"A Self-defining Game for One Player"

Invited Talk
Loughborough Conference on Cognition and Creativity
October 99

Harold Cohen
CRCA
UCSD

I once made a joke at a cocktail party to the effect that I would be the first artist in history to have a posthumous exhibition of new work. I should be more careful what I say at cocktail parties. The joke has been quoted – though not, need I say, actually discussed --much more frequently than anything serious I've ever said.

So I thought that perhaps I might take this opportunity to see what's behind the joke. Of course I was referring to the fact that my computer program, AARON, is currently capable of generating about a quarter-million unique, original images every year from now to eternity. Computing power increases over time, so pretty soon AARON could be providing several new, original paintings a year for everyone on the planet if it were, in fact, generating all those images on paper, which fortunately it isn't. It uses a mechanical painting machine to generate output in the real world, and I don't see its real world output ever getting much beyond one large painting per day. A good thing too.

 I said "New, original images." I did not say "creative." I use the word "creative," on those rare occasions when I use it at all, to refer to the ability of the individual – human right now, program potentially – to move forward, to develop, to introduce new material, whatever those rather imprecise terms mean. Which is to say, more precisely, that I believe the word properly attaches to continuous change, not to single events. There's no question that AARON has moved forward and developed over its thirty year existence, but the agency of change and development has been me, not AARON. Unless it can pick up from where I leave off, developing new knowledge and new levels of capability for itself, it will go on generating images that are original and different one from the other in the sense that two faces in the human population are different from each other, but all the same in the sense that there will have been no further development, no new material introduced.

Which leaves the question open, of course, as to whether we will ever be able to claim that a computer program is creative. What would be involved in giving a program that capacity? What is creativity actually like?

If we survey the work of any major artist, we get the distinct impression that he knew exactly what he was doing and knew exactly where he was going. Mozart always sounds like Mozart. Matisse didn't produce a Picasso on Monday and a Miro on Thursday, he produced Matisses every day. And while we could anticipate that whatever he did would be unmistakable as a Matisse, we could never predict by looking at what he did one day what he would do the next day or the next month or the next year. All we could say with some measure of confidence is that some element or some quality would eventually appear that was not there before and that did not follow so obviously from what went before that we knew it would happen. And we suspect that Matisse didn't know it would happen, either. When it does happen, it illuminates the connection with

what went before, but it says nothing about future moves, which remain as unpredictable as ever.

Does the individual know where he is going? If he did, wouldn't he simply go there? The individual appears to be directed less by what he wanted the single work to *be* than by what he wanted his work as a whole to *become.* But if the single symphony or the single sculpture figures only as a single step in a more complex and a more extended story, how does the form of that story affect the making of the single work? Or should we, rather, regard the story as merely a random walk from one interesting state to another? In that case how and why does the individual choose those particular states? And if the individual has reasons for choosing what he chooses, how can we still describe his progress as a random walk?

In the earlier days of my involvement with computing, I would frequently run across scientists who would look at what I did, and then tell me, "you should look at X – where X would be whatever was fashionable before fractals and neural nets – because X is really interesting." And at first I would think, what ***does*** this person think I do, that he would imagine I would find this X interesting? Then gradually I came to realize that my scientist had a notion of abstract interestingness that was independent of anybody being more or less interested. In this sense we might imagine a search space full of interesting things. But in the absence of some directing principle everything would be equally interesting and we would find ourselves a bit like Little Jack Horner with his Christmas pie. He put in his thumb and pulled out a plum – fractals, or neural nets or whatever -- and said, What a creative boy am I!

