CREATIVITY AND ARTIFICIAL EVOLUTION*

Margaret A. Boden
University of Sussex

Abstract: Some cases of creativity involve fundamental transformations, wherein a novel structure is generated which could not have been generated before. Biological evolution exhibits such cases, and many A-Life researchers (and artists) use genetic algorithms in trying to emulate them. Some critics argue that only physically embodied systems can be genuinely transformational, because only they are open to new types of information. However, when creativity in thought (as opposed to bodily morphology) is promoted by external events, these are usually events with semantic significance, not mere physical happenings. It follows that physical embodiment is less important than in the biological case, and that 'virtual' AI/A-Life could in principle produce creative transformations.

Keywords: creativity, transformation, open evolution, embodiment, AI, A-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: The Puzzle of Transformational Creativity

Creativity is the ability to produce ideas/artefacts that are *novel, surprising,* and *valuable.* Of the three types of creativity (combinational, exploratory, and transformational: Boden 2004), the most interesting is the third. That's because it's the most surprising. For in transformational creativity, new structures are created which were not only unexpected but which seem, on arrival, to be impossible. Indeed, they may still seem 'impossible' even after they've occurred.

Given the stylistic rules and constraints which were previously operative, these novelties simply *could not* have arisen. And yet they did.... How can this be?

In general, this involves the transformation of a previously-existing conceptual space, by modifying (or dropping) one or more of its defining dimensions. A conceptual space is a style of thinking--where "thinking" includes conceptualizing, theorizing, composing, painting, sculpting, dancing, cooking ... and so on. Each space/style is defined by a set of generative rules (not necessarily conscious) that both constrain and enable what can be produced. Whereas exploratory creativity involves the application, and perhaps the minimal 'tweaking', of these rules so as to generate stylistically acceptable structures never realized before, transformational creativity actually alters the style.

That is, it alters one or more of the defining rules. The greater the alteration, and the deeper the rule, the greater the stylistic change. In some cases, the change is so great that it may be difficult

(at first) even to see any relation between the old and new styles. When that's so, the new structure is especially likely to be rejected--even by members of the author's peer-group. (Pick your own examples from the history of science and/or art.)

The distinction between exploratory tweaking and transformational change is often unclear, although many specific examples of change may be unhesitatingly classified as the one or the other. Certain creative methods/heuristics are used in widely different conceptual domains: e.g. *iteration* and *consider the negative.* Iteration enabled (for example) the generation of a multi-frilled flamenco-skirt from a skirt with a single frill. Consider-the-negative enabled (for example) Friedrich von Kekule's conversion of string molecules, or open curves, to ring molecules, or closed curves, in discussing benzene (see Boden 2004: 25-28, 62-71, 261-263). In general, the former process delivers exploratory tweaks whereas the latter provides transformations.

With respect to computer models of creativity, the general public tend to assume that transformational creativity is impossible: "The program does what it does: no possibility of transformations, there!" They also assume, or so I've found, that combinational creativity is easily modelled. They're wrong on both counts.

Combinational creativity is, in fact, highly problematic for AI researchers. Granted, new combinations can be produced *ad infinitum* by the computer. However, its lack of linguistic subtlety and impoverished world-knowledge make most of these uninteresting. One of the best efforts in this area at present is Kim Binsted's joke-generating program (Binsted 1996). This relies on a thirty-thousand-word semantic network, whose built-in links represent relations not only of meaning but also of morphology, phonetics, and even spelling. These linkages are explored in generating riddles fitting various familiar templates, such as *What sort of X is Y?* and *What do you get if you cross an X with a Y?* Examples of jokes produced by the program include: *Q: What do you call a despressed train? A: A low-comotive* and *Q: What do you get if you cross a sheep with a kangaroo? A: A woolly jumper* (US readers should note that in British English, one meaning of the word "jumper" is sweater).

It would be misleading to say that Binsted cheated, even though she had to provide the links representing spelling and phonetics, and even though she also had to add the facts (world knowledge) that kangaroos jump and sheep bear wool. For fitting the schematic templates to specific network items is not a simple matter. Moreover, even this effort helps us to think clearly about just what's involved when people make, or understand, such jokes. Nevertheless, there's no prospect of actually building a riddle-generator sufficiently knowledgeable to produce the wide variety of jokes found, for instance, in Christmas crackers. And if we wanted even subtler combinations, the practical difficulties would multiply further. In short, while mere combinations are no problem at all for computers, *valuable* (interesting, funny, arresting, apt, ironic ...) combinations will remain in short supply.

Computer models are far better suited to exploratory creativity. Given that the relevant conceptual space has been computationally defined (rarely an easy matter!), the computer can explore it until kingdom come. In so doing, many structures may be generated which had never been created before. And most, or even all, of them will be acceptable to us, because the evaluation has in effect been folded into the program implicitly. After all, to say that a style is "accepted" is to say that instances generated within this style will be generally regarded as

valuable.

