STILLNESS AS AUTONOMY*

Margaret A. Boden; University of Sussex

Abstract
Stillness needn't mean stasis. But it could mean homeostasis (i.e. constancy in the face of environmental changes) – which is an example of autonomy. This is a key notion in certain forms of computer art. However, there are two very different forms of autonomy: physical and mental/intentional. Artists strongly influenced by A-Life pay more attention to the first. Some interactive artists prioritise the second. Evolutionary artworks have a greater degree of autonomy than other computer artworks. But the autonomy (freedom) of the artist is still key.

Keywords
Autonomy, computer art, interactive art, evolutionary art, artificial life

* This paper forms part of the research supported by AHRC Grant no. B/RG/AN8285/APN19307: "Computational Intelligence, Creativity, and Cognition: A Multidisciplinary Investigation".

I: Stillness, Stasis, and Homeostasis

Stillness can mean stasis. But it doesn't have to. It could, for instance, mean homeostasis. This is an example of autonomy – which, in turn, is a key notion in certain forms of computer art. Considered as autonomy, then, stillness is at the core of these artworks.

However, we'll see that there are two very different types of autonomy: one physical, and one mental. Artists strongly influenced by A-Life pay more attention to the first. Some interactive artists prioritize the second. The kind of aesthetic "stillness" that's achieved differs accordingly.

Homeostasis is the ability of living organisms to keep certain aspects of their physical state constant, irrespective (up to a point) of what's going on in the environment. One familiar example is the maintenance of blood-temperature in birds and mammals. It's not that the temperature of the blood never changes: it does. But incipient changes are soon compensated by other bodily changes (the diameter of blood vessels, secretions from the pituitary gland, etc.).

The end-result is the maintenance – the 'stillness' – of the blood-temperature. Homeostasis is a form of autonomy: the system does its own thing, achieving constancy despite varying environmental factors. The organism itself controls the various bodily perturbations that occur, directing them towards the final state which it wanted all along – 98.4° Fahrenheit, perhaps.

Birds and non-human mammals, of course, don't actually *want* a normal blood-temperature. Even human beings very rarely *want* that. Occasionally, however, they do – and then they take active, deliberate, steps to achieve it. As we'll see, this reiterates the point remarked above, namely that there are two very different kinds of autonomy.

II: Two Kinds of Autonomy

Autonomy is a complicated concept (Boden 1996). For there are three aspects of an organism's – or, more generally, a system's – behavioural control which are crucial to its independence. Moreover, these don't necessarily run alongside each other, nor keep pace even when they do.

The first is the extent to which response to the environment is direct (determined only by the present state of the external world) or indirect (mediated by inner mechanisms that depend on the system's previous history). The second is the extent to which the controlling mechanisms were self-generated rather than externally imposed. And the third is the extent to which any inner directing mechanisms can be reflected upon, and/or selectively modified in light of general interests and/or the current problem and context.

Clearly, then, autonomy isn't an all-or-nothing property. 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.

Even within A-Life, there are various types, and varying degrees, of autonomy. For example, the senses in which autopoietic systems or self-organizing networks are autonomous differ from each other, and from the sense in which situated robots are autonomous. (See Zeleny 1977, Linsker 1988, and Brooks 1991, respectively.)

For our purposes, however, the most important difference is that between autonomy *as ascribed to non-human systems* and autonomy *as ascribed to adult human beings* (though not to babies or infants). The latter has a name of its own: freedom.

Human freedom is commonly regarded as the epitome of autonomy. A-Lifers, who concern themselves with organisms well below *Homo sapiens* in the phylogenetic scale, rarely mention it. Occasionally, they admit that their work doesn't cover it (e.g. Bird et al. forthcoming: 2.1). But sometimes, their words imply that they confuse it with autonomy as such. That's a mistake. A-Life's examples of autonomy show varying degrees of independence from outside control. But none has the cognitive/motivational complexity required for freedom (remember the third aspect of autonomy listed above).

Traditional, or "top-down", methods of artificial intelligence – which are often disparagingly referred to as GOFAI: "Good Old-Fashioned AI" – have got closer to an understanding of freedom than A-Life has done. Freedom is best understood in terms of a particular form of complex computational architecture (Dennett 1984; Boden 2006: 7.i.g-i, 12.ix.b). It requires a range of psychological resources, which combine to generate decisions/actions selected from a rich space of possibilities. These resources include reasoning, means-end planning, motivation, various sorts of prioritising (including individual preferences and moral principles), analogy-recognition, the anticipation of unwanted side-effects, and deliberate self-monitoring. (It's because these capacities aren't yet developed in infants that they aren't truly free.)

In the paradigm case, the choice is largely conscious. But an action may be termed "free" because, given the computational resources possessed by the person in question, it *could* have been consciously considered by them, and the decision could have differed accordingly. (This usage is sometimes relevant for understanding 'autonomous' computer art, too.)