Is the ability to generate new material a necessary condition for creative behavior? Of course it is! But it falls far short of being an **adequate** condition. Nobody can doubt that the twelve monkeys sitting at typewriters are quite likely to generate new material, something no one has ever seen before; actually that's about all they will do. The question we usually ask is whether they will ever generate something we *have* seen before; specifically, the complete works of Shakespeare or even a single sonnet. If we view these poor, overworked monkeys simply as a device for clumsily searching the search space of all possible character arrangements, then obviously they are exactly as unlikely to generate a particular sonnet as they are to generate any other pre-determined character string of the same length. Plenty of new material, no ***significant*** new material; no way of having the monkeys observe the rules that would constrain the search space to the space of all legal sentences, for example, and certainly no way to have the monkeys dream up the rules for the production of world-class poetry.

Which then reaches to the heart of the matter. How does the human being dream up the rules for the production of world-class poetry? Step by step. Not by the random stubbing of toes against interesting possibilities. The search-space notion seems to me to lead to an extremely unpromising hypothesis for creative behavior, implying as it does that all possibilities exist already. Well, perhaps all prime numbers really do exist in the space of all prime numbers; perhaps all possible arrangements of three rectangles do pre-exist and perhaps it is interesting to find an arrangement I hadn't thought of. But I believe the creative act involves moving through a series of intermediate states that are significant to the degree that they are increasingly closer approximations to some weakly defined but strongly sensed end-state. **The creative individual**

doesn't find them because they aren't there to be found; they don't exist until he constructs them.

That's not what Michelangelo said, though. Michelangelo, so the stories tell, said that the figure was inside the block of stone and all he had to do was to remove the superfluous material. Did anyone ever take him literally, I wonder? Wouldn't that imply that Michelangelo's genius rested on spotting the blocks of stone that contained the best figures? When Mozart said that he had the whole symphony in his head before he started writing it down, do we take it literally to mean that he had composed the whole thing, chord by chord and instrument by instrument? Or do we suppose that he knew just enough about what he wanted that the notes would seem almost to write themselves?

Why do creative individuals have no better insights into their own processes? I assume it's because so much of what goes on in human creativity goes on in the non-conscious part of the mind; behind our backs, so to speak. It isn't just that we *don't* know how decisions are made, we *can't* know. But let's not jump to the conclusion that creativity is a property unique to the non-conscious mind, and consequently not possible for a computer program that doesn't have one. Just about anything with a nervous system exhibits *non*-conscious behavior. What is peculiar to the human brain is the curious property we call consciousness.

I've been asking a lot of questions. Do we know any of the answers? For the sake of argument , I will assume myself to be a conscious person with a modest level of creativity and no more insight into my non-conscious processes than the next person; and from that position I will venture to say that there are some answers I'm reasonably sure about. I'm reasonably sure that I do not stub my toes against interesting pre-existing possibilities, and I am more than reasonably sure that creativity rests upon a complex infrastructure of knowledge. Some of that knowledge has been acquired non-consciously; some of it has been acquired consciously but thoroughly internalized, so that I don't know I have it until it surfaces. I am quite sure that, however it is acquired, there is no creativity without knowledge, and that the way the individual's knowledge is deployed is critical to creativity. That would have to be as true for the creative program as it is for the creative person.

I also know for sure that a computer program is not the same kind of thing as a monkey with a typewriter and not the same kind of thing as a human being. In speculating about what computer programs can or cannot do we have to be especially careful not to confuse the goals with the hardware, because evidently we don't find it trivially easy to say what the goals **are** without tripping over the presence or absence of human-specific hardware. Goals are hardware independent, strategies are not. Instead of asserting that since a machine doesn't have the human's complex nervous system that enables its owner to do this or that, then the machine can't do this or that, we have to see whether we can devise strategies for doing this or that in terms of the hardware the machine **does** have.

One might reasonably have assumed, for example, that handling something as vision-dependant as color would be quite beyond the scope of a computer program. Yet over the past year or so AARON has been performing rather well as a colorist without the sensitive visual system that allows the human colorist to see what he's doing. The program's expertise involves a set of rules robust enough to generate good results under a wide range of conditions, but those

rules would be useless were it not able to build a durable internal representation of a complex image involving complex color relationships.