AI programs that reliably produce highly acceptable results in artistic domains include examples focussed on music (e.g. Cope 2001, 2006; Hodgson 2005), architecture (Koning and Eizenberg 1981), line-drawing (Cohen 1981; McCorduck 1991), abstract 3D imaging (Todd and Latham 1992), and colouring (Cohen 2002). As for science, the BACON program and its siblings (BLACK, GLAUBER, STAHL, and DALTON) written by Herb Simon's research group model scientific induction, and illuminate certain aspects of how many human scientists go about their work.

None of these, however, model transformational creativity. The art-focussed programs come up with many genuinely new 'thoughts' (pictures, music, architectural designs...), but all lie within the particular style defined beforehand by the programmer. As for the BACON suite, these only very rarely generate a new idea (e.g. a new way of expressing a familiar scientific law). Mostly, they *rediscover* already-known relationships. In other words, the BACON suite rely on the pre-existing thinking style, as an unchanged conceptual space. They have to be spoon-fed with the relevant questions, even though (by means of exploratory creativity) they can find the answers for themselves.

What their human namesakes had done was different, for they'd viewed the data in new ways. Indeed, they'd treated new features as data. To identify *mathematical* patterns in the visual input from earth or sky had been a hugely creative act when Galileo Galilei first did it. To look for numerical constants was another, and to seek simple linear relationships *before* ratios or products (search priorities that were built into the heuristics used by the BACONites), yet another. These programs were deliberately provided with the styles of thinking which Francis Bacon, Johann Glauber, Georg Stahl, and John Dalton had pioneered for themselves (and were roundly criticized accordingly--Hofstadter and FARG 1995: 177-179; Collins 1989). And what's true of BACON is true of most other automatic scientists. Admittedly, seven AI-aided discoveries have been sufficiently novel, and interesting, to be published in the relevant journals; in every case, however, the programmers were crucial in formulating the problem and/or manipulating the data and/or interpreting the results (Langley 1998).

But if transformational creativity wasn't achieved by any of those programs, might it be achieved some day by different AI systems? Has it, perhaps, been achieved already?

The common prejudice against the possibility of modelling transformational creativity in computers would dismiss this as a vain hope. Nevertheless, there are some *prima facie* candidates. In particular, there are the 'evolutionary' AI programs, based on genetic algorithms, or GAs (Boden 2006: 15.vi). For GA-programs *can* transform their own rules, so that structures can be produced in later generations which *could not* have occurred before. In short, a form of transformational creativity sometimes occurs in artificial systems.

GAs, of course, are inspired by biological evolution. That, over and over again, has produced fundamentally novel forms: vertebrates from invertebrates, for instance. These transformations in morphology are deeper, more surprising, than merely iterative changes--such as segmentation, wherein a structure is repeated many times without significant variation (think of millipedes, or earthworms). To be sure, some basic metabolic and genetic processes are shared by all, or

virtually all, living things (Boden 1999). And certain body-types crop up repeatedly. But they didn't crop up at all in the beginning. Biological organisms, then, result from a staggeringly creative process of transformation.

Indeed, it's often said that biological evolution is "open-ended", in the sense that it can come up with fundamental novelties. For example, it has created novel types of sensory organ. As well as producing *improved* eyes, ears, or noses, it has created *first-time* eyes, ears, or noses. And, of course, it has generated various types of communication and representation--including (in *Homo sapiens)* language and thought.

The question relevant here is to what extent 'evolutionary' AI/A-Life can do the same sort of thing. It's not obvious that it can. Indeed, the A-Life community recently identified "open-ended evolution" as the most important of fourteen "grand challenges" for the field (Bedau et al. 2000; Rasmussen et al., 2003). In other words, are the transformations produced by GA-programs genuinely *open?* Are they really transformations at all? Or are they merely deceptively impressive cases of exploratory creativity, where the possibility-space being explored has been implicitly pre-defined by the programmer?

II: When is a Transformation not a Transformation?

A number of people would answer the question posed in the sub-title above by saying "When it's derived by a GA." For they've argued that the *exploration* of pre-existing possibility-spaces is the best that GA-programs can offer us. True transformations, they say, are out of reach in principle. All the possibilities of genome-change (to use a deliberately neutral term) are built in from the start.

Are they right?

Well, some GA-programs certainly *appear* to be radically transformational. One such example is Tom Ray's *Tierra* program (Ray 1992), and another is Karl Sims' computer graphics program (Sims 1991).

