Colouring Without Seeing: a Problem in Machine Creativity

Professor Harold Cohen
Department of Visual Arts
University of California at San Diego

I have a very creative family; a fourteen-year-old who draws with unusual skill and has been playing the piano since she was six, a twelve year old who is equally at home in drawing, painting, sculpture and mathematics, and Zana, my three year-old daughter, who speaks two languages and has been drawing like a five-year-old since she was two-and-a-half. The mother of these three is a distinguished Japanese poet and writer. As for the only male in this feminist stronghold, I've been making my life and my living in art for fifty years, for about thirty of which I've been writing computer programs of sufficient ... something ... to explain why I'm writing this paper. Which brings me to the last member of the household. When I don't keep it busy with email, designing machines, writing conference papers and doing my tax returns, AARON is sitting quietly on my desk, generating original images at the rate of about one every two minutes (fig I).

And, most particularly when I'm watching it, I am aware of a couple of questions that need to be addressed: if I say that I have a creative family, and then I were to say that I have a creative computer program, would I mean the same thing by the word "creative"? And how far could I justify the claim that my computer program - or any other computer program - is, in fact, creative?

I'd try to address those questions if I knew what the word "creative" meant: or if I thought I knew what anyone else meant by it.

Back in the days of flower-power, the streets of Berkeley, California were lined with romantically attired flower-children selling tie-dye scarves and badly made silver jewelry, resulting, we were led to suppose, from the expression of their creativity. There's a lady in the United States who has built a media empire on showing people how to be creative by putting wreaths of rosebuds on their front doors and painting their dining rooms pink and yellow. Every week my mail offers me software which is guaranteed, for twenty-five pounds, or fifty, or a hundred, to increase my personal creativity. What can it possibly mean? Where does intelligence end and this, presumably higher, faculty begin?

For reasons that I hope to make clear, I have never, in fact, claimed that AARON is a creative program, and while we think all our kids are pretty smart, and the three-year old will probably get her PhD by the time she's six, we'd never use a word like "creative" to describe what all little children do if they 're given enough room to develop. "Creative" is a word I do my very best never to use if it can be avoided.

Having said all of which, I have to confess that when I think about the way Bonnard used colour, or the invention of anything as bizarre as superstring theory, or Faulkner's four ground-breaking novels written in the space of two years, or the astonishing prolificity of a Bach or a Mozart, I do find it hard to avoid the impression that there is indeed some "behaviour X" that appears to be distinct from intelligence, whether we choose to call it creativity or not, and whether I can find the line that separates it from intelligence or not.

So, if I can't talk about creativity, and before I can question what it might mean to say that a computer program exhibits, or might eventually be able to exhibit, creativity, I can at least try to identify the essential elements of behaviour X in human beings.

Well: any behaviour in mature human beings is pretty complex, and I don't have the knowledge or the professional skills that would be required to sort out the mess of genetic factors, environmental factors, personal experiences, ambitions and heaven knows what else that cause people to do the ordinary things they do, much less the extraordinary things. Fortunately I have access to a rather new human being (fig 2), in whom the determinants to behaviour are relatively uncluttered by unresolved experience. So I'd like to begin by looking to see whether part of her behaviour doesn't have many of the characteristics, albeit manifested on a very small scale, that we would hope to find in fully developed adult behaviour X. I'm going to look specifically at her drawing behaviour; which isn't meant to imply that drawing is by definition an example of behaviour X, but simply that behaviour X has to be manifested in activity, and drawing is one of those activities in which we can see, not only what has been done, but - if the drawer is a child and if we watch and listen carefully - track what happened while it was being done.

Early drawing usually proceeds out of the purely motor activity we call scribbling. At some point a round-and-round scribble will migrate outwards and become an enclosing form, and at some time rather later, the scribble will be omitted, leaving closed forms that are then available for the building of a variety of representational elements, (fig 3), made by the three-year old child of a friend, illustrates these stages quite clearly and is characteristic of what we would expect of a child of that age. The ability to distinguish between closed forms and open forms is one of the earliest cognitive skills to develop in the young child, so its appearance as a factor in drawing presumably waits upon the development of adequate motor skills.

By the time Zana was two she was constantly asking me to draw the people in her family. I obliged, somewhat reluctantly, with cartoonlike drawings which I couldn't show you - without ruining my reputation. Zana bypassed the scribbling stage entirely, and by the time she was about two-and-a-half she was drawing faces *(a, b, c)*. These were clearly done in imitation of the drawings I made for her, but while I would repeat the same drawing of each person over and over, her own drawings, many of them of me, showed a quite surprising degree of variation in how they were put together, reflecting an increasingly important role for own perceptions. One day she was busily drawing me when she stopped, stared at me very hard for a minute, then she stuffed two fingers into her own nostrils and said "you got two noses!". What a discovery! She then proceeded to draw in the nose for the very first time *(d)*. But what a strange configuration for a nose! Strange, that is, until one starts actually to examine the complicated formal structures around the nostrils, and then it seems that Zana got it very nearly right.

a

b

c

d

Here's another example of an adopted practice modified by personal experience. Zana developed a fondness for drawing her hand by putting her hand on the paper and tracing around it. She didn't invent the strategy; children at her school learn the simple geometric shapes by tracing around plywood cutouts, and I'm told that at Thanksgiving in the US every child is shown how to make a drawing of a turkey from the outline of a hand.