Vision frees the human colorist of constraints, wouldn't you think? Apparently not. If I were a computer program I might conclude that the human colorist **needs** to use vision to compensate for his miserable inability to maintain a stable internal representation. Imagine the poor human artist trying to make a painting by telling an assistant what colors to mix and how to mix them, without ever himself looking to see the intermediate results. He **needs** to use feedback from what he sees happening on the canvas in determining subsequent decisions. He spends a good deal of effort on adjusting color relationships, and he almost never gets it right first time.

The human artist has successive-approximation rules that require visual feedback from the external world, which the program doesn't have. The program has dead-reckoning rules that require an internal representation that is beyond the capacity of the human being. AARON now performs as well or better than most human painters with respect to color, and I have no reason to assume that it has reached the limit of its capability. What would be the grounds, then, for insisting that the human colorist is somehow better than the program? Presumably they would have to invoke the more general argument that the human being has the flexibility to change his notions, about color in this particular case – not that the human artist frequently **does** change his notions – while the program doesn't. That's easy enough to do, and I wouldn't challenge it; quite obviously AARON acquired its expertise through my own ability to reformulate what I knew about color in non-human terms. But it doesn't alter the fact that AARON is using non-human hardware to perform a task we would certainly have assumed required that hardware; just as the people at IBM were able to use machine-specific strategies to defeat the world's greatest chess master at his own game.

For the record, here's a summary of what was involved in developing AARON's color expertise; which, we should keep in mind, had to satisfy the human viewers of its work. First, the prior formulation of a theory of color use which sought to say, in general terms, what human artists use color for. Second: a consideration of how human beings satisfy the criteria imposed by the theory and speculation about how a vision-less program could satisfy those same criteria. Third: knowledge about color perception based derived from many years of my own experience as a painter. Fourth, and to supplement this knowledge, extensive data concerning the properties of the colored dyes the program would be using – what you see on the screen is in fact a simulation of the large-scale painting machine AARON uses in the real world – particularly with respect to how the brightness of a dye mixture changes with dilution. And, finally, the construction of a robust rule-set for generating color palettes for the program's images and for assigning individual colors to the different elements of the image. Although I said that AARON never changes its plans once it starts coloring, the rule-set is nevertheless feedback-driven at the level of designing a palette, since it has to juggle the inter-relationships of some number of colors in terms of their fundamental properties -- hue, brightness and purity.

I mean feedback here in the sense of the self-control exercised by an intelligent, knowledge-rich entity whose actions follow from intentions and which can assess to what degree its intentions are satisfied. Obviously it is a critical component in any intelligent system having to deal with an unpredictable environment, and a creative system can be no exception. Characteristically

feedback is manifested in programs by what we call production rules -- "If such and such is the case do the following…" A predicate followed by a consequent. Production rules imply two kinds of knowledge, then: the predicate implies that the program knows enough about the state of the world to decide whether the predicate is satisfied, while the consequent implies that the program knows how to perform the required action. None of this will come as news to anyone here, but before expanding upon what it actually means in the design of creative programs I want to talk about a rather less obvious issue that is perhaps even more fundamental to creative behavior. It has to do with criteria..

We would probably agree that creativity implies the development and application of criteria, but what are criteria? And what do criteria have to do with rule-based programs?

The movie critic writes "So-and-so's new movie simply doesn't sustain one's interest for two hours." "This director's work indicates a singular lack of moral awareness." Criteria are not just standards to be reached; they are standards to be reached with respect to particular issues. The critic's comments imply that a movie *should* retain the viewer's interest, that the director *should* be aware of the moral issues raised by his own work. But the critic rarely bothers to state explicitly **why** that should be. Why **shouldn't** one make movies designed for casual intermittent viewing, for example?

It is similarly true that criteria are almost invariably implicit in rule-based systems also, whether those systems are operated by people or programs. "If the sauce starts to separate, do the following…" is a rule, but the state of the sauce is the subject of the predicate, not the criterion. The criterion, which the cookbook author never bothers to state explicitly, is that sauces should have a smooth consistency and should remain smooth throughout their production.