Consider Ray's work, first. *Tierra* models the co-evolution of two or more species, where the individual "creatures" (his word) are strings of self-replicating computer code, competing for space in computer memory. This program is best known for having generated parasites, counter-parasites, and hyper-parasites. (A parasite, here, is a string of computer code that has lost its self-replicating portion but has evolved the ability to borrow the self-replication code included within other, longer, strings; counter-parasites are self-replicating strings that are able to protect themselves from such borrowing.) These new species weren't specifically built into the program, nor expected by the programmer himself. But another feature of *Tierra* is even more relevant at this point of our discussion. Namely, it appears to model what some biologists have termed "punctuated" evolution (Eldredge and Gould 1972).

Punctuated evolution denotes those cases where a new biological species appears in the fossil record *without* any line of near-similar fossil 'ancestors' preceding it. The biologists who named it (and, of course, today's proponents of creationism and/or "Intelligent Design") argued that this

sudden appearance of fundamentally novel forms--in our terms, of transformations--cannot be explained by Darwinian theory, which assumes that only gradual changes can happen.

Prior to Ray's work, all that an orthodox Darwinian could say at this point was that *there must be* some undiscovered fossils in the rocks, or that *there must have been* a line of intermediate forms, even if for some reason these didn't fossilize. However, those *musts* were professions of (Darwinian) faith rather than proven facts. *Tierra,* by contrast, proved that sudden phenotypic changes can indeed happen in a purely Darwonian way: that is, as a result of tiny mutations accumulating in the genotype over many generations. For Ray found that background mutations (that is, ones *not* causing a change in the phenotype) can sometimes accumulate over many hundreds, even thousands, of generations--only to cause a sudden change in the phenotype when a few 'last-minute' mutations occur and interact with them. In short, hidden gradualism (Darwinian evolution) can produce observable saltations--alias 'sudden' transformations.

Turning from the fossil record to computer graphics, let's now consider Sims' GA-generated images (Sims 1991). Here too, the results--the colours and shapes within the images--are not only unpredictable, but can also be shocking. For here, too, there may be no discernible family resemblance between members of the same 'lineage'. If that's not transformation, what is?

Sims' GA-program is indeed transformational--but it's *not* genuinely evolutionary, as that term is normally understood. One might say that the program is too transformational for its own good. For Sims' permitted (random) mutations include some that are much deeper than others, such as the nesting of an entire image-generation program within another one, or the concatenation of two such programs. Moreover, both of these can happen repeatedly, in which case the artificial genome acquires many-levelled hierarchy and/or greatly increased length. When nesting and concatenation are brought into play, there may be no visible resemblance even between images that are members of *successive* generations, never mind more distant ancestors. It's almost as though a cat had given birth to a puppy, or a fish, or even a buttercup.

The results, therefore, are so unconstrained that no fitness function, whether automatic or interactive, could reliably shape the generated images into some aesthetically preferred (or 'adaptive') form. Such shaping might take place over a few generations, if only relatively superficial mutations had happened to occur at each stage. But eventually, a deeper transformation (sic) would happen, sending the seeming 'progression' off course. (Or, to avoid the assumption that evolution is inherently progressive, turning the seeming evolutionary pathway into a dead-end.)

(Sims, being a computer scientist rather than a professional artist, doesn't care about that. If anything, he glories in the rampant unpredictability of the images. The sculptor William Latham, by contrast, does care--and deliberately avoids such unfettered variation: Todd and Latham 1992. He wants his interactive GA-program to stay within particular form-spaces that he finds aesthetically interesting. Consequently, no nesting or concatenation is allowed. Rather, the mutations change superficial parameters, and/or combine a small range of specific elemental forms. As a result, all of Latham's computer-generated images bear his personal 'signature'--an important matter for almost all professional artists. A few, however, deliberately try to *avoid* an authorial signature: see Boden and Brown, in preparation.)

In short, transformation doesn't guarantee evolution. That's clear in the biological realm, wherein deep genomic transformations are likely to be disastrous. Usually, they don't even give rise to a viable organism, still less one which can win the competition for reproduction. In the artificial realm too, Sims' program shows that transformations can destroy incipient signs of evolution. (This soon becomes evident if one plays around with his system, trying to satisfy one's own preferences--for colours, for instance.)

Nevertheless, our original question remains. Is AI-creativity confined only to exploration? Sims' program, after all, produces its surprising--and evanescent--'transformations' only by exploiting the possibilities allowed by the GAs defined for nesting and concatenation. Similarly, the apparently transformational *Tierra* can exploit only those types of mutation which Ray provided when he wrote it. So perhaps these two programs are only 'exploratory', really? Could any other AI systems involve genuine transformations, which *do* lead to fundamental (and sustained) changes comparable to those seen in biological evolution?

The objection that GAs can provide no more than the exploration of pre-existing possibility-spaces is especially telling when the length of the artificial genome is fixed. For in that case, any structural complexities which would require a longer genome simply can't arise.