In this case, then, it is a drawing strategy, rather than a form, that has been adapted from a different context. Two things are noteworthy in these drawings, one having to do with how she made them and the other concerning the drawings themselves *(e)*. The first, which illustrates the cognitive importance of closed forms in early development, is that Zana would always close the outline when she took her hand away and before she continued with the drawing. The second is the appearance of fingernails and joint-creases in the fingers. I know she didn't get that from anyone in the family, whose suggestions were always politely ignored, and it almost certainly didn't come from her school, since all of the other children of her age were still at the scribbling stage.

In fact, it is clear from the subject matter of these drawings that they reflect her own personal experience. Zana won't walk upstairs if she sees an ant on one of the steps, but she likes to pick up baby snails known to Japanese children as den-den-moshi-moshi, and let then crawl on her palm (*f*). Again, no one to my knowledge ever showed her how to draw a snail, and if they had. I'm sure they wouldn't have drawn them with legs. Some of Zana's snails have legs! This is not a mistake, you understand, though you may regard it as a failure to observe correctly, but clear evidence that when we draw, adults and children alike, we represent our internal models of the outside world, not the outside world "itself" - whatever that might mean. In Zana's experience, fairly recent for a three-year-old, you need legs for walking, and den-den-moshi-moshi unquestionably walked across her hand. I asked her about the legs, and she assured me that snails really do have legs.

e   f

Adaptation - that is, the recycling of forms and production technologies into contexts other than those in which the forms and technologies originated, is clearly characteristic of the young child's development, as it is in any manifestation of the adult's behaviour X. Thus, for example, the short straight line which began as the folds in hands began to make their appearance in other contexts, *(g, h, i)* A long line attached to a closed shape became a lollipop. Short lines radiating from a closed form became either a sun - I'm sure she was shown that one - or a lion. She had seen pictures of lions in one of her books, though, of course, not in a rendering that could have served as a model for her own.
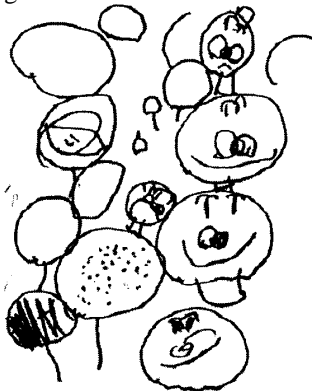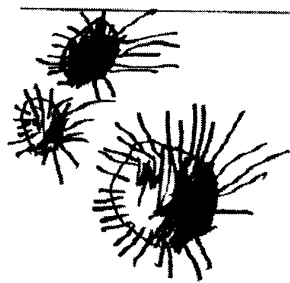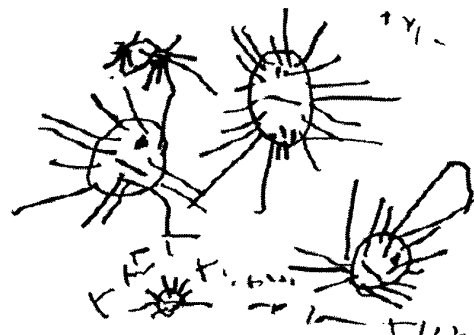
g   h   i

Books are an important part of Zana's life. Mummy reads the Japanese books and Daddy reads the English books. Zana knows all her English letters, and she can recognize the difference between English and Japanese texts. A couple of months before she was three she started to introduce "texts" into her drawings (*j*), which she would then ask me to read to her. None of that is very surprising if one reflects that pictures in childrens' books are always accompanied by texts, and parents are always able to read them.

Figure 1
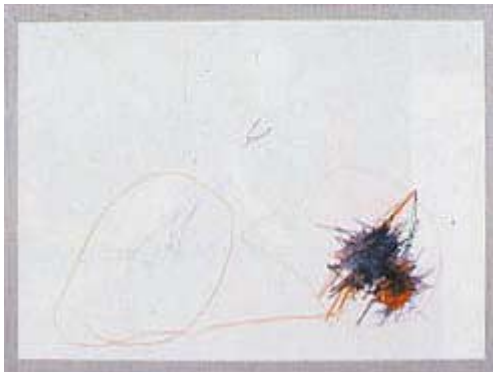

Figure 2
Zana drawing.

Figure 3
Drawing by Noah Macilwaine.

Figure 4
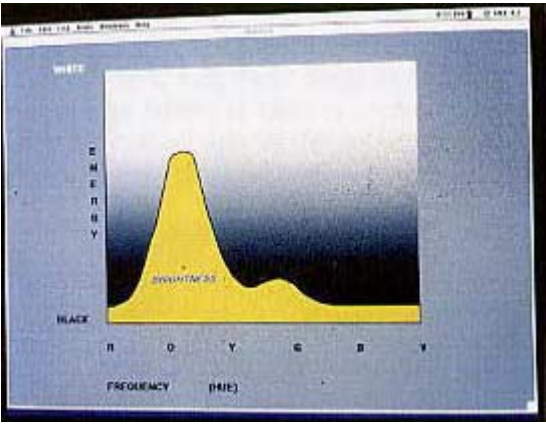Painting by Zana.

