Towards Artificial Creativity

Matthew Elton

CSRP 288, December 1993

Cognitive Science Research Paper

Serial No. CSRP 288

The University of Sussex
School of Cognitive and Computing Sciences
Falmer
BRIGHTON BN1 9QH
England, U.K.

Towards Artificial Creativity

Matthew Elton
School of Cognitive and Computing Sciences
University of Sussex, BRIGHTON BN1 9QH, U.K.
matthewe@cogs.susx.ac.uk

December, 1993*

Abstract

The production of novel and interesting entities may seem like a good starting point for a definition of creativity, but I argue that any definition of this kind does not take sufficient regard of the way in which an entity came to be produced. Only certain methods of production can lead to attributions of creativity. I suggest that creativity can be thought of as a two stage process of generation and evaluation, and go on to stress the role of evaluation. Two examples from the arts are provided to support this position. With a sketch theory in place, I turn to the notion of artificial creativity. I look at some philosophical worries that arise when attributing to machines abilities which are usually considered to be exclusively human, or at least biological. Some current computer systems are mentioned, and I indicate why I think some are better proto-models of creativity than others. Finally, I speculate on how we might move further towards artificial creativity.

1 Historical Roles

Strict definitions of creativity are bound to provoke much dispute about borderline cases, which, while useful in sharpening up our ideas, does not always help with the bigger picture. So, rather than begin with a definition, I shall begin with a comment on the broad shape any successful definition must take. In particular, I want to show that any proposed definition of creativity must consider very carefully the nature of the process that leads to the production of creative entities, and not just the quality (or intrinsic properties) of those entities themselves. I want to

*A version of this paper has been submitted to the journal *Languages of Design*.

emphasise that with certain processes, regardless of the values of their products, we shall not be inclined to say that creativity has occurred.

To show the importance of processes, and in particular the process of evaluation, we can think about the sort of definition which keeps quiet about them. So, a tentative, and initially attractive, first offering for a definition of creativity might be the production of something that is both novel and interesting. Here, interesting is meant to mean valuable according to some domain specific criterion. This sort of definition, though of the broad brush stroke kind, does capture something of importance about creativity. However, here is an illustration of how it breaks down.

Popular and scholarly opinion holds that Bach was a creative fellow, and that his musical compositions stemmed from his musical creativity. But maybe such opinion is in fact misplaced. Suppose a particularly thorough musical historian were to discover that, as well as producing many musical offerings, Bach was also a pioneering worker in the psychology of music. Extensive empirical research in the field, along with many hours of painstaking quantitative analysis, all without the aid of number crunching devices like computers, enabled him to develop an elaborate theory of good musical composition within several disjoint genres. The theory could be used to generate new compositions simply by instantiating free variables with random values. Now, having developed his theory, all Bach's attempts at publication were thwarted by the lack of any contemporary psychology journals. He sought solace in generating music, with dice and his theory, and passing it off as his own. To his delight these compositions were warmly received, but Bach kept quiet about his compositional technique for fear of public disgrace. Sadly, Bach could not appreciate his own music, being tone deaf. In fact, it was this unfortunate affliction that was the single most significant motivating factor in his psychological research programme.

If this story were true, what would we have to say of Bach's musical creativity? There would be no doubt that novel and interesting entities stemmed from Bach's hand. There would also be no doubt about his qualities as a research psychologist, the continuing success of his music being vindication enough of that. But he would not have been musically creative and his compositions would not be creative entities, resulting as they would have from a process that would not, could not, count as creative.

The aetiology, or causal history, of the fictional Bach's compositions is quite different from the aetiology of the real composer's work. The real composer was guided by his judgement as to which musical ideas worked, and which did not. But the fictional character exercised no musical judgement; this was left to the subjects of his psychology experiments. With the theory of good composition in place no more judgements were required.

The aetiology issue is particularly important in the arts, though sometimes

confusing.[1] If the infamous monkeys in the basement of the British Museum did produce a manuscript for Hamlet, and curiously presented it, and none of their other presumably chaotic writings, to the curator, then surely the play would be just as good as Shakespeare's? There is a sense in which the play's worth is judged by its intrinsic properties. On the other hand, there is a significant sense in which we would say of the manuscript that it was not *about* the human condition, let alone a Danish prince. This tension is less prominent in the sciences. Once a scientific or mathematical idea is put forward, it can be judged entirely on its own merits, independently of its aetiology. But, if the discovery of that scientific or mathematical idea is to count as a creative act, then the aetiology will matter. That said, the involvement of serendipity need not count against the creativity of a discovery, as in the case of Fleming's discovery of penicillin.

It is interesting to note that in the arts the aetiology factor can outweigh all others. So, in conceptual art, what matters is not what is on the canvas, which may be rather dull, but exactly how it got there. There is precious little aesthetic interest in a glass of water on a shelf. But some sort of case can be made for it being a work of art if its placement in a gallery is preceded by certain complex intentions.