However, we've already seen that Sims' program allows for a sudden lengthening of the genome, with a corresponding explosion in the possibility-space. And Inman Harvey's SAGA algorithm enables the length of the genome to vary in a more gradual manner (Harvey 1992). So structures that didn't lie within the potential of the original genome (more strictly: the original genome-plus-mutation-rules) can arise. As Harvey put it, there can be new "species", as well as new varieties.

Examples of SAGA-evolved forms produced at the University of Sussex include a simple feature detector in a robot's 'brain', broadly comparable to the orientation detectors found in visual cortex (Harvey, Husbands and Cliff 1994; Husbands, Harvey and Cliff 1995). This mini-network, which was sensitive to a light-dark gradient at a particular orientation, evolved unexpectedly from a controller-network that was initially random.

One can say that it was genuinely *evolved by* the GA, not merely (like Sims' surprising images) *generated by* it, because all its robot ancestors were selected by the same fitness function. This was intended by the programmers to result in the ability to navigate within a 'toy' environment. That environment happened to contain a white-cardboard triangle, and the feature detector evolved as part of an adaptive visuomotor mechanism. In other words, its connections to visual and motor units within the whole network enabled the robot to use the triangle as a navigation aid. This ability, once it had emerged, was selected/maintained by the fitness function.

Other SAGA-evolved forms included a network controlling sixteen robot motors--two for each of an octopod-robot's legs--in coordination, so that the robot could move coherently and 'appropriately'. This circuit took account of signals from four infra-red sensors, ten light sensors, and various sensory bumpers and whiskers (Jacobi 1998). Indeed, evolutionary robotics in general shows many examples of random networks being transformed (sic) into functional sensorimotor circuits (e.g. Beer 1990).

For all that, some researchers still insist that AI can't get beyond exploratory creativity. Even varying the length of the genome, they say, doesn't avoid the problem, for the possibility-space--albeit huge--is still inherently fixed by the nature of the program. For example, Paul Hodgson--who has written some impressive programs, both exploratory and GA-based, that improvise (in real time) in the manner of Charlie Parker--argues that deep *stylistic* change, such as that achieved by Parker himself, necessarily lies beyond the scope not only of SAGA but of *any* GA program of the current type (Hodgson 2002, 2005). (Such transformations might be possible, he feels, if he himself added a great deal more musical information to constrain the changes allowed; but even so, the nature of the transformation would have to lie within the generative potential of the human-authored program.)

Hodgson's not alone: other musicians responsible for GA-based composer-programs agree (Gartland-Jones and Copley 2003). Moreover, in artificial GA systems the fitness function itself is fixed, whereas in nature new fitness functions can arise. Within AI, such objectors conclude, truly transformational creativity--such as we see in biology and in human thought--is impossible.

III: Informational openness

An AI-enthusiast might try to meet the challenge just remarked by saying that, likewise, an organism's DNA (plus the mechanisms of meiosis and mutation) defines a finite--though enormous--set of possibilities. So if DNA genomes can create genuine novelties, why can't artificial genomes do so too?

The answer often given by critics of orthodox AI is that *embodiment,* and environmental *embeddedness,* are crucial. That is, biological form doesn't result only from DNA. For DNA exists in a richly varying physical environment that can affect it in unpredictable, evolutionarily fruitful, and sometimes transformative ways. It follows (on this view) that 'virtual', or purely programmed, AI can't give us what we want--namely, genuine transformations. Only physical models, such as certain types of (situated) robot, could do that.

This is now an increasingly popular position within A-Life, but it dates back to an A-Life pioneer of the mid-twentieth century, William Ross Ashby (Ashby 1952a,b; Boden 2006: 4.viii.c-d). (It's an intriguing historical fact that a copy of Ashby's 1952a paper was among the few such items possessed by Alan Turing; it's now held in the Turing Archive.) Ashby's "law of requisite variety" stated that the variety in a system's regulator must be at least as great as the variety in the system being regulated. So if the system is to outstrip the possibilities inherent in its regulator, it must be informationally open. That is, it must be capable of being affected by events that weren't specified, whether explicitly or implicitly, within the regulator.

Where the regulator is a program, this means that the programmed system must sometimes be affected by the external world. An interactive program, such as Sims' or Latham's image-generator, could be said to be "affected by the external world" in the sense that it's the human being--that is, a force external to the system--who selects the individuals for reproduction at each generation. However, the transformations (if any) that happen are generated by prespecified GAs, as we've seen, so lie within a predefined space of possibilities. A truly open system (or so the proponents of embodiment claim) is one which is *physically* embedded in the environment so

that its *physical* nature may be affected--i.e. altered--by causal interactions with states or events in the external (material) world.