Figure 5

Figure 6

Figure 7
Screen Image 1992.

Figure 8
Painting from fig. 9.
Oil on Canvas 1993.

Figure 9
Screen Image 1993.

Figure 10
Machine painting.
Dyes on paper, 1995.
Collection: Dr. Thomas Gruber.

Figure 11
Machine painting.
Dyes on paper, 1995.
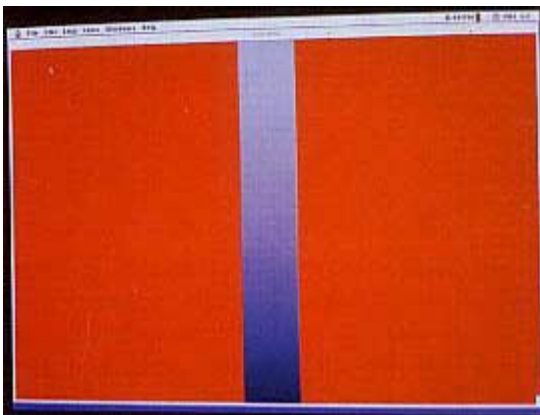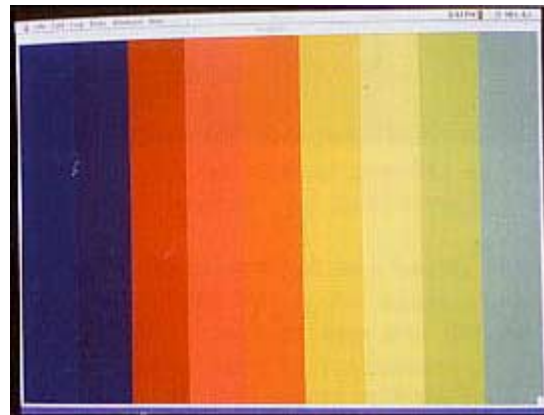

Figure 12
"Mother and Daughter"
Oil on canvas, 1998.

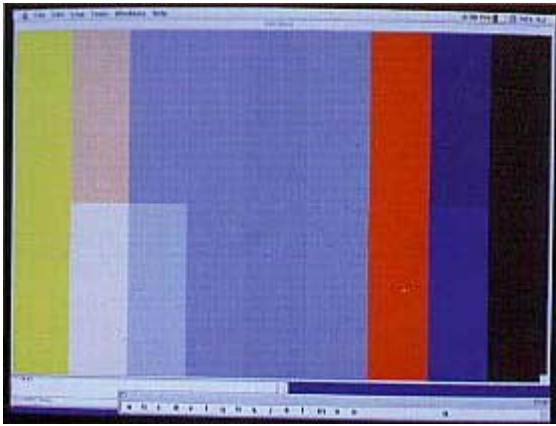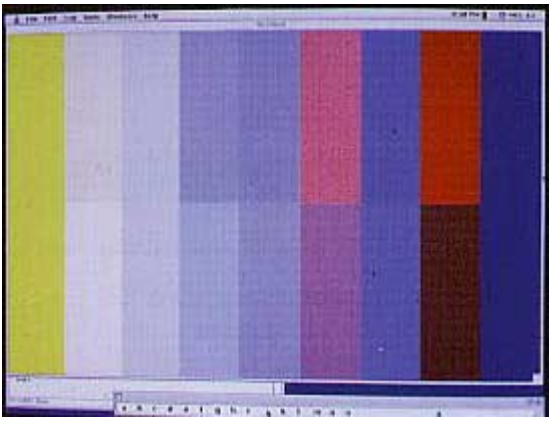Figure 13

Figure 14

Figure 15
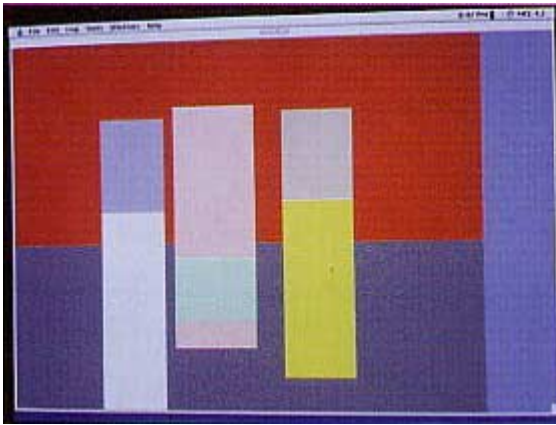

Figure 16

Figure 17

Figure 18

j

Now if Zana had been a mature adult artist and not a three-year-old, she would perhaps have noticed that her preoccupation with text-and-image had generated what was for her a quite new kind of composition quite different from the centralized faces. Of course, she wasn't at all interested in composition, but what of that? Every artist has known the experience of finding something happening in his work that he hadn't intended to happen, but which nevertheless causes a change in direction. As earnest observers, we would say that Zana's new kind of composition emerged, unbidden and unintended, from her preoccupation with text. And Zana, who probably couldn't care less about emergent properties, might be on her way to New York with a show of new paintings.