In sum, the claim of this section is that any decent account of creativity needs to take the particular history of the entity seriously—it must be shown that the process which produced the entity was one of the right kind.

2 Creative Processes

Talk of the importance of particular processes in creativity is nothing new. Four stages are commonly distinguished: preparation, incubation, illumination, and verification. I prefer a simpler division consisting of generation, subsuming preparation and incubation, and evaluation, subsuming illumination and verification. Generation without evaluation, I contend, just could not count as creativity. Think about a remark which can inadvertently be interpreted as a fine witticism. The remark, witty though it may appear to be, is inadvertent and not creative. But the creative wit, grateful for her good luck on this occasion and realizing the potential future value of the comment, may save it for later, genuinely creative, use.

While some methods of generating product may lead to more successful evaluations, I think there are good reasons to place as much, if not more, emphasis on

[1] **Aetiology:** The aetiological properties of an entity are those concerned with its causal history. Note that entities can be classified solely by their intrinsic properties, i.e. properties fixed by their immediate physical makeup. Such classification ignores the difference between the Mona Lisa and an exact replica, between legal currency and perfect counterfeit. For these reasons we often take non-intrinsic (or relational) properties, such as an entity's causal history, very seriously.

the evaluation process. For example, even the creative worker who follows a strict technique, or working practice, is involved in evaluation. If not during or at the end of each creative cycle, then she will have at some point engaged in the evaluation of her generative methods. And note that, if this is not the case, if no judgement has ever been exercised, then we may well not want to call the worker creative. She may just have hit on a method that produces a rich stream of valued entities, but if she has not been involved in evaluation then she is lucky, rather than creative. Or one might say she is practising a craft, rather than practising an art. We often think of the practitioners of craft as creative people. I think we do so because they make very many careful evaluations as they proceed with their work, even if they do not often generate and evaluate altogether new ideas.

For evaluation, some sense of what is interesting and novel needs to be defined relative to a particular domain. For most domains it is possible to construct a theory, or a grammar, of which entities are broadly successful.[2] Such theories or grammars can be used both in the generation of entities, and also to assess whether entities generated by other means conform to the standard. Examples of theories and grammars of good composition are common in music [20, 18]. The fictional Bach's theory of good composition would take this kind of form. Another example would be the grammar discovered for the design of the Frank Lloyd Wright's prairie house. It turns out there is a precise way to specify the range of possible designs.[3] In all cases where a domain specific grammar or theory can be constructed, success is defined within a formula (with or without free variables). As we proceed I shall contend that this form of evaluation is too static to form the basis of an account of creativity.

Nonetheless, a sound knowledge of the domain at issue does seem important for evaluation. Knowledge of the current state of the art—an awareness of the *Zeitgeist*—is surely essential for evaluating any given entity. Even if your project involves breaking the rules, it will still be understood against the prevailing background of entities which conform to domain norms; entities cannot remain isolated

[2] **Grammars:** A grammar is a way of specifying the legal combinations of a set of symbols, such as a set of notes on a score. These symbols in turn are interpreted as specifying entities in some domain, such as music or architecture. Generally the aim of a grammar is to specify only those symbols sets which specify interesting entities in the domain, i.e. enjoyable musical compositions or effective living quarters. In general a grammar either generates many uninteresting entities (because it is over-general) or misses very many interesting possibilities (because it is too specific).

A grammar defines a conceptual space. See Boden [1] for a discussion of conceptual spaces, and how creativity can be understood as a process of exploring and transforming them.

[3] **The Prairie House Grammar:** Frank Lloyd Wright's Prairie Houses are all unique, but they all have a common architectural feel. By careful analysis of the canon, Koning and Eizenberg constructed a grammar which describes all the extant houses, and shows how new variants can be designed. Some of the grammatically correct houses which can be generated will be more interesting than others, but the grammar defines a space of characteristically Lloyd Wright designs. See [16] for the details.

from their context of creation. And note again how aetiology plays a role. New works which conform to the style of Bach are, providing we know the difference in history, simply less interesting than those written by Bach himself. Consider also Borges tale of Pierre Menard [2], the man who recreated the text of Cervantes's *Don Quixote*. The new work, intrinsically identical to the original, i.e. word for word the same, has a meaning very different from the old. And it does so because it has a very different history of production.

In addition to the creative context, there is often something more fundamental in evaluation. Some pictures, melodies, or poems appear to have more *intrinsic value* than others. We have a direct, perhaps visceral, response to them, which we are largely unable to analyse. Presumably there are evaluative criteria at work here, but we are unable to understand them. They might be unconscious, or perhaps altogether non-conscious. In fact, it often seems impossible to untangle the web of influences that are involved in evaluation. There is some notion of a basic, visceral aesthetic, some notion of a culturally conditioned, but roughly stable, aesthetic, and then there are the important effects of the current cultural milieu.