Where the regulator is DNA (plus natural selection), physical interactions with random and/or previously irrelevant external events--including mutagenic cosmic rays--can lead to a progressive adaptation of organism to environment, wherein radically new senses and forms of behaviour (and new behavioural 'problems') appear. Where the regulator is a program, informational openness requires--analogously--that the system's physical states sometimes be affected by causal interactions with seemingly irrelevant states or events in the external world. For example, a robot programmed (as a result of GAs) to draw lines of certain sorts will sometimes be influenced by physical factors such as friction, or accidental jolting--so that the evolutionary process is even less predictable than would otherwise be the case (Bird et al. 2006).

The problem with (most) artificial evolution, then, is not that the system is artificial, but that it is informationally closed. Artificial systems that *are* open to environmentally triggered change may show transformations that are truly radical.

This was exemplified half-a-century ago by the exceptionally creative cybernetician Gordon Pask (1959; Boden 2006: 4.v.e). One of his "evolutionary" chemical machines grew threadlike crystalline structures, dynamically balanced between the deposition and re-solution of ions. These were supposedly analogous to concepts, since they could discriminate sounds of different pitch (50 or 100 Hertz), or the presence/absence of magnetism, or differences in pH level. Pask's chemical threads weren't *initially designed* in order to discriminate pitch (or magnetism, or pH), but *naturally arose* in a way that made this discrimination possible. In other words, they were a primitive example of the self-organization of *new* perceptual abilities, creating *new* perceptual dimensions--which happens from time to time during phylogenetic evolution (Cariani 1993, 1997).

A few of Pask's contemporaries realized the significance of his result. Warren McCulloch, for instance, said (in his 'Preface' to Pask 1961) that Pask had illuminated "the central problem of epistemology", namely: where do our concepts *come from?* But no system-builders followed in Pask's footsteps. His electrochemical ear was seen as a mere curiosity, and was soon well-nigh forgotten. The world would have to wait for 35 years before its theoretical significance was realized (Cariani 1993--and see Boden 2006: 15.vi.c), and for almost 50 before a comparable example occurred (Bird and Layzell 2002).

Even that example was unexpected. Jonathan Bird and Paul Layzell (2002) found themselves confronted with a primitive radio antenna (a "radio wave sensor"), which picked up and modified the background signal emanating from a nearby PC monitor. They had been using a technique for evolving circuits in hardware (Thompson 1995), in order to evolve an oscillator--not a radio. But, to their surprise, some of their devices exploited unconsidered and/or accidental aspects of the physical environment to produce radio reception.

These aspects included the aerial-like properties of printed circuit boards, the proximity of the PC, the order in which the analogue switches had been set, and the fact that the soldering-iron left on a nearby workbench happened to be plugged in at the mains. This is a mixed bag: Bird and Layzell concluded (p. 1841) that the open-ended evolution of novel sensors requires a

physical device, whether natural or artificial, whose primitives are sensitive to a wide range of environmental stimuli.

IV: Transformations in Psychology

Radio-receivers, orientation detectors, and even biological eyes and ears are all very well. But what about transformational creativity in the specifically human realm, such as (sometimes) occurs in music, painting, literature, or scientific thought? Could these transformative novelties have arisen in real life, or could they be emulated in AI systems, thanks to serendipitous *physical* interactions--like those which took place in Pask's glass containers, or in the Sussex robot's toy environment, or on Bird and Layzell's laboratory bench?

In *Homo sapiens,* the "environment" is not only physical/biological, but sociocultural too. Some of the external factors that affect the organism--better: the organism's mind, the person--are driven by and/or cause physical events that implement, or are interpretable as, *thoughts.* Speech-sounds, and marks of ink on paper, are the most obvious cases. But other cases include totem poles, commemmorative statues, and painted cathedral triptychs reminding worshippers of the story of Christ. All those material objects have semantic content, in the form of some cultural significance. It's relatively rarely that a *merely physical* stimulus leads to interesting (valuable) creative thinking in science or art.

There are exceptions, however. Marcel Proust's memory-triggering madeleine is one: the mere tasting of the cake (a physical object with no semantic content) triggered a host of memories and productive thoughts in Proust's mind. Another concerns a jazz-drummer with Tourette's syndrome (Sacks 1985), who was a fine enough musican to be able to use the random noises caused by uncontrollable tics in his hands as the seeds of improvisations which *could not* have happened otherwise. And it's highly probable that some of the deeply surprising (transformationally creative) leaf-arrangements and ice-sculptures created by the artist Andy Goldsworthy were grounded in serendipitous perceptions of fallen leaves or melting ice in their natural state. (A research-programme in which I'm currently participating is attempting to evolve robots that will draw aesthetically acceptable marks, thanks partly to accidental interactions with their physical environment: Bird et al., 2006; Brown 2006/7. It remains to be seen whether we'll succeed.)