So it turns out that the writers of rules make certain assumptions about the issues in relation to which their rules attempt to define standards. We are all assumed to know what the issues are and we are all assumed to agree that they are important; to the degree that they never need to be discussed. That's fine for the critic, but a good deal less so for the creative individual who moves forward into territory that is, by definition, unpredictable; not uncharted but un-constructed. And for anyone involved in writing a creative program the distinction between a rule and what implicitly informs the rule, between a predicate and a criterion, is critical. One can write explicit rules; there is no special reason why a program should not re-write its own rules. The problem is that what would **require** the rules to be re-written would have to be a shift in the program's criteria, and these are never explicitly stated in a rule-based program any more than they are explicitly stated by the critic or, for that matter, by the human artist.

We must then suspect that the limit on a program's creativity is not determined ultimately by its ability to modify its own code, but by its ability to modify its own criteria.

Where do they come from, these criteria? Who determines the range of issues that defines an artist's criteria?

Obviously they are culturally determined to a large degree. It is comparatively rare, that is, for the work of a single individual to enforce a major shift of criteria for the entire culture. At the same time creative behavior must involve

an increasing differentiation, of whatever degree, of the individual's criteria from those of the culture at large. The process of acculturation begins almost as soon as the child leaves the womb, and my guess is that individuation begins not much later, by the time the child is able to construct something very like a rule-based system for itself. "If I cry, I will get something nice and warm in my mouth and then in my tummy." The connection between behavior and result – predicate and consequent – gets less simple as the child gets older, but the criteria which drive the rule-based system are clearly related to maximizing the child's influence over its environment to its own advantage.

She can elicit praise by making a drawing, as she could earlier by turning somersaults or by drinking from a cup. Praise is enjoyable. But sooner or later she'll decide it might be interesting to drink from a spoon, and in the competition between messy experiment and tidy praise, praise comes in second to the more important goal of gaining skills and enlarging experience.

My daughter, Zana, is nearing her fourth birthday, and the praise-seeking phase, in which she would spend almost as much time showing us her latest drawing as in making it, is all but over. She draws for long periods without showing the drawings to anybody and she discards them as soon as they are done. And increasingly her drawings are directly related to recording and to reviewing her own experience.

For example: she got quite excited last Easter about Easter eggs. She took part in a couple of Easter egg hunts and she helped her two sisters to color their own Easter eggs, using the traditional method of floating several non-water-soluble colors on a bowl of water and dipping the eggs to pick up the colors. Zana had never used color previously other than in the sense of choosing a red pencil rather than a blue pencil to make a line drawing. Now she was clearly fascinated by the complicated pattern of colors on each Easter egg, and result was that for several months color became part of the content of her drawing rather than an optional and not very important part of her drawing practice. Older subject matter was recycled in terms of color. Noses took on a life of their own, for example, and new subject matter dealing specifically with color began to appear, beginning with Easter eggs themselves and going on to subjects like rainbows.

Zana must have seen many more pictures of rainbows than actual rainbows, so I was quite surprised to find her drawing them vertically with the ends tucked in. But, I had to remind myself, drawing is never a simple photograph-like transformation of the retinal image. The information that enters through the eye helps to form a mental model, and it is always this internal model that the individual draws. (I have watched students in life classes draw for ten or fifteen minutes at a time without looking at the model, while at the same time I've been able to demonstrate that they couldn't store up enough visual data to support ten seconds-worth of drawing.) Zana's failure to get the rainbow right is actually very like the condition of the art student who struggles to get the nose or the foot right with the model actually sitting in front of her. She can see it, but she can't form a clear enough internal model and when she externalizes that model it fails to correspond to the outside world.

In this case I was persuaded to do the line drawing for her – something I rarely do -- because it was clear that my contribution was simply a format for the real issue, which was color.