3 Art, Anti-Art, and Ambience

I shall say more about the nature of evaluation later. But first, I want to provide some support for my stress on evaluation, and my relative indifference to schemes of generation. Note that elaborate schemes of generation are clearly important. My point is that, in themselves, such schemes cannot capture what is essential about creativity. In the following two sections, I discuss two cases where artists explicitly adopt approaches which stress evaluation. My wider claim is that, implicitly, this is how all creative workers must behave, if we are to call their activities creative.

3.1 Dada

Dada was a radical movement in art which began in the early years of the twentieth century. It turned art on its head by rejecting old orthodoxies and continually reinventing itself. Pictures made from magazine scraps, poems made from sounds instead of words, and cabarets where the performers heckled the audience were just some of the features of the movement. Many of the products of the Dada artists were important, not for their intrinsic value, but because of the statement they made. Such was the case with Duchamp's infamous urinal. But whilst this was often the case, and is one of the more readily characterisable aspects of the movement, the people involved were genuine artists and not mere charlatans. Much of their work, such as Kurt Schwitter's brilliant poem Anna Blume [22, (140–141)], maintains its value today.

Dada would try any approach in its quest for the new, the exciting, and the

unpredictable. The deliberate use of chance elements was just the kind of working method which brought delight to the artists of the movement. Hans Richter describes how fellow Dadaist Hans Arp came to chance *by* chance.

> Dissatisfied with the drawing he had been working on for some time, Arp finally tore it up, and let the pieces flutter to the floor of his studio on the Zeltweg. Some time later he happened to notice these same scraps of paper as they lay on the floor, and was struck by the pattern they formed. It had all the expressive power that he had tried in vain to achieve. How meaningful! How telling! Chance movements of his hand and of the fluttering scraps of paper had achieved what all his efforts had failed to achieve, namely *expression*. He accepted this challenge from chance as a decision of fate and carefully pasted the scraps down in the pattern which chance had determined. [22, (51)]

Richter adds: "Was it the artist's unconscious mind, or a power outside him, that had spoken?" I think Arp overstates the case when he claims that the scraps of paper "had achieved ... expression." Arp found that the scraps generated a certain kind of response in him, and made them expressive by choosing to turn them into an art work.

Around the time of Arp's *Nach dem Gesetz des Zufalls (according to the laws of chance)*, many of the Dada artists began making deliberate use of chance in their work. The deliberate use of chance was regarded as freeing the artist from the barriers of causality and conscious volition. It was a way of accessing new thoughts, ideas, and experiences that more methodical working methods would automatically exclude.

Whenever a methodology is too firmly fixed, the resulting work is in danger of becoming staid and unexciting. Moreover, a chance-free methodology tends to force the artist's intention out into the open; the Dadaists felt that they could free their unconscious minds by exposing themselves to the unpredictable. They believed that chance allowed them to access this new and powerful mystery recently promoted by Freud. Richter remarks:

> By appealing directly to the unconscious, which is part and parcel of chance, we sought to restore to the work of art something of the numinous quality of which art had been the vehicle since time immemorial, the incantatory power that we seek, in this age of general unbelief, more than ever before. [22, (59)]

Dada relied on chance to keep the work and the directions of the movement unpredictable, but a line needed to be drawn somewhere. Richter describes how there was division in the movement as to how to walk the perilous tightrope between order and chaos, and this division (amongst others) led to collapse. When

the audience come always to expect the unexpected, chance fails the artist. At this extreme state art becomes anti-art, Dada, and finally nothing. Too much reliance on any methodology will eventually cease to prove interesting and, whilst maintaining their anarchistic streak, the Dada movement tempered their onslaught against causality and order:

> Proclaim as we might our liberation from causality and our dedication to anti-art, we could not help involving our *whole* selves, including our conscious sense of order, in the creative process, so that in spite of all our anti-art polemics, we produced works of art. Chance could never be liberated from the presence of the conscious artist. [22, (59)]

Richter might add that it is not desirable that it should be.

3.2 Eno

A creative worker may still be successful, even though she works with production tools that yield many failures. If she has a high rejection rate, based on critical evaluation techniques, the overall creative system will still yield successes. It might be suggested that such an approach is inefficient, but that would be a misplaced criticism. An efficient but boring system has nothing to boast of. The clear advantage to the high risk approach is that when something interesting does come along it may be more novel and more interesting than the products of a safer system. Brian Eno is an artist who takes this lesson to heart.

Eno is an experimental musician whose career has encompassed the kitsch glam rock of Roxy Music [23, 24], proto-punk synthesizer pop [5, 6], minimalism [12, 13], ambient music [8, 3, 9], sound and light installations, and an impressive string of production credits for other musicians, such as David Bowie, Talking Heads, and U2 [10, 25].

Although, by his own account, Eno is a not a musician, he aggressively declares this an asset. It provides him with an opportunity to exercise a favourite aphorism: "Exploit your weaknesses." He makes extensive use of the modern recording studio, which allows many tracks to be built up in layers, tracks to be recorded many times, and recording at different speeds (thus allowing complex parts to be slowly picked out by the inept Eno). By such means he can construct complex musical pieces without too much reliance on his skill as a performer.