And that observation leads us directly to what I am sure is a critical element of behaviour X. The individual has to find something in his work that he never consciously put there. Whether these emergent properties result from the juxtaposition of disparate conceptual elements, or whether they result from the technological complexity of the mode of production, the individual must find properties in the work that were neither consciously sought for nor intended, but which may nevertheless lead subsequently to new conceptual structures, new ways of using the technology, or even suggest a modification of existing, consciously-held, goals.

At three, Zana has neither the adult's rich conceptual framework, nor the mature artist's formidable array of technological know-how. As I've tried to show, properties emerge in her drawing in part in an attempt to incorporate her own perceptions of the world, but principally through the adaptation of simple strategies to satisfy new purposes. Also, I should add, from her sensitivity to the physical properties of the materials she's been given. Like any other child, Zana will behave quite differently if she's given a brush and a pot of paint to the way she behaves with a pencil or a coloured marker (fig 4).

So I will argue that emergence is a necessary component of behaviour X, but not that it is a sufficient component. There is the further requirement that becomes clear when we consider the difference between a three-year old and a mature adult. Zana hasn't *noticed* that she has made a new kind of composition. For behaviour X to manifest itself for the mature artist, the individual must also notice that something has emerged, and be prepared to act upon what that something suggests.

Behaviour X is thus not manifested in a single unexpected outcome but rather in a capacity for continuous self-modification.

To summarise, then, we have three elements, all of which have to be present in the performance of some activity:

First, emergence, which implies a technology complex enough to guarantee it. And I want to stress that while for present purposes I am using terms like "technologies" and "materials" in relation to the physical materials and production strategies of the plastic arts, I mean them in the broadest sense to include language and thought itself.

Second, awareness of what has emerged. And I have to put aside what necessarily would be a lengthy discussion on the nature of awareness. If by awareness we mean conscious awareness, then the truth is that mature artists are often not aware of what emerges in their work, and consequently they often don't know why they act. But that is to be expected; only a very tiny part of all the data presented to the senses is ever advanced to consciousness — which isn't to say that it disappears without a trace — and there is ample evidence that most of what goes on in the brain, including decision-making, is not conscious at all. The second element is awareness, then, but not necessarily conscious awareness, of what has emerged.

Third — without which, in fact, we could hardly know that awareness had occurred — a willingness to act upon the implications of what has emerged; and this surely implies a distinctive psychology and motivation in the individual.

And to these three I'll add a fourth, which hasn't been mentioned so far because my three-year old mini-subject has so little of it: knowledge. Having a great deal of knowledge about one's chosen technologies doesn't guarantee that behaviour X will kick in for the individual - consider how knowledgeable many thoroughly academic artists are — but behaviour X will not kick in without it. My behaviour X is not to be equated with, or confused with, either innocence or ignorance.

How much of this - and I have no doubt there is more - can we expect to find manifested in a computer program? Let me begin with a short and distinctly provisional answer concerning these four elements.

On the first: I can see no problem about the increasing complexity of computer programs other than the programmer's ability to keep track of what the programs are doing. On the third: of course a program can be written that will act upon anything we want, provided we can put into code what that is. On the fourth: I can see no reason in principle why a computer program should not be given an arbitrarily large body of knowledge about a particular technology, even a technology as complex as painting. As to the second element, the program's awareness of properties that emerge, unbidden and unanticipated, from its actions... well, that's a problem.

But of course it's not the only problem. It should be clear from the way I've answered the question that the answers I give for the computer program *don't* mean just what they would mean if I were talking about a human being. It may be true that we can give a program an arbitrarily large body of knowledge about physical properties, of oil paint for example, but it can't be the *kind* of knowledge I acquire viscerally as I mix the paint with a palette knife. It may be true that the program can be written to act upon anything the programmer wants, but surely that's not the same as the individual human acting upon what he wants himself. Isn't free will of the essence when we're talking about the appearance of behaviour X in people?

Perhaps: if we assume that free will emerges, when it emerges at all, as a rather unusual property of the nervous system; in human beings, that is. What else could it emerge from: in human beings, that is? I don't doubt the existence of free will, but I do reject the assumption of exclusivity, and my provisional answer to the question - how much of my four elements of behaviour X can we expect to find in a computer program - was intended precisely to focus attention upon the fact that a computer program is not a person. AARON is an entity, not a person; and its unmistakable artistic style is a product of its entitality, if I may coin a term, not its personality.

In any computer program that purports to parallel human behaviour there is a line, above which the claim that the program is doing what human beings do can be verified in straightforwardly functional terms: either the program produces equivalent results or it doesn't. Below the line we're dealing with implementation, and we can rarely claim that the program does things in the same way that human beings do them. In functional terms AARON does what human artists do: it paints pictures. But if, as I've suggested, behaviour X is manifested in a capacity for continuous self-modification rather than in the resultant objects, then we need to look for it below the line: largely in the non-conscious processes of the human being on the one hand, and, on the other, in the program structures which can be devised for the machine's intrinsic and distinctly nonhuman capabilities.