Freedom can be compromised by everyday threats and bribery, or – more deeply – by relatively exotic circumstances. Under hypnosis, for instance; or when someone obeys hallucinated instructions; or when brain-damage undermines action, so that the person repeatedly forgets what goal/sub-goal they were following, or performs plan-steps in the wrong order (e.g. licking the letter instead of the stamp). Those "exotic" phenomena have been theorized and/or modelled in partly GOFAI terms (Boden 2006: 7.i.h-i; 12.ix.b). In short, the third aspect of behavioural control isn't modelled in A-Life but does feature in (some) GOFAI.

Another way of putting all this is to say that the autonomy of non-human organisms is purely physical, whereas the autonomy of adult human beings is both physical (blood-temperature, again) and mental (or, philosophers would say, intentional). Physical stillness, and mental stillness – where, in each case, the "stillness" isn't stasis but involves continual effort and change. (The goal-forgetting example is an especially clear case of the *lack* of constancy, or stillness, in the afflicted person's behaviour and thinking.)

In the context of computer art, this analysis raises the question of which type/types of autonomy are being attempted (or perhaps even achieved?) by the artists in question.

III: Autonomy and Computer Art

It follows from the discussion above that the "autonomy" one can ascribe to the computer artist and to the computer artwork are fundamentally different. Both involve various types of independence, and both can result in some form of stillness, or constancy. But there are important distinctions to be drawn.

Consider interactive art, for instance. In general, the computer systems involved here are reactive, or responsive. They do what they do, and change from moment to moment, as a result of environmental input. They are thus "autonomous" in the same sense that A-Life's situated robots are autonomous.

In practice, what happens is a function – usually, a direct function – of the viewer's behaviour. (The term "viewer", here, is a shorthand: it could also be a listener, or a viewer-and-listener, or even a sniffer or a feeler...). The relevant input may be the movements of the viewer's body, even including the movements of the eyes, or perhaps the sounds emitted/caused by the viewer. Or it may be something more abstract, such as the density of viewers located in a particular area of the gallery floor, or the rise in temperature due to the accumulation of body-heat in the room. It may even depend to some degree on past events: if the artwork has a memory of past interactions, the cue-response rules may vary accordingly.

In principle, non-human input could be relevant too: the temperature outside the room, or shadows in the sunshine, or the density of autumn leaves falling in the wind. Indeed, some artists have produced video-installations that are partly driven by impersonal changes such as these, as well as by the actions of the viewer. However, this approach lessens the viewer's sense of engagement with the artwork concerned. And that sense of engagement is crucial to truly *interactive* art. The characteristic feature of interactive art – as opposed to traditional paintings or sculptures – is that the viewer doesn't merely engage with it intellectually, or even emotionally: they *change* it, too.

(That's true also of traditional music or drama, if we consider only the audience – although there's a form of audience-interactivity in John Cage's pregnant 'silences', where individual listeners interpret/fill the silence in their own way. In these genres of art, however, performers/interpreters are involved too. 'One and the same' concerto or sonata may sound very different if played by different musicians, and/or under different conductors. Indeed, Baroque composers typically omitted the decorations, leaving those to be freely decided/played by the performer. And, notoriously, different directors/actors draw very different things out of *Hamlet*. So a 'static' work by a 'dictatorial' artist, if music or drama is in question, enables 'artwork-altering' interactions on the performer's part, if not on that of the audience.)

Whether the viewers of interactive art can decide *how* the work will change is another matter. If the viewer is able to follow an agenda, producing a desired end-result in the artwork in question, their sense of engagement (control) will be much greater than if they can cause only unpredictable, apparently random, changes in the piece. Some people, however, may prefer to luxuriate in the experience of caused-but-uncontrolled unpredictability.

It is the computer artist who freely (sic) decides just how unpredictable the artwork's responses will be. They also decide just what the environmental cue/cues will be, and just how the system will respond. Whether the viewer is able, with experience, to discover what the particular artwork's cue-response rules are depends on whether the computer artist wanted that to be so. One can give an aesthetic rationale for wanting viewers to have this capacity, which could give them some voluntary control over what happens to the artwork; and one can equally justify not doing this, so that the changing artwork remains unpredictable (Boden in press). In the first case, the autonomy (independence) of the artwork is lessened, and the autonomy (freedom) of the viewer is increased.

What about the autonomy of the computer artist? They exercise their freedom in deciding what sort of artwork to produce, and why. Even a relatively 'formless', or unpredictable, artwork is a result of the artist's free choice – which isn't to say that they will have foreseen all that the artwork will do. But they may want to engender certain fairly specific types of experience in the viewer. In other words, the artist may be aiming for a particular aesthetic effect to arise constantly (sic) as a result of the interaction. Since the viewer's behaviour too is autonomous (in the strong sense), that constancy can't be guaranteed – unless the system is very highly constrained, not to say boring. But, with skill and understanding (of the human audience as well as of the computer artwork itself) on the artist's part, it can be approximated. In such cases, the artist's autonomy is over-riding the autonomy both of the artwork and of the viewer.