The studio approach allows Eno to build music production systems involving various delay loops and electronic effects. These can generate whole pieces of music with the minimum of intervention, for example *Discreet Music* [7]. Eno, like the Dadaists before him, self-consciously embraces unpredictability in his working methods, and, luckily for us, he has often discussed these working methods publicly. A famous aphorism (and *Oblique Strategy*—see below) is "Honour thy error as a

hidden intention." So although he often works with modern electronic synthesizers, he prefers his equipment to be unreliable. The earlier model of the famous DX7 synthesizer is favoured, because a programmer error in the original design meant that one oscillator acts erratically.

The interest in error is not in any way mystical, and it can be justified by analogy with genetic mutation.[4] In evolution it is the *errors* in genetic duplication which allow organisms to develop and change. As already noted, with such an approach must go an acceptance of a high failure rate. Eno comments: "People would probably be surprised to know my own rejection rate of my work. I must produce a hundred times the amount of music I release." [25, (74)] This high discard rate helps explain how Eno can argue his case as an artist, despite his self-proclaimed lack of technique. Eno has less control over his systems than the conventionally trained musician has over his or her chosen instrument. He does his work as an artist by creating systems of production (which may involve none, one, or a number of people), and by carefully selecting the results of these creative experiments.

In order to further this approach, Eno and the artist Peter Schmidt developed and published a useful creative tool [11]. *Oblique Strategies* are methods for dealing with crisis points in the creative process, and involve an explicit appeal to chance. When the work is not going well, and no acceptable solutions come to mind, the artist or group may elect to use an *Oblique Strategy*. The strategies come in the form of a pack of cards, with each card bearing a particular aphorism, or piece of cryptic advice. When a strategy is required, a card is picked at random. The message on the card must be interpreted so as to apply to the current situation and, once the interpretation has been made, the advice must be followed. Eno remarks that the cards need not be treated mystically, they simply offer the chance of a new perspective or new approach. (See [25, (77–79)] and [10, (98–99)].)

The strategies are half *I Ching* mysticism and half a collection of useful heuristics.[5]

[4] **Genetic Algorithms:** Eno's approach can be compared with the use of genetic algorithms in artificial intelligence research. Rather than design programs by hand, they can be evolved. Each program (phenotype) is specified by a symbolic string (genotype). A population of programs is randomly created and set to work on a task. The most successful programs are then selected for breeding, in which their symbolic strings (genes) are used to generate the next population of programs. Genetic algorithms may make an important contribution to artificial creativity, but the extent to which they will do so depends on how a program's success is determined (the fitness function).

[5] **I Ching:** The *I Ching* [19] is an ancient Chinese text that is used to predict the future, and to offer advice on personal problems. The entries are arranged in 64 divisions, indexed by hexagrams. Each hexagram consists of six lines; each line may be broken or unbroken. When an answer to a question is needed, or a prediction is required, certain chance techniques, such as the tossing of coins, are used to select a hexagram, and to assist in the interpretation of the chosen entry. Unlike the *Oblique Strategies*, the *I Ching* claims it can literally predict the future. Nonetheless, it can be used in the same heuristic way as the *Strategies*, with its attendant mysticism adding ritual flavour.

Either way, they introduce an unpredictable element into the creative process. When you are stuck and don't know where to go, when you have examined all the options and found them wanting, then you need something to break you out of your current set of patterns. This is the function of the *Oblique Strategies*. The unpredictability is important. Simply to have a set of written down heuristics, which you examine in a fixed order to try to help solve a problem, would still place you in some danger of becoming stuck in a rut. It is partly the shrouded mysticism and the sense of ritual which give the cards their value. But it is mostly the chance factor that they introduce.

The cases of the Dadaists and of Eno show that artists who positively embrace chance as part of their working methods can be very successful. And I think this reflects the importance of evaluation over generation. In these extreme cases, the generative techniques maximally employ chance factors, which might ordinarily be seen as the antithesis of creativity. In comparison, a rigid process, with as much chance as possible eliminated, is most likely to produce staid and mediocre results. And even if a system is a reliable generator of good results—like the fictional Bach's system—we want to be sure its performance is being constantly checked and assessed by a qualified party. Without such evaluation, it is inappropriate to argue that a system is a creative one.

4 Talking About Machines

The two case studies should have helped to vindicate my stress on evaluation over generation. I now turn to the specific question of artificial creativity. In this section I consider the very idea of computer based behaviour being described in psychological terms. In the following sections I move on to discuss whether a machine can act as a *qualified* evaluator.

