ON THE MODELLING OF CREATIVE BEHAVIOR

Harold Cohen

November 1981

P-6681

The Rand Corporation
Santa Monica, California 90406

ABSTRACT

The introduction to this paper discusses the notion of human creativity, and raises the question of designing a "creative" computer program. Creativity is assumed not to imply the possession of special mental equipment: a theory of creativity should be a theory of intellect which accounts for normal performance and enhanced performance in the same terms. Art-making is described as a form of creative behavior which demonstrates the importance of non-rational features. It is argued that the central feature of "enhanced" intellectual performance is the individual's ability to modify, by the manipulation of internal representations, his/her own mental structures.

The processes of representation constrain the actions of the representer, and thus what he/she is capable of representing. Part Two examines the anatomy of Representations in technological terms: the means, the skills, and the theory of operation (of the representation process) which the individual may bring to bear, and the constraints which result. It is proposed that representations represent lower-order representations (internal models), not the external world, and that the making of external objects plays a role in "checking" internal representations of explicit information, is shown as a culturally-modulated phenomenon distinct from evocation, which draws upon more inherently human capacities.

Part Three describes a program designed to investigate the interaction of a primitive internal model of world objects with a "representational technology"--the technology by means of which the

internal model becomes externalised. It is shown that the external representation takes on characteristics which derive from the technology and not from the internal representation. The economy which the program exhibits is shown to result from its use of meta-data--the encoding of the data embodied in the internal representation which facilitates its processing through the representational technology.

Part Four develops a fuller model of creativity based upon the associative character of memory. It is argued that representation-building involves the introduction of "counter-purposive" material, and that the inherent "noisiness" of representations is a positive element in creative performance.

Part Five considers broad design specifications for a program which would exhibit creative behavior. The proposed program will deal with "visual imagination" as the highest level of internal representation: thus the human visual cognitive system could provide the basis for the program's representational technologies. It should be capable of self-modification, through the progressive development of higher-order representations out of lower-order ones. Its output should exhibit the "completeness illusion" characteristic of visual imagination.

CONTENTS

PART ONE: INTRODUCTION

In recent years, a great deal of work in Artificial Intelligence (AI) has concentrated on two areas of research: the automation of rationality and the building of expert systems. These approaches, and the strategies in the representation of knowledge that have developed along with them, have provided a powerful characterization of some of man's higher intellectual functions, one that has proved quite appropriate to the now-orthodox problems of AI. The success of these approaches and these characterizations has, of course, defined the "now-orthodox" problems of AI.

Historically, to the degree that it is exemplified by the "expert systems" approach, the AI enterprise represents the convergence of two earlier, differently motivated trends. On the one hand were those researchers who saw in the new discipline a new way of considering the curious workings of the human mind. On the other were those who saw the computer as a tool of such broad application that it might be effectively applied to tasks previously the exclusive domain of human intelligence, and they recognised no obligation to the human mind as a prototype for their own endeavours. The convergence of these philosophical/psychological and engineering considerations has rested upon a growing consensus that AI should be concerned with intelligence, not uniquely with human intelligence [1]. This consensus has cleared the way for a pragmatic bypassing of the fact that we know relatively little about how human intelligence works.

It is, of course, the case that intelligence is rooted in the particular characteristics of the organism or mechanism possessing it. Even if we wanted to claim that some products of intelligence, e.g., laws of logic, are universal, we would still need to recognise that knowledge, the material with which intelligence concerns itself, is certainly not universal, but that its acquisition is profoundly dependent on the organism's acquisitional modes. Building this knowledge into systems of belief about the world must therefore also be dependent on the character of the organism: an organism equipped with radar rather than with binocular vision would not have generated Euclid's parallel line postulate.

Thus, whether or not the AI community will further free itself from the human prototype in generating machine-based intelligent systems, it has not freed itself from human considerations with respect to knowledge and the representation of knowledge. The automation of rational processes may require that we consider only the nature of rationality itself, without concern for the human intelligence that has so far been its only known manifestation. But, at least to the present, the knowledge with which any intelligence must deal has been acquired by organisms with eyes and arms and legs and cultural forms like language and social institutions, but with no particularly marked rationality. In general, an AI program that involves the re-embodiment and manipulation of bodies of specialised knowledge is likely to be less a propositional calculus than a set of highly prescriptive, and patently pragmatic, rules. The tasks that AI programs are meant to serve remain human tasks in the sense that they are meant to be performed on behalf of humans and

within the cultural context. For what other reason has the community made the massive investment toward providing the computer with "eyes" and with natural, i.e., human, language?