With some notion, then, of what we should be looking for, and something about the terms in which we should be looking, I'd like to turn to particulars. AARON began as a drawing program almost thirty years ago, and it has evolved, thanks to my own continuing involvement, from something resembling late Paleolithic cave painting to figurative painting: portraits, as it were, of "imagined" people (fig 5). For most of its history I've used some part of AARON's output to make paintings for which I supplied the colouring. The possibility of having AARON do its own colouring grew for me over the years from a casual curiosity about what would be possible and what wouldn't, to an unavoidable imperative, which took several years from inception to its current partial satisfaction. Since AARON as a whole is rather too complex to see how it shapes up with respect to behaviour X - it is currently about a megabyte and a quarter of LISP code - I'd like instead to examine the implementation of this most recently developed part of the program - colour - in some detail.

How does one write a program to perform as an expert with respect to colour?

Well, not the way the human colourist does it, obviously. The human colourist has an extremely refined visual system, and whatever theories he may come up with, he is unavoidably dependant upon visual feedback. The fact that the computer has no such system doesn't mean that colour expertise is impossible for a program, it means simply that one has to build a system on the capabilities the program has. And which capabilities, we should remember, the human colourist doesn't have. No human artist could compose an entire colour scheme in his head, with his eyes closed, and then write down the mixing instructions for someone else to follow; which is, roughly speaking, what the program needs to be able to do.

But one has to build a system to do what, exactly? We can't expect to get very far in modeling expertise in any area

unless we have some notion about what the expertise entails. Unfortunately, and for very familiar reasons, very few experts know consciously what their own expertise entails. As a painter. I'm just your average expert in this regard. I've never subscribed to any particular theory of colour, and I've never been one of those whose colouring procedures have been either systematic or consciously manipulated, certainly not to the point where, like Albers or Seurat, I could have formulated a theory of my own.

Colouring for me has always involved a lot of sitting and staring; a little thinking about structural considerations that might influence how to make choices, but in the main not consciously deciding what to do next. Sometimes I might find a colour name popping into my head - usually the name of a colour I haven't used in a while - and when that happens it may act as a trigger to action. But that's the exception; for the most part I don't know what causes me eventually to stop staring and get to work. I go to my paint table, squeeze some amount of paint from each of two or three tubes, mix them up, and recognize that the result is the colour I want. Evidently my non-consciousness has been pretty busy while I've been staring at the painting; it has not only provided the colour that I want, but also the program for generating it.

("Hold on," says the skeptic; "how do you know it's what you want? How do you know you're not fooling yourself into believing you want what you get?" That's obvious, isn't it? What I want relates to the needs of the painting. Do you imagine those needs can be satisfied by squeezing random amounts of paint from randomly selected paint-tubes?) For those of you whose experience of colour is limited to the computer display, colour is an abstraction with components like wavelength, amplitude and phosphorescence. The painter doesn't paint with abstractions, he paints with paint; and any painter will recognize that my non-conscious deliberations must rest upon an extensive body of knowledge; about colour-as-abstraction, certainly, but also about the stuff I squeeze from the various tubes, which gets its colour from widely differing physical materials, so that no two of them have quite the same physical characteristics. It would take a tube-full of cerulean blue to have the same colouring effect on a mixture as a thimble-full of ultramarine, for example, and nothing will give alizarin crimson the same opacity as any of the iron oxide colours. So knowing just how much of each colour to squeeze to make a mixture must rest upon a great deal of experience of mixing and using these particular physical materials.

I'm almost inclined to think there's a difference between knowledge that has been internalized to the point where the expert no longer knows he has it, and knowledge that was not consciously acquired in the first place, so that the expert never knew he had it. If that is the case, then expertise in colour is surely of the second kind. What I do know is that it took me a painfully long time after I started to fuss about having AARON do its own colouring before I was able to see any light at all, before I was able to isolate a single element in my own knowledge. And I know that this was a piece of internalized, rather than intrinsically non-conscious, knowledge, because it was something I'd been passing on to my painting students for most of my adult life.

It was that the most important single element in colouring is not colour at all, but brightness.

Let me explain that in a little more detail. I referred earlier to colour-as-abstraction, as opposed to colour as a property of physical material. The primary abstractions, whether we're talking about computer displays or paint, are hue, brightness and purity. The curve in the graph shown in Figure 6 plots the energy of the component frequencies of a colour sample, so hue refers to where the sample falls on a spectrum - the horizontal axis -. brightness is the total energy and is thus represented by the area under the curve; and purity refers to the bandpass characteristics of the sample - that is to say, how much of the available energy falls 'within how narrow a spectral band. I'm using a simple physical model here so that I can attach clear meanings to the terms I use, because these three properties go under different names in different colour theories. Words like brilliance, intensity and saturation are often used, either instead of what I've called purity, or for some hard to measure combination of brightness and purity. And I'm using the word "brightness" as a more expressive term for what is called "tone" in the UK and "value" in the US; independent of its hue and its purity, how light or dark is the sample?