If we want to talk about creativity and artificial intelligence (AI) in the same breath we need to be quite sure that any definition, be it strict or otherwise, does not automatically exclude the possibility of an AI system being creative. Psychological terms, such as "intelligence," "understanding," "consciousness," and "creativity" are so closely bound to the behaviour of people that it can be difficult to extract abstract definitions which capture these ideas independently of human activity. "Intelligence" is a term sufficiently impersonal that non-anthropocentric definition (and application) seems possible. But other terms, perhaps because they involve a wider spread of human activity, may sit much less comfortably when applied to non-human entities: "consciousness," "greed," and "love" are just such terms. And perhaps "creativity" is rather like this too. If so, efforts towards artificial creativity will be in vain.

Or, rather, any model which is less than a complete model of mind will be seen to be inadequate as a model of creativity. And even a complete model of

mind might not be illuminating in just the way we require. If we were to gain a complete neurophysiological understanding of the brain, there would still be many questions about the mind unanswered. And, of course, it is possible that the sorts of illuminating explanations for psychological phenomena that we seek may simply not be available. The best explanation for, say, consciousness, might involve no more than pointing at the neurophysiological architecture of the brain.

Unfortunately, because we have such a dim understanding of the mind in general, it is hard to know, in any particular case, whether or not abstracting out a particular psychological attribute is a reasonable thing to do. But if we proceed with the assumption that it is, then at least we have the chance of finding out that we were right. If the assumption proves to be wrong, we shall still have gained some insights on the way.

The hope is that the power humans have to be creative is somewhat like the power they have to do mathematics. Working out how our brains actually do mathematics problems is bound to be very difficult, but the activity of doing mathematics can be described in an abstract way. Once we have the abstract description, we can understand how we can describe what a person does and what a computer does, when they add two numbers together, as doing the same thing. To build a machine to do addition we do not have to build a machine that simulates the way the brain does addition, just build a machine which conforms to our abstract description of what addition is.

Is it likely that such an abstract description of the creative process can be discovered? Perhaps not—creativity may involve too many aspects of the human mind all at once. But there do seem to be some sorts of processes which can definitely be excluded from the creativity race, and others which look like reasonable competitors.

A final note for this section: while philosophers may accuse AI researchers of being too generous with the term creativity, there may be well founded counter claims about the generosity of the attribution with regard to people. If a person is the generator of entities, the temptation to regard this activity as creative is much greater than in the case of a machine producing similar results. The Bach of our story, even if his theory of what made a good musical composition was exactly right, did not count as creative because he did not exercise any judgement. Indeed he was incapable of appreciating the music at all. Had he discarded many (or any) of the pieces, we might well be prepared to restore his status as a musically creative individual.

5 Some Current Systems

Johnson-Laird [14, 15] has written a number of computer programs that generate or compose music, including a model of jazz bass line improvisation. The program

works by following a set of fixed rules that encapsulate a grammar of acceptable bass lines. At each stage the fixed rules, drawing on a small working memory, indicate a set of acceptable options for the next note. A random number generator is then used to select a note from this set. This approach works well, and shows how jazz improvisation can be achieved rapidly, and without knowledge of long ranging musical structures. But what of the creativity of the system?

It should be noted that any given system can be viewed in many ways, and, depending on how it is viewed, different parts may be seen as generator and different parts as evaluator. Despite this apparent freedom, it is usually possible intelligently to extract distinct functional modules when analysing a system, especially when the modules show a certain degree of autonomy. The natural way to view the jazz composer is as a system which simply generates. Although we could view the system as a generator and evaluator, the "evaluative module" will never alter its behaviour independently of the "generative module". That is to say, even if we view the system as generator and evaluator, the evaluator module operates using rigid criteria, criteria which do not change and develop over time. If, for example, the system generates the same jazz line twice, it will not notice this; it will not get bored. For my purposes, we are only really justified in describing a module of the system as an evaluative module, if that module is somewhat de-coupled from the rest of the system. One way in which this de-coupling will be indicated is by its criteria of evaluation changing and developing over time. In the case of the jazz improviser, the real evaluation is done ahead of time by Johnson-Laird, for the evaluation lies in the preparation and revision of the set of rules. If the jazz improviser is creative in any way, the creativity belongs not just to the computer system, which is a mere generator, but to the computer system and the evaluator, i.e. Johnson-Laird *and* his program. As Johnson-Laird points out, given the speed with which the skilled performer can improvise jazz, it seems likely that this is the strategy she follows. That is she "programs" herself with a grammar, since she could not have time to undertake thorough evaluation during performance.[6]

Any evaluative criterion that is static will have limited value in a field where values are changing all the time (something which is especially true of the arts). An evaluative criterion may have a reasonably long life, and hence be of practical value, but to nest well conceptually with creativity it is necessary at least to make gestures towards the possibility of evaluative development and change. Evaluative dynamics are an issue both within a particular generative cycle, and also over a

[6] **Self-Programming:** This model of explanation is widely applicable. When we become proficient at sport, or when we learn to drive, it very much looks as if we "program" ourselves in a similar way to the jazz improviser. During these activities we need to operate, to some degree, on auto-pilot, since we do not have the cognitive resources to evaluate our performance in real time. That does not, however, undermine the role of evaluation in learning such tasks. (For a discussion of learning to become expert see [4].) Whether our choices of the options given by the grammar are entirely random is, however, another question.

set of cycles. The first demand is that even within the process of generating a particular entity the process of evaluation must be dynamic. The second demand is weaker, suggesting that over time evaluative criteria will have to change, rather as fashion changes.