By now drawing has become a general skill for Zana – I mean that she is no longer limited to drawing particular things in a particular way as she was a year ago – but she still has limited technical resources and it is her **conceptual** grasp of the external world, finally, that limits what she can do. That becomes clear when she tries to draw something she's never drawn before. She called me from Japan this summer to tell me that her Mummy had bought her a new hat for when she went out in the sun. When I asked her if she would make a drawing of the hat and fax it to me she responded that she couldn't draw hats. I told her I was sure she was smart enough to figure it out, and she did, in fact, send me this (very peculiar) drawing. It was only later that I saw the hat itself and realized that she was drawing a circular straw hat, and that the jagged lines were intended to represent the woven straw.

No doubt she will in due course draw a hat that reflects her increasing conceptual clarity about what a hat is. But if one of her siblings shows her how to draw a hat cartoon fashion there's a better-than-average chance that the cartoon hat will persist for her and the conceptual un-clarity will persist also. This last proviso is important, because it implies -- what in fact seems to be the case -- that representational strategies are re-enacted without modification for as long as they satisfy an unchanged conceptual state. For a very long time Zana indicated ears by little horizontal marks breaking the boundary of the face at about the right level. Only very recently did she begin to draw ears as closed forms. Eyes were little vertical marks, drawn, characteristically, at the top of the head. Actually, this is so persistent a behavior for children that it may simply reflect the fact that from a small child's upward-looking perspective the adult's eyes really **are** at the top of the head. And it persists far beyond childhood; students learning to draw from life almost invariably place the eyes much too high in the head.

Watching my daughter draw I am quite sure that drawing is an important tool for the individual, child or adult, in developing conceptual clarity. To paraphrase Wittgenstein, we only know what's in our heads by the images we make; the image serves as a reality check; to show us what we believe and thus to enable modification of those beliefs. Not everyone does use it in this way, of course, and in real-life situations it's hard to disentangle from other considerations. For the professional artist in particular, the practice of art places so much emphasis upon style, that the more fundamental cognitive function is confused with trying to achieve a satisfactory – that is, an art-acceptable, original, wholly acculturated -- appearance. But essentially the same thing is true for children also. Every so often I find allusions in her drawings to the way other children in her school draw and she is affected, to a much greater extent actually, by watching video programs of her favorite cartoon characters. That part of her experience is pre-drawn, of course, and that makes it that much more difficult to track her cognitive development through her drawing.

I remarked earlier that knowledge is implicit in any rule-based system because the program couldn't examine the state of predicates it didn't know about and couldn't perform the consequent actions unless it knew how to perform them. What we see of cognitive development in the child indicates, as I've suggested, that knowledge of the world does not mean a collection of facts, but a complex internal model that has been built to represent the world. It's pointless to talk about whether that internal representation is "correct" in relation to what is being represented; it is what the individual has, and it will change only with the accumulation of experience. But life is experienced across a broad spectrum.

Even when Zana draws her favorite cartoon characters there is experimentation going on that is not prescribed by the original. Here, for example, is a series of Zana's drawings of Anpanman, beloved of all Japanese children, made over a period of about an hour. Anpanman is forever round, and Zana is experimenting with alternative possibilities; yet the mode of drawing – the delineation of closed forms -- remains constant. Compare that mode with how she made a recent drawing of a stuffed animal which was sitting on the table in front of her while she was drawing it. This "filled-in" mode is something she has only used about four times over the past year, and in each case it has been when she has been drawing from direct observation. In this case it is probably significant also that the toy was newly acquired so that she hadn't had time to fix on a representational mode. My guess is that the filling-in represents Zana's awareness of the solidity, the physicality, of the subject. But her awareness is inside her head, not mine, and I can do no more than guess.

In summary, then; it would seem that what can be externally represented depends upon a complex interplay between the patterns of knowledge constituting the internal representation and what representational skills the individual can call into play. All very human. To what extent might we hope to find similar patterns in a computer program?

