and Thagard, P. R. (1986), *Induction: Processes of Inference, Learning, and Discovery* (Cambridge, Mass.: MIT Press).

Kugler, P. (1992), Talk given at the Summer School on 'Comparative Approaches to Cognitive Science', Aix-en-Provence (organizers, J.-A. Meyer and H. L. Roitblat).

Langton, C. G. (1989), 'Artificial Life', in C. G. Langton (ed.), *Artificial Life: Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems* (Santa Fe Institute Studies in the Sciences of Complexity, Proceedings, 6; Redwood City, Calif.: Addison-Wesley), 1–47.

Lestel, D. (1992), 'Fourmis cybernetiques et robots-insectes: Socialité et cognition à l'interface de la robotique et de l'éthologie experimentale', *Information sur les Sciences Sociales*, 31/2: 179–211.

Libet, B. (1987), 'Are the Mental Experiences of Will and Self-Control Significant for the Performance of a Voluntary Act?', *Behavioral and Brain Sciences*, 10: 783–6.

McFarland, D. (1989), 'Goals, No-Goals, and Own-Goals', in Montefiore and Noble (1989), 39–57.

Maes, P. (1991) (ed.), *Designing Autonomous Agents* (Cambridge, Mass.: MIT Press).

May, R. (1961), *Existential Psychology* (New York: Random House).

Meyer, J.-A., and Wilson, S. W. (1991) (eds.), *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior* (Cambridge, Mass.: MIT Press).

Montefiore, A., and Noble, D. (1989) (eds.), *Goals, No-Goals, and Own-Goals* (London: Unwin Hyman).

Newell, A., and Simon, H. A. (1961), 'GPS—A Program that Simulates Human Thought', in H. Billing (ed.), *Lernende Automaten* (Munich: Oldenbourg), 109–24. Repr. in E. A. Feigenbaum and J. Feldman (eds.), *Computers and Thought* (New York: McGraw-Hill, 1963), 279–96.

—— —— (1972), *Human Problem Solving* (Englewood Cliffs, NJ: Prentice-Hall).

—— Shaw, J. C., and Simon, H. A. (1963), 'Empirical Explorations with the Logic Theory Machine: A Case-Study in Heuristics', in E. A. Feigenbaum and J. Feldman (eds.), *Computers and Thought* (New York: McGraw-Hill), 109–33.

Nisbett, R. E., and Ross, L. (1980), *Human Inference: Strategies and Shortcomings in Social Judgment* (Englewood Cliffs, NJ: Prentice-Hall).

Reynolds, C. W. (1987), 'Flocks, Herds, and Schools: A Distributed Behavioral Model', *Computer Graphics*, 21/4: 25–34.