An increasingly popular genre is evolutionary art, inspired by A-Life and the self-organizing powers of biological evolution (Whitelaw 2004). This employs genetic algorithms (GAs) to make random alterations in the images/sounds produced and (often) to select the 'best' at each generation, as in natural selection in biology. Evolutionary artworks have a stronger claim to autonomy than the more traditional computer artworks do, because the system's detailed performance (arising from its evolutionary history) is less dependent on human choice.

There are different kinds of evolutionary art, however, wherein the influence of the viewer varies. As a result, the degree of autonomy that can be ascribed to the artwork, and/or the program, varies too.

On the one hand, an evolutionary artwork may run in complete independence of the viewer. In such cases, *both* the nature of the images/sounds produced *and* the selection at each generation are wholly automatic. As in the non-evolutionary cases, the autonomy of the human programmer is crucial – even though the results are less predictable (because random mutations are involved). But the autonomy of the viewer is irrelevant, since it can't affect what the program does.

On the other hand, the evolutionary artwork may be interactive – in one or both of two senses. First, in a work of art (e.g. a video-installation) that responds to the movements of the viewer, the cue-response rules themselves may evolve as the engagement proceeds. So one and the same movement has different effects in different generations. Second, the selection of images/sounds at each generation may be done by the viewer. This gives the viewer a better chance of (freely) moulding the artwork to his/her own preferences.

In principle, there could also be a third type of interaction, wherein the viewer can choose the type/s of mutation to be used. For instance, suppose that the artwork allows not only for superficial parameter changes (such as altering numerals in an image-generating procedure), but also for the hierarchical nesting of an entire image-generating procedure into another and/or for the concatenation of such procedures (cf. Sims 1991). The viewer might be able to 'switch off' the nesting and/or the concatenation at will (sic) – in which case, the newly-evolved images will be much less varied, much less surprising. The viewer's reason for limiting the degrees of freedom possessed by the artwork in this way might be that they find the images of generation *x* aesthetically interesting, and want the program to explore only images of that general type rather than jumping into a very different (e.g. hierarchically more complex) space of possibilities. (That choice was made by the sculptor William Latham, to produce otherwise-unimaginable images within a clearly recognizable aesthetic style: Todd and Latham 1992.)

Even if the selection at each generation is done automatically (not interactively), the artist's freedom constrains the autonomy of the evolutionary artwork itself. For it is the artist who chooses the fitness function, or functions. These are the ways in which the evolutionary program measures the success of the phenotype in the Darwinian sense. (This choice is usually made 'once and for all', but it could vary during the development of the program.)

Indeed, many people see this point as crucial. Any GA-system's performance, they say, is implicit in (wholly determined by) its fitness functions and mutation rules. Ultimately, then, it's the human who is the creative force. Human autonomy (on this view) trumps computer autonomy, every time.

IV: Robot Artists?

This tension – between the autonomous variation of the system and the artist's free choice of fitness functions – is being explored in a project on creativity and evolutionary robotics at the University of Sussex (Bird et al. forthcoming). The robots in question have a retractable pen, with which they can make line-drawings as they move. These aren't representational drawings (of houses, or men ...), but merely abstract marks – which sometimes have aesthetic value.

Two key questions for the project are: (1) whether the robots can evolve to draw aesthetically acceptable lines that *do not* reflect the personal signature of the computer artist involved (Paul Brown), and (2) whether the relevant fitness functions can specify only very basic properties of the lines, such as (for instance) whether an existing line is or is not about to be crossed. If so, then some degree of creativity would seem to be attributable to the autonomous robots themselves. By contrast, if the fitness functions have to include relatively 'weighty' feature-specifications, then not only will the drawings display Brown's signature, but *his* creativity (and that of his roboticist colleagues) will be entirely responsible for them.

Suppose that the answer to (1) turns out to be *Yes*. In that case, the artist's autonomy will be largely hidden. An artist's signature is a form of constancy, or stillness, in their oeuvre. If it doesn't appear in the drawings done by an evolved robot, then the autonomy (weak sense) of the robot will – to that extent – have avoided, or even overcome, the autonomy (strong sense) of the human being. Indeed, the reason for using actual robots rather than artificially 'clean' simulations is that unpredictable noise in the real world – such as friction on the wheels, or obstacles on the floor – may affect the drawings that are done.