It is one of AARON's most valuable assets that it has been under continuous development for more than twenty-five years. It remains remarkably simple compared to the complex internal models of even the very young child, but there has been sufficient time for the program to take on a complexity no younger program could have. It has acquired nothing for itself, so I don't use terms like cognitive development when I'm talking about its history, although I do wonder occasionally how much difference it makes how knowledge is acquired; after all, most of what we know today comes from books and television sets, not from direct experience. But, in any case, how often do I need to remind myself that a program is not a person? What should we expect the machine equivalent of human cognitive development to be like? Since I've been at pains in developing AARON to build upon, never to abandon, its accumulating structure, I can at least show now that there is a connection between AARON's strategies for drawing and its internal model of its limited external world.

In the early days I tried to base AARON on what I saw my own very young children doing. They were at the scribbling stage, and specifically at the stage where a round-and-round scribble migrates outwards and becomes an enclosing form. Eventually the child allows the scribble which generated the enclosing form to drop out, leaving an outline that can then serve as a basic element in representing objects. Superficially AARON went through a similar transformation, but for the child the closed form appears to acquire an independent existence no longer requiring the original internal scribble, while AARON still today has to draw it, "in imagination" as it were, in order to generate the closed form.

So at some point AARON's development diverged from that of its human archetype, and I was obliged to devise strategies for an alien entity that could generate results that would not be alien to the human viewer. That raised all kinds of problems and in many respects it would certainly have been easier to start over with a new basic structure. For example; as AARON moved on from its early "primitive" phase to explicitly representational drawings, its knowledge about what it was drawing continued to relate specifically, and exclusively, to the core figure from which outlines are generated; no longer scribbles now, but

articulated internal structures controlling posture. Surfaces, in the sense of the physical boundaries of real-world objects reflecting light to our eyes, don't exist for AARON. If the word "surface" has any meaning, it means simply the areas bounded by outlines on the picture plane. That might not be too obvious from the appearance of AARON's images, until one realizes that they are quite devoid of the shading that would indicate the variable play of light on surfaces. In practice, it means that the program can't deal with occlusion by the standard hidden surface methods of computer graphics, and I've been obliged to develop a quite non-standard method for representing the changing state of the drawing. But AARON is AARON; and I felt very strongly that whatever problems came up I would have to deal with them in AARON's terms. Or perhaps, and thinking now about the problem of coloring without seeing, I simply had no choice.

So, in more detail, how does AARON's knowledge constrain what it is capable of doing? How does the structure representing the current state of the fluidly-developing drawing affect what can be done subsequently? As a simple example: if the program wants to draw a second figure, what does it need to know about where it had placed the first figure? But perhaps that isn't as simple as it sounds.

To begin with, in any representation of the physical world dealing with where things are, the word "where" has, simultaneously, two distinct meanings. It refers to the location of the representation of the figure within the frame of the picture, and it refers also to the location in the real world of the real figure, which, with respect to the viewer, gives rise to its two-dimensional location on the picture-plane. However, the two are not necessarily compatible; a perfectly reasonable distribution of objects in the real world does not necessarily result in a satisfactory distribution on the picture-plane. And in fact there's no direct way of predicting much about what the picture will look like from what the world looks like. That's why Polaroid makes a lot of money selling instant film to professional photographers; they need to know what the picture will look like while they are still in the process of arranging their models and their lights and their camera tripods.

Before AARON had an external world to represent, it's compositional strategy was framed in terms of the way information density changed across the picture plane, information levels being assigned somewhat arbitrarily to outlines, line-junctions and so forth. That all broke down when the program became overtly representational, because it was rather obvious, firstly, that eyes and noses have a very different information content from line junctions and closed forms: and secondly, because representations are constrained by the way the external world behaves in a way that closed forms and scribbles are not. I don't think I understood that very well at the time, and without any other compositional mode at my disposal I persuaded myself that since anything we see through a window looks perfectly reasonable, so a simple perspective rendering of that frame should be equally reasonable. I have always maintained that the most universal rule of composition is "put it where you can find space for it." But I quickly discovered that that rule isn't as simple as it sounds, either. There proved to be a big difference, in terms of reading, between a figure whose torso is obscured by foliage and one whose head is obscured by foliage; the one seems to be hiding while the other seems to be hidden. And the program had no control over what the viewer might then read as the dramatic content of its drawing.