It isn't hard to see *why* brightness is more important in controlling a complex colour scheme than hue is. We have colour vision, yet the eye functions predominantly as a brightness discriminator. Consequently, and as anyone who has looked at a photographic negative knows, it's very hard to read an image when the brightness structure is distorted, while we have no difficulty with colour reproductions in which the colour bears little relationship to the original. Once when I was teaching a course on colour at the Slade School, I had the students take turns in repainting parts of one of their paintings, using arbitrary colours, but with the single constraint that the replacement had to be of the same brightness as the colour being replaced. And after a couple of hours work everyone except the student whose painting we used - and whose soul was a little the worse for wear — agreed that nothing significant had changed very much.

Based upon this single scrap of expert knowledge, it didn't take long before AARON was able to generate simple but quite acceptable colour schemes (fig 7). By the end of a few weeks I was actually using AARON's colours as well as its drawings to make paintings (figs 8&9), and the days of staring and head-scratching were over, at least

temporarily. If I wanted to know what colour to paint something, I just looked at a slide of how AARON had done it. That marked the beginning of a rather difficult period for me, however, because my goal had always been to have AARON exercise colour expertise in the real world, with real materials, not simply on a computer screen.

There's a small industry now devoted to resolving the fundamental difference between colour mixing on a screen and mixing with physical material. Everyone who uses Photoshop or one of its cousins wants to be sure that what finally gets printed on paper is just what was done on the screen, My own problem is the reverse: I want to be sure that what I see on the screen represents what I can get on the paper. But in either case the translation from one to the other is not so much extremely difficult as literally impossible; the best one can hope to do is to produce a convincing equivalent.

The cathode ray tube display uses only three primary colours - red, green and blue - and they mix additively, which is to say that each new colour component added to the mix increases the energy - and hence the brightness - of the mixture. Physical materials - paint, or dyes - act as filters, and each new added component reduces the energy of the mix. The classic demonstration of the difference between additive and subtractive mixing, which always seems to astonish people who have never seen it before, is that if we were to put a green filter in one projector and a red one in another, we would get yellow on the screen where the two projections overlapped. Almost the same is true with any electronic display - a computer display or a television set - except that the tiny spots of red and green mix in the eye rather than on the screen itself. Based only on experience of physical materials, one would correctly anticipate that the pigment we call red would filter out everything except the red light, the green pigment would filter out everything except the green light, and we'd be left with a sort of dirty brown. Which is just what happens.

The best kept secret of this little party trick is that actually the two methods of mixing produce the same *hue;* it is just that the one produces an energy-enhanced sample while the other produces an energy-degraded sample. And that points to the major difficulty in translating from one to the other, which is predicting the brightness of the result.

While I was doing the initial work on colour on the screen I was also building AARON's first painting machine and selecting a palette of the dyes I proposed to have it use. I made about twelve hundred samples of mixtures of the dyes, borrowed a rather expensive gadget from my university's optical department, and measured the samples for their red, green and blue components and for their reflected brightness under standard lighting. Lots of data; but when the painting machine was finally finished and I tried to translate the mixing rules I had generated for the red, green and blue components on the screen into the red, green and blue components that identified the dyes, it became clear that the two had very little in common.

Even worse, and because I had not seen my way clear to making a dozen or more dilution samples for each of my original twelve hundred mixtures, I really had no reliable data on dilution and thus no adequate control over brightness. The paintings generated during an exhibition at Boston's Computer Museum (figs 10 & 11) were consequently quite varied in quality - when they were good they were very good, but when they were bad they were ... well, not horrid exactly, but not what I'd hoped for. Eventually it occurred to me that rather than making a large set of dilution samples for each colour, I need only determine what dilution would be required to produce the lightest version that was still identifiably the same colour. If the dilution-brightness function proved to be acceptably linear, which it did, then AARON could calculate the dilution for any required brightness of any colour.

I was able also to establish a coherent and reliable mapping between what I saw on the screen and what I would see on the paper. I simply scanned in the dye samples and let the Macintosh software tell me what it thought the component values were. The result of these two strategies is that in making colour specifications, AARON is limited now to the samples it scanned, with appropriate dilution, and I can have some confidence that the colours it uses on the screen are actually mixable from the dyes.

I've identified brightness control as the most important key to colouring, regardless of whether the implementation is done by the human colourist or by the program and we might pause for a moment to consider how each goes about his —or its— business. As I noted, the human colourist relies heavily upon a sophisticated visual system. The possibility it affords of continuous feedback from the work in progress encourages the development of *ad-hoc* strategies so effective that they become essentially the norm; it's comparatively rare for the painter concerned with colour to "get it right" first time. For a program without a visual system, on the other hand, visual feedback isn't an option, and consequently the program has to work from a *priori* plans robust enough to produce satisfactory results without subsequent modification, no matter what chance juxtapositions arise in the course of their implementation.

In short, the program can't do what the human colourist does, and the human colourist can't do what the program does; below the line, that is. But even though they are doing it in entirely different ways, they are, in fact, doing the same thing - that is, controlling the brightness structure of the image.