There has been some success in manipulating significant bodies of expert knowledge by machine--though nothing to compare with what the average human intelligence manipulates. However, the strategies used for the representation of knowledge inside computer programs have little presumed similarity to the strategies the human expert develops for storing knowledge in his own memory. Nor would any claim be advanced that there is any similarity. Machine-internal representations are designed to facilitate machine-specific operations. The forms of internal representations in man's mind develop in an ad-hoc fashion, as part of the process of acquiring knowledge, and those forms are as likely to constrain as to facilitate subsequent intellectual activity. In the fullest sense, an expert is not someone who merely possesses an extraordinary body of knowledge, it is someone who exhibits some enhanced (but not necessarily more rational) mode of intellectual performance in deploying that knowledge. But questions about the nature of expertise, about how the expert acquires, stores, and deploys the knowledge which the knowledge engineer seeks to re-embody in program form inevitably have been left on one side.

It is certainly true that the characterizations of intelligence arising from the work of the past decade have proved to be powerful, and major achievements rest on them. It is certainly <u>not</u> demonstrably true that all higher intellectual functions in man must yield to the same characterizations. How important is the lack of correspondence between

machine-based systems and human systems? Is there some unavoidable limit on a program's ability to perform human tasks unless it written to emulate the way humans perform them? Will a program ever "have a good idea," for example, other than in the same sense that humans have good ideas? Since we know little about the nature of creativity in man, could a program perform creatively unless it was written to elucidate the nature of human creativity? Questions as to what machines can and cannot do do not turn on the innate powers and limitations of machines themselves but on our ability to characterize what it is we want them to do.

The central concerns of this paper are the nature of human creativity and the problem of designing a "creative" program. Before proceeding, I need to say several things about the assumptions underlying these concerns, their nature, and my motivations for undertaking this enquiry.

First: the word "creativity" is used here to denote a range of enhancements to intellectual performance which are strongly generative; that is to say, they emphasize the speculative building of new mental structures rather than the analysis of existing structures. These generative modes are found most readily in the arts, where speculation is not constrained in the same ways that it is in the sciences. Nevertheless, there is no reason to suppose that what we observe in the arts is in any fundamental way different from what we find in the sciences.

However, the use of the word "creativity" should not be taken to imply the existence of some clear and accepted theory of creativity

ready to be embodied in a program. There isn't one. Considering the fact that creativity has been a source of fascination from earliest times, the lack of adequately rigorous definition seems remarkable. The large literature on the subject is predominantly anecdotal, more enthusiastic than precise, and records the common belief that individuals displaying creativity in marked degree possess abnormal resources or act under supernatural guidance. They have "genius," they are "inspired," the muse sits on their shoulders. All too frequently, "creative" individuals who go on record describe their experiences as "mystical"--as if they, or their readers, might better understand anything labelled in this way.

With the concept of creativity so ill-defined, it is hardly surprising that it has received almost no attention within the AI community. But there is another reason also: the essentially analytic character of orthodox AI strategies has left the researcher without powerful tools for modelling its essentially generative character. With the exception of "AARON" [2], a program of mine that modelled image-generating processes in art-making, and its successor "ANIMS," dealt with later in this paper, no significant attempt to model creative behavior is on record.

Second: in considering the nature of creativity, it is tempting to simply assert that some individuals use their intellectual resources with vastly greater efficiency than most people do. That, of course, flies in the face of the largely unsupported popular belief that creative people have special mental equipment. However, we are in an area of speculation where there may be reasons for preferring one

construct to another but no reliable evidence upon which to choose. In this case, my preference is unequivocal. Little as we know about creative behavior, it seems quite likely that no individual has ever approached his/her potential limits intellectually, and we have no idea where those limits might be. Proposing that creativity is impossible without special equipment pointlessly ignores the possibility that there is a continuous spectrum of potential efficiency in the use of resources and, thus, that any individual may move across that spectrum.