At that stage, AARON had only a sort of two-and-a-half dimensional model of the figure, which it would construct from the head down, just as any human artist

would do. Individual figures looked plausible until I attempted to place them within a physical context, where the programs lack of control over where the feet finished up became quite evident and with embarrassing regularity I would find a foot half-way up a tree that was supposed to be behind the figure.

I responded to that problem in part by providing AARON with a fully three-dimensional model of the figure and in part by having AARON constructing its figures from the feet up, finishing up with the head: something no human artist would do.. In planting the feet firmly on the 3-space ground-plane I solved the misplaced-feet problem, but found that I had introduced a new one in its place: namely, that where the head would finish up was now largely a function of posture and the figure's orientation in 3-space with respect to the viewer. The result was that two figures that had been side by side in the real world would frequently end up with one head partially hidden behind the other.

By this time it was becoming clearer to me that the problem of placing representational elements in the picture couldn't be solved by placing the real-world elements in the real world and then rendering them in perspective. And, of course, a little reflection showed that it never had been solved that way. The fully developed perspective of Western art was always preceded by sketches, the primary purpose of which was to consider and manipulate "composition" -- that is, the distribution of elements **on the picture-plane** -- independently of where the real-world elements might have been in the real world. That was true even during the High Renaissance, when "correct" perspective was mandatory and rectangular solids – buildings, tessellated floors, furniture and so on – were habitually constructed by rule rather than by observation. AARON, conversely, was not considering composition at all; the elements fell into place in the straight perspective rendering as a result of the program's control over posture, and that afforded only the coarsest-grain control over placement.

The next phase of AARON's development thus required a significant increase in its knowledge of the picture-plane and how to place things in relation to it. That may sound like a non-issue – after all, what is there to say about a picture-plane? In fact – and as I've tried to indicate -- there is a big difference between an empty frame which is held up to the world to receive whatever happens to be out there, with the bits falling where they may; and a rectangular space whose very rectangularity plays a dynamic role in the artist's determinations; because if the artist has control over the way elements are placed in relation to the frame he also has a good deal of control over the way the image is read and understood by the viewer. I'm not talking here about composition in terms of geometrical strategies that have been used by a few artists from time to time – the oft-cited but rarely-used golden section, for example -- for placing distinctly non-geometrical entities like heads and bodies within the frame. AARON's problem was not that it didn't know anything about the golden section, but simply that it didn't know how to find space for what the viewer was supposed to see.

In its current version AARON comes much closer to the human way of doing things. It still constructs its figures as three-space objects viewed, as in an orthodox perspective, from a position within the space; plausibility requires that much. But then it has the freedom to manipulate the placement of the figures in relation to the frame. This involves more than simply shifting the elements around, of course, because the relative sizes of the initial unmodified elements depends upon their distance from the viewer in three-space. The sizes they are to have in the picture is determined by where the program wants the figures to

be and how much of the frame it wants each one to occupy, but the relative sizes still needs to be maintained if plausibility is not to be sacrificed.

In practice the two-space points in the figure are determined by the perspective rendering and then scaled to fit whatever space the program wants to assign. I'm talking here about the points in the core figure; the outlines are generated only after the figure is in its final position. The difference between a head-and-shoulders portrait and a full-length figure is principally a difference in scaling: in each case the bridge of the nose may be placed roughly three-quarters of the way up the frame, say, while either the neck – for the portrait -- or the feet – in the case of the full-length figure -- is placed somewhere close to the bottom of the frame.