As for scientific creativity, think of Alexander Fleming's accidentally abandoned Petri dish: the unexpected physical stimulus (the clear areas visible in the dish) led him to infer that there might be a bactericidal substance there. Even Kekule's benzene ring, according to his own report many years later, was triggered by the specific shapes of the flames dancing in his hearth. The 'joining' of two leaping flames prompted him to draw a visual analogy to a tail-biting snake (that is: to a normally-open *string* that's unexpectedly closed to form a *ring),* which in turn led to his transformative ideas on chemical structure. (This example, wherein a specific analogy prompted a general transformation, illustrates the fact that the distinction between the three types of creativity is an analytic one: in practice, they may occur together.)

Clearly, then, the various forms of unpredictability that characterize creative thought (Boden 2004: ch. 9) include some serendipitous events that are *physical* rather than meaningful.

However, we must remember Louis Pasteur's famous observation that "Fortune favours the prepared mind". Only a highly accomplished jazz-drummer could see how to relate random taps to musically meaningful schemas for improvisation. Similarly, only an expert chemist (already aware that the neighbour-relations between the atoms within a molecule are important) could see the potential chemical relevance of a topological change from string to ring. In other words, the physical event must be *understood as* meaningful and/or *assimilated into* meaningful schemas, if it is to have its creative effect.

Typically, however, the unexpected stimuli that lead to transformational thought are 'thoughts' themselves. (I put this term in scare-quotes because I intend it to cover thoughtfully produced artefacts, such as music or sculpture, as well as examples of speech and text.) Remember Immanuel Kant's being, as he put it, suddenly "awoken from [his] dogmatic slumbers" by reading David Hume. Or think of Henry Adams' (1900) highly creative response to the contrasting artefacts he encountered in the cathedral and the turbine-hall.

Some of these thoughts come into the mind from the external world: other people's remarks, sentences read in books, items heard in concert-halls, sights seen in art-galleries or turbine-halls or even fire-places.... Their transformational effect in the relevant person's mind may be more or less immediate (recall Kekule's reaction "But lo! What was that?", on seeing two flames suddenly linking together). Alternatively, they may be stored away in the recesses of the memory, to be resuscitated only when the relevant creative process (the painting, poetizing, composing, theorizing ...) gets under way.

Yet other transformation-triggering thoughts come not from the external world as such, but from that part of the person's internal world which appears to be *external to the specific task in mind.* In other words, they are generated (and maybe also stored away for future use) within the person's own mind, on the basis of already-existing thoughts and mental associations. These "thoughts and mental associations", of course, include various highly general exploratory and transformative heuristics--such as *iteration* and *consider the negative* (both mentioned above). But even in those cases, previous *outside* influences--the person's social interactions, their education and reading, and their cultural experiences in general--are crucial. No man, after all, is an island. Our language and cultural communication are guarantors of that.

The implication is that plausible, non-cheating, AI models of *psychological* transformational creativity should be open (sic) to semantically interpretable events and representations from outside, and/or to internally generated representations ('thoughts') that are *prima facie* irrelevant.

How might such external and/or 'irrelevant' representations arrive, apart from their being deliberately provided by the programmer? Well, the program might be able, for instance, to communicate with other programs carrying out broadly comparable tasks but with a different 'educational' experience--including different data and/or heuristics. It might even be able to trawl the web, maybe googling for human-composed texts/music that could turn out to be helpful for the task at hand.

That possibility was foreseen by Simon, who suggested that future versions of his machine-discovery programs (Langley et al. 1987) might eventually be able to 'read' and learn from scientific papers, as human scientists can. Indeed, he described scientific discovery as a matter of

"social psychology, or even sociology", because publication provides a "blackboard" that hugely extends the individual scientist's memory (Simon 1997: 171f.). At present, he said, machine discovery systems "are still relatively marginal participants in the social system of science". We might be able to hook them up to data-mining programs, to reduce their reliance on human beings for providing their data and problems. However, "Even this is a far cry from giving machines access to the papers, written in a combination of natural, formal and diagrammatic language, that constitute a large part of the blackboard contents" (173).

Alternatively, a program might spend its down-time in generating--and storing--novel representations in a relatively unconstrained manner, for possible future use in as-yet-unforeseen creative tasks. Novel *combinations* would be especially easy to produce, given that the program would be making no judgments on their 'relevance'. (As we saw in Section I, evaluating their interest would be much more problematic.) But more or less adventurous, and unfocussed, *explorations,* and even conceptual *transformations,* could take place too.

However, enabling AI systems to utilize externally or internally generated representations in those ways is a tall order. If programs are ever to be able to respond fruitfully--and sometimes transformationally--to serendipitous 'thoughts' rather than mere key-words, or to accidentally encountered musical structures/techniques rather than mere note-sequences, huge progress will be needed in AI. Specifically, this will require significant advances in research on associative memories, on the recognition of analogy, on natural language processing, on musicology and comparable domain-specific studies, and on knowledge representation in general.