Of course, human autonomy is still crucial. After all, it was Brown himself who decided that the project should aim at losing his signature. Moreover, humans designed the robots in the first place. And, not least, it was they who evaluated the acceptability of the drawings at each 'generation' (more accurately: at the end of each experimental batch of 600 uninterrupted generations) and – if necessary – adjusted the nature of the fitness functions accordingly.

For example, an early fitness function produced robots that made marks within a very small region and then moved back and forwards over them repeatedly. To prevent this, the team added a component to the fitness function that rewarded the robots for making marks spread across the whole arena – which resulted in robots which, in effect, did wall-following. As the next step, a further component was added which rewarded marks spread over a large proportion of the arena. The overall result was that the robots would turn away from walls at an angle, and mark the centre of the arena as well as its edges.

Another key question is (3) whether unpredictable input from the external world (e.g. friction on the wheels, or encounters with obstacles on the ground) will engender line-drawings of a kind not expected by Brown, *and not wholly implicit in the mutation rules and fitness functions*. Other work by a team-member has shown that radically unpredictable results, such as the formation of novel types of sensor, sometimes occur due to fortuitous interactions with seemingly irrelevant aspects of the physical environment (Bird and Layzell 2002). Some philosophers have argued that this is the only way in which fundamental novelties, as opposed to gradual improvements, could have developed in biological evolution – and that *purely programmed* A-Life is limited accordingly (Cariani 1992). That's why the Sussex team chose to evolve line-drawing robots, rather than programmed agents displaying on a computer screen.

Suppose, for argument's sake, that aesthetically acceptable line-drawings do evolve thanks to such physical interactions. In that case, the creative responsibility of the human artist will be somewhat less, and that of the robots (better: of the robots-plus-environment) will be increased. To be sure, we're still speaking of strong-sense versus weak-sense autonomy. These robots aren't evolving decision-processes comparable to those underlying human freedom. But the human being's creativity, *considered as the origin of certain specific aspects of the final-generation drawings,* is highly indirect, if not irrelevant.

In short, to regard human artists as the undisputed creators of *all* evolutionary art may be to speak too soon.

Credits

Helpful comments were provided by Jon Bird and Paul Brown.

References

Bird, J., and Layzell, P. (2002), 'The Evolved Radio and its Implications for Modelling the Evolution of Novel Sensors', *Proceedings of Congress on Evolutionary Computation*, CEC-2002, 1836-1841.
Bird, J., Stokes, D., Husbands, P., Brown, P., and Bigge, B. (forthcoming), 'Towards Autonomous Artworks', *Leonardo Electronic Almanac*, to appear – probably January 2007. (Available from jonba@sussex.ac.uk.)
Boden, M. A. (1996) 'Autonomy and Artificiality', in M. A. Boden (ed.), *The Philosophy of Artificial Life* (Oxford: Oxford University Press), 95-108.
Boden, M. A. (2006), *Mind as Machine: A History of Cognitive Science*, 2 vols. (Oxford: Oxford University Press).
Boden, M. A. (in press), 'The Aesthetics of Interactive Art', in C. Makris, R. L. Chrisley, R. W. Clowes, and M. A. Boden (eds.), *Art, Body, Embodiment* (Newcastle: Cambridge Scholars Publishing).
Brooks, R. A. (1991), 'Intelligence Without Representation', *Artificial Intelligence*, 47: 139-159.
Cariani, P. (1992), 'Emergence and Artificial Life', in C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen (eds.), *Artificial Life II* (Redwood City, CA: Addison-Wesley), 775-797.
Dennett, D. C. (1984), *Elbow Room: The Varieties of Free Will Worth Wanting* (Cambridge, Mass.: MIT Press).
Kember, S. (2003), *Cyberfeminism and Artificial Life* (London: Routledge).

Linsker, R. (1988), 'Self-Organization in a Perceptual Network', *Computer*, 21: 105-117.
Sims, K. (1991), 'Artificial Evolution for Computer Graphics', *Computer Graphics*, 25 (no. 4): 319-28.
Todd, S. C., and Latham, W. (1992), *Evolutionary Art and Computers* (London: Academic Press).
Whitelaw, M. (2004), *Metacreation: Art and Artificial Life* (London: MIT Press).
Zeleny, M. (1977), 'Self-Organization of Living Systems: A Formal Model of Autopoiesis', *International Journal of General Systems*, 4: 13-28.

Biographical Notes

Margaret A. Boden is Research Professor of Cognitive Science at the University of Sussex. She is a member of the Academia Europaea, and a Fellow of the British Academy and of the American Association for Artificial Intelligence. In 2002 she was awarded an OBE "for services to cognitive science." Her writing has been translated into 20 foreign languages. Her latest books are THE CREATIVE MIND: MYTHS AND MECHANISMS (2nd edn. 2004) and MIND AS MACHINE: AHISTORY OF COGNITIVE SCIENCE, 2 vols. (2006).