When the program comes to draw a second figure, which is further away in the real world and thus smaller in the original perspective, it uses a slightly different strategy. Instead of deriving the scaling factor from two points in the figure and where it wants them to be on the picture plane, it uses only one point and calculates an appropriate scaling factor by modifying the scaling that was used for the first figure according to how far away the second figure is supposed to be. In this way it is able to keep the size relationship consistent with that of the perspective rendering even though both figures are different in both size and placement to the original perspective.

So much for the vertical component of the composition issue, in which the program needs only to know the scaling and the placement of a couple of key points in the first figure in order to place the second figure plausibly. But the vertical component is only half the story. The other half – the horizontal placement – shows that the knowledge-dependency of the program's strategies currently leaves something to be desired. Of course, control could be exercised quite trivially by allocating separate parts of the picture plane to the individual figures, but that would severely limit what the viewer might read as the interpersonal relationship between the figures. The closer two figures are to each other the more obvious it becomes that the program has no fine-grain control. It becomes increasingly likely that part of the second figure's head might be obscured by part of the first figure, and the program certainly couldn't decide, for example, to place the second figure with its right ear just to the right of the first figure's left ear.

I use the absence of fine-grain control here simply to underscore the issue of knowledge dependency, not because there's anything inherently difficult in providing that level of control. The reason the program doesn't have it is actually quite trivial: when the program is drawing the second figure it has already discarded the fine-grain information it had about the first figure when it was drawing it: it can't relate anything to the first figure's left ear because it doesn't know where the first figure's left ear is. There are almost a thousand control points in AARON's prototype figure, and each point is represented by a data structure requiring about fifteen words. AARON dates from a time before memory was the cheapest part of a computer, and the fifteen thousand words are recycled after each figure is completed to make room for the representation of the next figure.

There's no longer any good reason why each figure should not retain all the data that was generated in drawing it, beyond my own disinclination to go through a megabyte and a half of lisp code to make all the necessary changes. But until I do, the limit on what the program can do towards meeting its own objectives –

no! towards ***formulating*** its own objectives -- is constrained by how much it knows about what it has done. If I wanted to go further, to have the figures in actual physical contact – holding hands, for instance – there would clearly be no option but a wholesale revision of the data structures retained by the program to represent the drawing in progress.

No big deal. But then, my ambitions for the program aren't going to be satisfied by having the figures holding hands. Increasingly, the big challenge for me has been to give the program much greater autonomy than it has presently; and that, without a doubt, involves moving the problem of knowledge and its representation to an entirely different level; and that looks to me like a very big deal. AARON still does each painting as if it has never done one before. Autonomy implies that it needs to learn something from its own actions; which implies, in turn, that AARON should have a long-term memory. But what is the program supposed to remember? Retaining the initial scenario it constructs for each image doesn't tell it enough to differentiate one example of that scenario – two woman with potted plant or whatever -- from another one. Storing the entire final image would be equally useless, since in recalling it the program would be faced with having to make sense of an array of colored pixels as if it had never seen them before, much less made them.

Somewhere between the extremes there has to be a representation tailored to the use AARON might then make of the knowledge it contains. And in building that representational architecture, remembering and recalling have to proceed in lock-step. The program won't be able to learn from past actions unless it can access what it needs to know, while at the same time there's no value in committing to memory knowledge of past performance it can't subsequently use.

Well, the fact that I don't know how to solve this problem right now doesn't mean that it's intractable, and I'm working on it. But evidently there's a long way to go before AARON can go on to eternity producing new, original images in the deeper sense that I've tried to indicate. Which reminds me that I started this talk with a joke about being the first artist in history to have a posthumous exhibition of new work. Maybe. I have to confess to feeling less like a potential immortal than like Moses looking out over the promised land he will never enter. Aaron did, if I remember correctly. I don't mean that I think the final stage of AARON's development is beyond my reach, though it certainly may be. I mean that if AARON ever does achieve the kind of autonomy I want it to have it will go on to eternity producing original AARONs, not original Harold Cohens. Apparently the joke is on me.