I'm not at all sure that this will be feasible in practice, except in toy (and largely pre-ordained) examples. Indeed, I doubt that AI will ever attain the full richness and subtlety of human thought, and therefore of human creativity (Boden 2006: 7.iii.d, 9.x.e). But it's not impossible *in principle.* (I'm assuming here that human thought is a matter of complex computation, or information-processing; for a host of arguments/evidence to this effect, see Boden 2006.) The major obstacle is the mind's complexity, not the computer's lack of embodiment.

V: Conclusion

Informational openness has contributed to transformative biological evolution, and to some artificial versions too. In all cases, this evolution has occurred entirely within the *overall* possibility-space. However, there's no philosophical cheating here. Any novel structure, whether biological or psychological, is of course made possible by--and accordingly constrained by, and confined within--the totality of pre-existing conditions, environment included. So in biology, new sensory organs don't spring out of nothing: typically, they are adaptations of pre-existing structures, which may or may not have had some other function (e.g. the bones of the inner ear evolved from jaw parts, which in turn evolved from gill-slits). But this is just to say that creativity doesn't happen by magic. Creation *ex nihilo* is an illusion, although creativity is not.

In sum: I've argued two main points. First, that AI systems which are informationally open can in principle exhibit truly transformational creativity--and some have already done so. Second, that in order to model this type of creativity as it occurs in arts and science, AI-researchers will have to consider interactions with stimuli that are meaningful, not merely physical. A-Life's

emphasis on embodiment is well taken. But novel sensors (and motor capacities), intriguing though they undoubtedly are, aren't enough.

References:

Adams, H. (1900), 'The Dynamo and the Virgin', in H. Adams, *The Education of Henry Adams* (Boston: Houghton & Mifflin, 1918), 379-390.

Ashby, W. R. (1952a), *Design for a Brain: The Origin of Adaptive Behaviour* (London: Wiley). (2nd edn., revd., London: Chapman, 1960.)

Ashby, W. R. (1952b), 'Can a Mechanical Chess-Player Outplay Its Designer?', *British Journal for the Philosophy of Science,* 3(9): 44-57.

Bedau, M. A., McCaskill, J. S., Packard, N. H., Rasmussen, S., Adami, C., Green, D. G., Ikegami, T., Kaneko, K., and Ray, T. S. (2000), 'Open Problems in Artificial Life', *Artificial Life,* 6: 363-376.

Beer, R. D. (1990), *Intelligence as Adaptive Behavior: An Experiment in Computational Neuroethology* (Boston: Academic Press).

Bird, J., Stokes, D., Husbands, P., Brown, P., and Bigge, B. (2006), "Towards Autonomous Artworks', submitted to *Digital Creativity.* (Available from jonba@sussex.ac.uk.)

Boden, M. A. (1999), 'Is Metabolism Necessary?', *British Journal for the Philosophy of Science,* 50: 231-248.

Boden, M. A. (2004), *The Creative Mind: Myths and Mechanisms.* 2nd edn., expanded/revised (London: Routledge).

Boden, M. A. (2006), *Mind as Machine: A History of Cognitive Science,* 2 vols. (Oxford: Oxford University Press).

Boden, M. A., and Brown, P. (in preparation) 'Personal Signatures and Computer Art'.

Brown, P. (2006/7), 'From Systems Art to Artificial Life: The Drawbots Project at Sussex University', *CMCA: Computational Models of Creativity in the Arts* (Workshop at Goldsmiths College and DANA Centre, London, May 2006). Official publication in *Digital Creativity,* 18(1), 2007.

Cariani, P. (1993), 'To Evolve an Ear: Epistemological Implications of Gordon Pask's Electrochemical Devices', *Systems Research,* 10 (3), 19-33.

Cariani, P. (1997), 'Emergence of New Signal-Primitives in Neural Systems', *Intellectica,* 2: 95-143.

Cohen, H. (1981), *On the Modelling of Creative Behavior.* RAND Paper P-6681 (Santa Monica, Calif.: RAND Corporation).

Cohen, H. (2002), 'A Million Millennial Medicis', in L. Candy and E. Edmonds (eds.), *Explorations in Art and Technology* (London: Springer), pp. 91-104.

Collins, H. M. (1989), 'Computers and the Sociology of Scientific Knowledge', *Social Studies of Science,* 19: 613-624.

Cope, D. (2001), *Virtual Music: Computer Synthesis of Musical Style* (Cambridge, Mass.: MIT Press).

Cope, D. (2006), *Computer Models of Musical Creativity* (Cambridge, Mass.: MIT Press)

Eldredge, N., and Gould, S. G. (1972), 'Punctuated Equilibria: An Alternative to Phyletic Gradualism', in T. J. M. Schopf (ed.), *Models in Paleobiology* (San Francisco: Freeman Cooper), 82-115.

Gartland-Jones, A., and Copley, P. (2003), 'The Suitability of Genetic Algorithms for Musical Composition', *Contemporary Music Review,* 22(3): 43-55.