To some degree, this kind of dynamic is present in Lenat's AM (Artificial Mathematician) system—a program which generates "interesting" mathematical conjectures—and also in his later EURISKO system [17]. AM has a "worth-slot" for each of the concepts it is modelling. The value of the worth-slot is constantly being increased or reduced, depending on what the program can find to do with the concept. A concept may grab the attention of the program for a short while, before being discarded as not so useful after all. This might be because the concept is no longer producing interesting conjectures, or because other concepts are performing better. The evaluative techniques used by AM, then, are quite sophisticated in an interesting way. They have the property that what is interesting today may not be interesting tomorrow, and even if it is, it may be interesting for different reasons.

AARON, Harold Cohen's picture generating system, has another interesting evaluative approach [21]. In the earlier versions of the program it would begin to draw at a random place, but as soon as it began to draw it would "look" at what it had done and take that into account in determining its next move. When a particular part of the drawing was complete, AARON would examine the picture to see if more was needed, in general by looking around to see if there was some free space. The extensive feedback, which operates on several hierarchical levels from the actual drawn line to the major items in the picture, means that the evaluative criteria are not readily captured in simple terms. The criteria, while thoroughly deterministic, are a moving target.

Programs like AM and AARON demonstrate, at least in a sketchy way, that the generation-evaluation model which I have been supporting can be successfully mimicked by a computer system. In important respects AARON and AM have a much more complex behavioural profile than Johnson-Laird's jazz line improviser, which essentially composes by rote. (Of course, this is no more than it was designed to do). Both AARON and AM are constantly evolving so that what they have already produced affects the way in which they will continue to produce in the future. (Though note that AARON's evolving evaluative response is forgotten once each picture is completed, i.e. its evaluative dynamics only take place within creative cycles, and not across them.) These systems exhibit sufficiently interesting behaviour, and achieve that behaviour in a sufficiently interesting way, for us to consider the question of whether we might reasonably call them creative.

6 Cultural and Visceral Aesthetics

The stress in this paper rests on evaluation rather than generation. Of course the sophistication of a generator is important, and for certain creative endeavours—such as writing a novel—the requirements on a generator are enormous. But relatively simple generators can produce interesting results. The artists discussed in this paper, guided by chance rather than self-conscious intent, have sometimes used quite crude generators, and have relied on their evaluative powers to filter the good from the bad. If we are aiming for artificial creativity, then our main worry concerns not what the machine can produce, but the way in which it evaluates its productions. If we find its evaluative approach sufficiently interesting, then we might move towards attributing creativity to it. But just how sophisticated does an evaluator need to be?

The answer must partly depend on the kind of entities being generated. If the entities are mathematical conjectures or scientific theories, then it may be that the evaluator can be relatively simple, and within the powers of today's machines. In these areas, evaluative criteria are relatively well defined and largely static in relation to the new entities. However, artistic evaluation is a more complex issue. Human artistic judgements may encompass many aspects of our experience, and it is unrealistic to expect current or near future machines to be able to encompass all of this. A machine that can evaluate a novel in the way a human being can is a very long way off. So far, our machines are just too ignorant and too narrow minded.

Part of the problem here is just getting enough cultural information into the machine. In comparison with human information gathering resources, the best contemporary machines are astonishingly limited. If we are to take a system's evaluations seriously, we need to be impressed that the evaluation is, to some reasonable degree, an informed judgement. To appreciate most, if not all, human art, it is necessary to be embedded in human culture, and probably a particular culture. However, there is no reason why we cannot build machines which have, in their own way, some kind of cultural understanding within which they can evaluate artistic productions. Indeed, Harold Cohen's AARON program has a certain style of drawing, and expresses certain preferences for form as it constructs its works. Work in analogy [26] shows how a machine can make links between one entity and another, and in this way deepen its understanding of what it is trying to evaluate.

With this sort of sophistication, a machine can build up some sort of evaluative map, and even develop tastes and preferences. Perhaps these preferences will arise out of a series of chance decisions, and so some machines may have bad taste and others good. More likely, because the criteria set will be very different from our own, it may only be possible to say that machines differ in their preferences. But if their judgements arise out of sufficiently rich evaluative processes then we may find ourselves ready to accept them as genuine judgements, even if their aesthetic

framework is very different from our own.

7 Towards Artificial Creativity

One of the reasons why the whole notion of a creative computer seems so outrageous is, I think, that we are still very much inclined to look at computers in too crude a way. We think of computer systems as consisting of a set of rigid rules operating over a set of changing data. The rules are static, and the way in which the data can vary is clearly marked out. Often, because we have access to a program listing, we can gain a clear idea of the rule set in operation, and the range of data which it can process.