Brightness control is a key factor in colouring and, as we've seen, it was alone enough to make AARON a competent colourist; not enough, however, to sustain an expert level of performance. Something else was needed, presumably involving the other two components, hue and purity. That something else was not dug out from my internalized expert knowledge; in fact, I doubt it would have occurred to me at all had I not been engaged in this work. It is that while one's overall response to an image - one's emotional response, if you will - is substantially determined by the inter-relationships and the juxtapositions of masses of colour, the legibility of the image is determined primarily by what happens at the edges of forms.

It is very clear, of course, that we don't *need* colour to build a legible image; brightness control alone will do that, as any black-and-white photograph will demonstrate. But if we want to use colour, for its ability to extend the range of the responses it can evoke in the viewer, for example, then the use of colour has to be subject to a complex set of constraints, most notably by the need to retain adequate separation between forms. And we have not just brightness alone, but also hue and purity at our disposal in maintaining adequate separation.

We can see now why it is so difficult for the painter to get the colours right first time, why he needs feedback from the image in progress to adjust the hue, the brightness and the purity of the colour areas until the proper separation has been achieved. And finally, as often as not, making use of localized darkening around the edges of forms — a kind of simulated Mach-band effect — to get what he was unable to get from the adjusted colour areas alone (fig 12). AARON gets no feedback, is unable to make adjustments, and at the present time is unable to handle localized darkening. An ideal dead-reckoning strategy would involve a combining function to generate suitable simultaneous values for the three components, but it doesn't look as though such a combining function can exist. We simply don't know how to weight the three components in arriving at an overall score, and it's pretty clear that weighting couldn't be constant across the spectrum. We might expect, for example, that increasing the separation on any of the three axes would result in increased overall separation. But we find, in fact, that increasing the difference in hue between two colours only increases legibility if there is adequate separation in brightness. If the two colours are adjusted to near-equal brightness, the eye has more difficulty sorting out where the edge is as the hue difference increases, not less (fig 13).

As a first approximation, then - and I should say that I've not yet moved on to a better one - I decided to have the program provide adequate separation on each of the three axes independently. This is the sort of thing a program should be able to do in one go, if at all, while the human being can only do it by continuous adjustment.

Let me explain how the program's implementation proceeded.

My twelve hundred samples had been made in the first instance by mixing pairs of dyes in seven equally spaced samples from one pure dye to the other. About a third of them I considered to be too drab to be worth using and they were discarded. The remaining eight hundred or so I distributed by eye around a circle divided arbitrarily into a hundred and eighty positions. These dyes are not manufactured to cover the spectrum evenly, unfortunately; there are big gaps in the greens, for example, and no strong blues, while there are several colours in the red-to-yellow part of the spectrum. The result was that in some cases there were no samples to occupy one of the hundred and eighty positions, while in others there were anything up to six samples. These were then sorted, also by eye, according to decreasing purity, and without any regard to their undiluted brightness.

All of the data representing mixing instructions on these eight hundred samples - measured brightness, estimated purity and dilution factors - is available to the program, and it can now associate these instructions with physical locations. We might think of this as a crude three-space model, in which brightness would be an independently variable third dimension controlled by dilution. There is an important limitation here, however, which directly affected the design of the implementation, as you will see. It is, simply, that dyes are made lighter by adding water, but there's no way to make them darker than their undiluted original state.

Within this limitation, AARON is able to produce any brightness it requires of any of its eight hundred available colours. The point of the exercise is not to deal in terms of individual colours, obviously, but in terms of sets of colours that satisfy the separation criteria as I've described them-, to function as a colour chord generator, and to assign the components of its chords to particular elements within a painting.

The first pass involved using equal spacing of the chord components around the circle, with only the purest sample taken from each position. By setting the spacing differently, the program could generate a colour scheme all within a narrow part of the spectrum or spread across the entire spectrum, but obviously the relative brightness and the relative purity of the elements of the chord was quite arbitrary; it depended simply upon what particular set of samples happened to be chosen. A variation on this strategy (fig 14) used an additive series for the distances between samples rather than constant spacing with slightly more interesting results; close grouping of parts of the chord along with greater distances between other parts. But, again, purity and brightness remained uncontrolled.

In order to assert the essential control of brightness, the next step was to sort the chords by brightness, then divide the range from lightest to darkest into equal steps and try to adjust each component to the brightness required by its position in the sequence (fig 15). That worked to some extent, but of course a colour couldn't be adjusted if it was already lighter than what was required, and in that case it was simply left unchanged.

That led to the final step (fig 16). The program would scan through the lower purity levels at the position in question, and then around its neighbours, until it found a colour dark enough to be diluted to the required brightness. The result was a guaranteed minimum perceptual separation between any two components of the chord, which the program was able to demonstrate with simple diagrammatic renderings of a couple of figures in a room (fig 17).