Given that, I advance the axiom that a theory of creativity should be a theory of intellect which accounts for normal intellectual performance and extraordinary performance in the same terms, that is, in terms of the acquisition, storing, and deployment of knowledge. It should not be necessary to explain extraordinary performance as "superhuman."

Third: consonant with the above, this paper presents a theory of creativity only in the sense that it advances a general view of mental activity that lends itself to examining and modelling creative behavior. This view of mental activity has arisen speculatively and on introspective evidence. In the absence of explicit and examinable evidence of the workings of the mind--as opposed to its output--it is hard to see how else one might proceed. One cannot "prove" a theory to be correct by writing a program. But in this case, I believe that the theory that emerges suggests that computer programs can, in principle, function creatively. The reader should regard this paper as an extended hypothesis, significant to this degree: (1) that it provides a plausible account of something innately unexaminable, (2) that the account offers

a good fit to a range of intellectual problems, and (3) that it supports further speculation.

This is not to say that the account of mental activity given here takes poetic license with what can be observed of the mind's activities. In fact, any adequate theory will need to be extremely broad-ranging in its scope. Since the individual functions (creatively or otherwise) in relation to a world external to himself, the theory will need to consider the part played by the external world in acquiring knowledge and determining the forms in which it is stored. Since creative activities normally involve the production of external objects (utterances, images, actions, and so on) the theory will need to consider the part played, for the individual, by the production of those objects, not because of their intrinsic worth but because their production appears to be an important element of creative behavior.

Once these objects enter the external world they take on an autonomous role as items of communication, and while the problems posed by communication do not fall within the scope of this paper, the theory should consider the curious fact that extremely creative individuals appear not to be well understood by their own societies in their own times. Finally--though also not within this paper's scope--the theory should examine not only the "how" of creativity, but also the "why." What provides the driving force for enhanced performance? It may be that the creative person is one driven constantly to revise his views of the world and his beliefs, while most people seek rather to reinforce whatever views they happen to have acquired. Of course, this view of creativity requires no superhuman explanation to support it: the drive

to self-revision may be a perfectly normal intellectual function, no less so because most people don't habitually exercise it.

Fourth: as I implied before, the central concern of this paper is modelling creative intellectual activity. But why is this endeavor more than academically interesting or important? I can answer that best by explaining the context out of which it arises.

My previous research has been in the modelling of art-making behavior, employing strategies strongly emphasizing the cognitive underpinnings of image-making. As a practising artist, I base my work on many years of experience as a painter. "AARON" is the program that best exemplifies this earlier research. It has shown itself capable of generating extremely diverse, even bizarre, output; yet it is not capable of the purposeful self-modification that characterizes creativity. It has become apparent that the programming structures it employs will not lead to that end.

I believe that the theoretical issues involved in constructing a program that will use image-making as a means of self-modification provide insight into wider issues of creativity. However, some explication is necessary to establish the relationship between art-making and these more general issues and the anticipated advantage of research ostensibly directed to art-making.

For most observers, art-making appears to be a game without a goal, a game that is played but not won. If that were the case, art would have very little in common with science, and it would be difficult to explain why it has always been among the most highly-respected of human occupations. It is more revealing to view art-making as possessing a

strongly hierarchical structure in its pursuit of goals. The goals to be satisfied through any local action may appear as meta-goals, and the artist may pass through a hierarchy of considerations in seeking them. The making of a single brush mark may address the question "shall I paint this flower red", "shall I paint this flower", or "shall I paint any thing at all?" Thus, the same ostensible act may represent a striving to capture the color of a tiny natural event or a striving to overhaul fundamental beliefs about the nature of art itself.

Viewed in this light, it is obvious that art-making has a great deal in common with science; but it is a great deal less structured than science generally is, and it is less constrained by concepts like "relevance," "correctness," and "evidence." Artists are not more creative than scientists, but the intellectual methods made available by art allow for irrationality and "hierarchical thinking," while the methods of science discourage them and make their examination difficult. Moreover, while there are obvious discrepancies between the day-to-day practice of science and the way that practice is represented in the literature, it may be inevitable that the modelling of science-like behavior by computer programs should favor that representation: it is easier to deal with because it lacks precisely those creative features which make the difference between creative science and ordinary science.