Harvey, I. (1992), 'Species Adaptation Genetic Algorithms: A Basis for a Continuing SAGA', 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), 346-54.

Harvey, I., Husbands, P., and Cliff, D. (1994), 'Seeing the Light: Artificial Evolution, Real Vision', in D. Cliff, P. Husbands, J.-A. Meyer and S. W. Wilson (eds.), *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior* (Cambridge, Mass.: MIT Press), 392-401.

Hodgson, P. W. (2002), 'Artificial Evolution, Music and Methodology', *Proceedings of the Seventh International Conference on Music Perception and Cognition* (Sydney), 244-247 (Adelaide: Causal Productions).

Hodgson, P. W. (2005), *Modeling Cognition in Creative Musical Improvisation.* D.Phil. thesis, Dept. of Informatics, University of Sussex.

Hofstadter, D. R., and FARG (The Fluid Analogies Research Group) (1995), *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought* (New York: Basic Books).

Husbands, P., Harvey, I. and Cliff, D. (1995), 'Circle in the Round: State Space Attractors for Evolved Sighted Robots', *Journal of Robotics and Autonomous Systems,* 15: 83-106.

Jacobi, N. (1998) 'Running Across the Reality Gap: Octopod Locomotion Evolved in a Minimal Simulation', in P. N. Husbands and J.-A. Meyer (eds.), *Evolutionary Robotics: Proceedings of the First Workshop on Evolutionary Robotics, EvoRobots-98* (Berlin: Springer), 39-58.

Koning, H., & J. Eizenberg. (1981), 'The Language of the Prairie: Frank Lloyd Wright's Prairie Houses', *Environment and Planning, B,* 8: 295-323.

Langley, P. W. (1998), 'The Computer-Aided Discovery of Scientific Knowledge', in S. Arikawa and H. Motoda (eds.), *Discovery Science,* Proceedings of the First International Conference on Discovery Science, Fukuoka, Japan (Berlin: Springer), 25-39.

Langley, P. W., Simon, H. A., Bradshaw, G. L., and Zytkow, J. M. (1987), *Scientific Discovery: Computational Explorations of the Creative Process* (Cambridge, Mass.: MIT Press).

McCorduck, P. (1991), *AARON's Code: Meta-Art, Artificial Intelligence, and the Work of Harold Cohen* (New York: W. H. Freeman).

Pask, G. (1959), 'Physical Analogues to the Growth of a Concept', in D. V. Blake and A. M. Uttley (eds.), *The Mechanization of Thought Processes* (London: Her Majesty's Stationery Office), vol. 2: 877-922.

Pask, G. (1961), *An Approach to Cybernetics* (London: Hutchinson).

Rasmussen, S., Raven, M. J., Keating, G. N., and Bedau, M. A. (2003), 'Collective Intelligence of the Artificial Life Community on Its Own Successes, Failures, and Future', *Artificial Life,* 9: 207-235

Ray, T. S. (1992), 'An Approach to the Synthesis of Life', in C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen (eds.), *Artificial Life II* (Redwood City, CA: Addison-Wesley), 371-408.

Sacks, O. (1985), *The Man Who Mistook His Wife For a Hat: And Other Clinical Tales* (New York: Simon & Schuster).

Simon, H. A. (1997), 'Machine Discovery', in J. Zytkow (ed.), *Machine Discovery* (London: Kluwer Academic), 171-200.

Sims, K. (1991), 'Artificial Evolution for Computer Graphics', *Computer Graphics,* 25 (no. 4): 319-28.

Thompson, A. (1995), 'Evolving Electronic Robot Controllers that Exploit Hardware Resources', in F. Moran, A. Moreno, J. J. Merelo and P. Chacon (eds.), *Advances in Artificial Life.* Proceedings of the Third European Conference Artificial Life - ECAL95 (Berlin: Springer-Verlag), 640-656.

Todd, S. C., and Latham, W. (1992), *Evolutionary Art and Computers* (London: Academic Press).

********** END OF MIT VERSION **********

*** To add to longer version:

Something on latham's freeze and adapt is **genuine** evolution, though not autonomous of course .....

**** OLD VERSION OF CONCLUSION, with THREE main points: To be RE-INSTATED:

In sum: I've argued three main points. First, that AI systems which are informationally open can in principle exhibit truly transformational creativity--and some have already done so. Second, that informational openness aids all three types of human creativity. And third, that to model this type of creativity as it occurs in arts and science, AI-researchers will have to consider interactions with external stimuli that are meaningful, not merely physical. Novel sensors (and motor capacities), intriguing though they undoubtedly are, won't be enough. **** END OF OLD VERSION OF CONCLUSION

*** After "No man, after all, is an island" have a brief thingy out Andy's extended mind/self .....