I've left out most of the details of what remains, even with the details, a fairly simple system, one of the most interesting features of which is the astonishing proliferation of options that presented themselves at every stage of development. For example, I mentioned that in the first step it would only use the purest colours available, but obviously it could also choose to use the least pure, which then yielded entirely different effects; or to choose colours which varied as much as possible in purity, again with different results. Or it could select a starting point for the additive distribution in order to generate chords with strong groupings in particular parts of the spectrum. Or, I said that in the final stage it would look for colours that could be diluted, while actually it has the option of giving preference to brightness or to purity as the controlling factor. And finally, once the chord is applied to a representational framework, it has to deal with semantic issues. With respect to the composition as a whole, for example, the program has the option of placing dark figures against a light ground or light figures against a dark ground. And since some colour choices are mandated by subject matter - AARON will never choose to paint faces green or purple, for example - it may choose to generate a separate chord to deal exclusively with flesh tones.

At one point I estimated that AARON had no less than seventy-two distinct paths to follow, and by that time, even though I could follow the history, I could no longer keep track of how an individual image had come about. Which condition is, of course, exactly what one would look for to satisfy the requirement of emergence.

And so we come to the current state of the program (fig 18) and the need to assess its success, with respect both to the presence of behaviour X and more generally to the quality of its work.

Let me begin by reflecting that AARON is able to do what expert human beings do, and do it to a significant level of expertise, *without* the visual system upon which human beings rely and *without* the full range of experiential knowledge which they bring to bear, in this case to colouring. It is noteworthy also that the response its work is capable of evoking in the viewer appears not to be too badly constrained by the program's own lack of an emotional life. At the very least, these facts should raise interesting questions that are clearly not being addressed by those who concentrate only on whether the machine can be said to have intelligence, or consciousness, or awareness, or, in the final fallback position, a soul.

However, my goal in this paper was not simply to demonstrate AARON's prowess as an image maker, but specifically to look for evidence of behaviour X within the program. I have, in fact, demonstrated a colour technology complex enough to guarantee emergence. And, to the degree that this technological complexity rides on a body of expert knowledge, we should conclude that AARON has expert knowledge of colour and of this particular range of colouring materials. I don't doubt that it's far short of what I have myself, and it's unlikely ever to cover quite the same range, simply because it has been acquired quite differently from the way I acquired mine. Nevertheless, I think it gives reason enough to infer the possibility of much greater knowledge in the future.

Now we come to the hard part. AARON is not a prescription for picture-making; the complexity of its structure guarantees that the program itself cannot predict exactly what will happen as a painting proceeds - that is, after all, what we mean by emergence - and a high proportion of its decisions have to respond to the state of the painting at the time the decisions are made. That level of response sounds promising, but it doesn't satisfy the second criterion for behaviour X, which requires that AARON is aware of what emerges, unbidden, from the exercise of its own production technology. AARON does respond to some part of what emerges, but that's because there is a range of properties AARON is always watching out for - they define what the human artist would call the problematics of current artmaking – the things worth paying attention to. In that sense they define what AARON *is*. But AARON can't notice something *outside* that range; it couldn't notice, for example, if the foreground elements of the painting occupied exactly half the overall space, or if the colour of a background turned out to be precisely complementary to the colour of the subject's shirt. And, by the way, humans wouldn't notice those things consciously, either.

Or, more correctly, AARON *could* notice those things, but only if they were included in the range of issues it was written to watch out for. Certainly, we can anticipate that its performance will become increasingly varied and sophisticated as its range of noteworthy events increases, but the criterion for behaviour X is satisfied only if it can notice something that wasn't included. Catch 22. We can always include more, and what we want always requires something we didn't

include. But of course! I did say at the outset that behaviour X is not manifested in objects, but in the capacity for continuous self modification. And if, as I suspect, most human beings don't do either better or differently in this regard, we should probably conclude that their performance doesn't satisfy the requirements for behaviour X either.

So, finally AARON fails on the third criterion also: the ability to act upon the significance of emergent properties. AARON's biggest single limitation is the fact that it has no capacity for self-modification. And why, one may ask, after more than twenty-five years of work, is that still the case? Because I've always considered self-modification to be a pointless programming exercise unless it could be driven by the existence in the program of higher-level criteria and the program's ability to judge to what degree it was satisfying them. You must understand that I'm not talking about rules for good composition; what I mean by higher level is criteria pertaining to painting-as-verb, not to painting-as-object. After fifty years of painting I am still unable to externalize the criteria for painting-as-verb which I must obviously exercise myself, much less see how to build that capacity into a program.

So, as the author of a program capable of generating a quarter-million original, museum quality images every year into perpetuity, I will award myself an A for effort, but no cigar. I don't regard AARON as being creative; and I won't, until I see the program doing things it couldn't have done as a direct result of what I had put into it. That isn't currently possible, and I am unable to offer myself any assurances that it will be possible in the future. On the other hand I don't think I've said anything to indicate definitively that it *isn't* possible. Many of the things we see computer programs doing today would have been regarded as impossible a couple of decades ago; AARON is surely one of them.

In the final analysis we follow our dreams, not the all-too-obvious, but ultimately proven-wrong, reasons why our dreams won't work. If you want the machine to be a device for enhancing your own personal creativity, whatever that might mean, it's yours for twenty-five pounds or whatever if you believe the hype; until the next release, at least. If I want autonomy for the program; if that notion is at the core of everything I do, then perhaps the day will come when I can proclaim AARON to be the most creative program in history. Or someone else can make the claim on behalf of some other AARON. Stay tuned.