But, while many computer programs are best understood in this way, it is not the case that all are. To understand a system in this way, we have already made some arbitrary decisions. We have decided what we are going to consider fixed, calling it the program, and we have decided what varies, calling that the data. If the program-data distinction works well for a system, it is easy to recover an efficient rule-like explanation of the system: we simply look at the program listing. When we can do this we can class the system as "implastic". Most man-made machines are implastic systems; the machines of physics and chemistry are implastic by stipulation; word-processors, databases, and compilers are implastic by design. The best way to understand these systems is by considering a corpus of fixed rules as being the main influence on their behavioural profile. Of course, their behaviour is affected by the data which they process, by their inputs. But, with implastic systems, we can keep separate the role of the program and of the data.

In contrast, there is a class of system where the program-data distinction is pathologically unclear. These systems can be called plastic.

An example is a self-modifying program, one which tracks some of its own internal states, and alters its own rules of operation as a result. Such a program can still be understood as implastic, but it is more natural to conceive of it as a series of implastic systems, changing through time. Similarly, a machine that is actively involved in the world often cannot be explained properly from an implastic perspective. This is because the various complex feedback relationships between its outputs and its inputs tend not to be clear from the program itself. They are revealed only when the system is actively running in the world.

A hallmark of an implastic system is that one can make sense of it being "reset" to some initial state. You can reset your video recorder so that it is in the same state as it was when it left the factory. You can do the same to most computers. But this idea makes no sense with a person. You cannot discard a collected lifetime of experience, nor would you want to. A plastic system can be like a person in this respect. The features which are important for understanding what it does, and why, may be more to do with its particular history than with its basic functional

design.

If you ask an implastic system to add the same two numbers a hundred times it will go through the same state transitions on each occasion. With a highly plastic system you might expect (though it need not be the case) such a request to cause a change in the state transition sequence during the trial. The system may become "bored". And if this is so, then it is clear how a fixed program-data divide will fail to describe the system's performance.

Note that there will always be some lower level of analysis, say one which falls below the level of the addition operation, in which the system's performance can be explained by a fixed program-style account; this is true of brains as well as artificial intelligence systems. So systems do not fall into the two classes: plastic and implastic. Rather, there are two perspective from which systems can be seen. All systems can be seen from the implastic perspective, but some are best understood from a plastic one.

One final example: A connectionist network, simulated as a virtual machine on a serial computer, can be looked at from both the plastic and the implastic perspective. From the point of view of the programmer who is implementing the simulator, the system is implastic, with the simulator program acting as the basic explanatory reference point. But the program user is not especially interested in that perspective on the system. Rather, she is interested in seeing the system as a connectionist network. Connectionist networks are well-behaved; their operation is governed by fixed, if often unintuitive, sets of rules. But, when a network is in the process of being trained, there are no fixed rules that govern its behaviour with respect to the domain over which it is learning. The rules, those which we find it natural to say the system follows, change from learning cycle to learning cycle.

The simulator programmer is happy to reset her simulation, and run her programs from a fixed initial state time and time again. Resetting the system does not affect its fundamental nature, it does not affect its identity. But the connectionist modeller is more cautious. If she resets the system then the machine which she had created through training is literally destroyed. She can only recover her lost work by retraining the system from scratch.

It is not appropriate to delve deep into these issues here. But I think that we are relatively naive at exploiting the plastic perspective with respect to artificial systems, and too readily slip into the assumption that there is only the implastic view. This may account for the reluctance, in some quarters, to accept the workings of the computer as a useful metaphor for the workings of brains, the machines which realize minds. By thinking about computers in a more flexible way, and especially in stressing their ability to be highly sensitive and adaptive, we can help our intuitions warm to the idea of artificial creativity.

8 Conclusions

Machines are capable of generating novel and interesting entities, but this activity cannot be counted as creative unless we are confident that the machine is exercising some sort of judgement. If we are to move towards artificial creativity then we need to concentrate on developing evaluative systems. A key feature of such systems is that they must be dynamic, i.e. their evaluative criteria must change and develop. Even if we can develop systems with sufficiently interesting evaluative approaches it will be difficult for us to think of them as creative. But if such machines do not meet the criteria set out here, then it will be quite impossible for us to think of them in that way.

Acknowledgement

This work was made possible by SERC grant, award ref. 91313730. Earlier versions of this paper were presented at an International Symposium on Creativity and Cognition, Loughborough, April 1994; The University of Brighton; and The University of Sussex. I am grateful to my audiences on all those occasions. Thanks should also go to Andy Clark, Ron Chrisley, Ronald Lemmen, John Cupitt, and especially Margaret Boden.

9 Glossary

artificial intelligence The science of constructing (or attempting to construct) machines that exhibit intelligent behaviour. This is usually achieved by carefully designed computer programs. More recent approaches, including connectionism and genetic algorithms, allow the environment to play a part in the design.

connectionist network A type of computer system whose design is very broadly inspired by neural network architecture of the brain. Connectionist networks consists of a set of nodes linked by variable strength weights. Systems are "trained" to produce certain output patterns in response to certain input patterns. Having been "trained" in this way, the system is then capable of generalising its pattern recognizing ability to new cases. Unlike traditional computer programs, connectionist networks also respond well to noisy or corrupted inputs.

evaluative dynamics The changing and developing of evaluative criteria over time. Too much flexibility in your evaluative dynamics makes you a fashion victim, too little makes you conservative. A system which lacks any evaluative dynamics, it is claimed here, cannot be creative.

genetic algorithms A programming technique in which computer programs evolve, rather than being designed by hand. This technique is powerful, and can come up with quite unexpected solutions to problems.

implastic perspective A way of looking at computer systems which treats them as following a fixed set of rules operating over a changing set of data. In implastic explanation the rules are the dominant explanatory base. All systems, even human beings, can be viewed from an implastic perspective, but some are more usefully seen from the plastic perspective.

plastic perspective A way of looking at computer systems which acknowledges that sometimes a fixed set of rules is an inadequate way of describing the system's behaviour. While it is acknowledged that such explanations are always available, the plastic perspective rejects such explanation in favour of a less rule bound understanding.

self-modifying program A program that alters some of its own rules in the course of its operation.

virtual machine When a computer system is programmed to behave as if it were another computer system, something which is easily done, the resulting system is described as a virtual machine.

References and Discography

[1] Margaret A. Boden. *The Creative Mind: Myths and Mechanisms.* Weidenfeld and Nicolson, London, 1990.

[2] J. L. Borges. Pierre Menard, author of Don Quixote. In D. A. Yates and J. E. Irby, editors, *Labyrinths: Selected Stories and Other Writings.* Penguin, Harmondsworth, Middlesex, 1970.

[3] Harold Budd and Brian Eno. *Ambient 2: The Plateaux of Mirror.* Editions EG EGS 202, 1980.

[4] H. L. Dreyfus, S. E. Dreyfus, and T. Athanasiou. *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer.* Free Press, New York, 1986.

[5] Brian Eno. *Here Come The Warm Jets.* EG Records ENO 1, 1973.

[6] Brian Eno. *Taking Tiger Mountain (by strategy).* EG Records ENO 2, 1974.

[7] Brian Eno. *Discreet Music.* Editions EG/Obscure EGS 303, 1975.

[8] Brian Eno. *Ambient 1: Music For Airports.* Editions EG EGS 201, 1979.

[9] Brian Eno. *Ambient 4: On Land.* Editions EG EGED 20, 1982.

[10] Brian Eno and Russell Mills. *More Dark Than Shark.* Faber and Faber, London, 1986. With commentaries by Rick Poyner.

[11] Brian Eno and Peter Schmidt. *Oblique Strategies.* Opal, London, 1975. (Limited edition boxed set of cards.).

[12] Robert Fripp and Brian Eno. *No Pussyfooting.* Editions EG EGS 102, 1973.

[13] Robert Fripp and Brian Eno. *Evening Star.* Editions EG EGS 103, 1975.

[14] Philip N. Johnson-Laird. Reasoning, imagining, and creating. *Bulletin of The British Psychological Society*, 40:121–129, 1987.

[15] Philip N. Johnson-Laird. Jazz improvisation: A theory at the computational level. In P. Howell, R. West, and I. Cross, editors, *Representing Musical Structure*, pages 291–325. Academic Press, London, 1991.

[16] H Koning and J Eizenberg. The Language of the Prairies: Frank Lloyd Wright's Prairie Houses. *Environment and Planning B*, 8:295–323, 1981.

[17] Douglas Lenat. The role of heuristics in learning by discovery: Three case studies. In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach*, pages 243–306. Tioga Publishing Company, California, 1983.

[18] F. Lerdahl and R. S. Jackendoff. *A Generative Theory of Tonal Music.* MIT Press, Cambridge, MA, 1983.

[19] Da Liu. *I Ching Coin Prediction.* Routledge and Kegan Paul, London, 1975.

[20] H. C. Longuet-Higgins. *Mental Processes: Studies in Cognitive Science.* MIT Press, Cambridge, MA, 1987.

[21] Pamela McCorduck. *Aaron's Code: Meta-Art, Artificial Intelligence, and the Work of Harold Cohen.* W. H. Freeman and Company, New York, 1991.

[22] Hans Richter. *DADA: art and anti-art.* Oxford University Press, New York, 1965.

[23] Roxy Music. *Roxy Music.* Atco SD 36-133, 1972.

[24] Roxy Music. *For Your Pleasure.* Atco SD 36-134, 1973.

[25] Eric Tamm. *Brian Eno: His Music and the Vertical Colour of Sound.* Faber and Faber, London, 1989.

[26] Stella Vosniadou and Andrew Ortony. *Similarity and Analogical Reasoning.* Cambridge University Press, Cambridge, 1989.