This distinction provides the deepest motivation for the current research. Programs which assume and follow the rationality of intelligence must be limited to ordinary achievement--just as an exclusively rational human intelligence would be limited and would stop short of invention, speculation, and adaptation. Programs which

recapitulate the ad-hoc rules of expert behavior are similarly non-adaptive: they are necessarily task-specific and highly specialized, and it has not proved possible, so far, to generalize from particular ad-hoc rule-sets to meta-rules of broad application.

A person is not creative because he/she performs familiar tasks extremely well. Nor is he/she creative by virtue of being prolific, though prolificity is certainly a common characteristic of the creative mind. The single feature by which the creative mind can be identified is its ability to modify its own beliefs and to generate new belief structures. Speculation must stand at the heart of creativity, whether for people or programs. The limitations of current programs may be obscured temporarily by the increasing size and speed of computers, but, in the long run, only an understanding of how human beings "have good ideas" will serve as the basis for a computer program capable of having good ideas.

Overview

To reiterate, a theory of creativity should be a general theory of intellect capable of accounting for both creative performance and normal performance in terms of the same mental resources. Consistent with this view, this paper is not directed primarily at the examination of creativity itself, but at advancing a view of mental activity which might then lend itself to the examination, and to the modelling, of creative performance.

The essence of this view is that all levels of mental activity involve acts of representation, to the degree that mental process can be

characterised as a continuous, free-running, representation-building process. The view has three central features.

The first is that what is represented is not the external world, but some lower-order internal representation.

The second is that the processes of representation as such may be seen to possess clearly defined characteristics which constrain the representer in the building of internal representations as in the building of external representational objects. The representational modalities to be found in the making of external objects are certainly artificial, and not natural: their characters may reflect innate human propensities, but their forms are culturally determined, and acquired by the individual within a cultural context. These modalities are highly purposive and by no means interchangeable.

It is thus possible to discuss representational technologies as artificial systems and to extrapolate from what we can see of these technologies in the external world to wholly internal processes. On this basis we can speculate about the characteristics that representational technologies would require in order to support creative activity.

Third: the making of an external representational object is a stage of, and is continuous with, mental process as a whole, and serves a clearly definable function within that process. Briefly, it serves to fix, by accretion, elements of internal representations which are transitory and partial. As the basis for an operational model, this view offers a unified treatment of what would otherwise appear to be an extremely baroque organization.

The Scope of this Paper

It will be evident by now that central to the argument to be developed in this paper is a rather unorthodox notion of representation--unorthodox, at least, from the standpoint of AI. Section II amplifies this notion by discussing the nature of representational technologies and the anatomy of representations. This section also deals briefly with cultural issues which bear upon the individual, not so much to comment on the uses made of his external representations as communicative artifacts, but to assess the implications of the fact that representational technologies are culturally acquired.

Much of the speculation about modelling creative behavior grew out of the writing of a program, ANIMS, which generates "visual" representations of animals. It may be regarded as a prototype for a more complex program in a number of respects. The short paper which discusses it, "Drawings by a Know-Nothing, Almost," is included as Section III. Section IV considers a model of mental activity in greater detail as a support for the views of creativity developed here. Section V discusses design issues involved in the operationalising of such a model.

It will be as well to make clear that this paper has been written in speculative, not in declarative mode. For me, it constitutes a jockeying for position, an attempt to bring a number of conceptual considerations together into a single informing principle. It will end at the point where work on an instantiation of that principle can begin.

PART TWO: THE ANATOMY OF REPRESENTATIONS

The Class of Representations, and High-class Representation

For the AI researcher, a "representation" is the form (machine-internal representation) in which a body of knowledge is stored in a computer. This form is dictated by the operations the program is to perform upon or with that knowledge. For the artist, a "representation" is a set of marks (on a piece of paper, for example), the form of which is dictated by the desire to express some attitude about the outside world.

It may seem at first glance that the two disciplines use the same word to express quite different things, if only because the first "representation" is wholly internal (to a computer) and has no corporeal reality, while the second is a physical artifact which exists in the real world. It is undeniable that there are dissimilarities between these two "representations." However, all representations are more "like" other representations than they are "like" the things represented, and there are general assertions which may be made about all of them. For example, we may assert that the form of each of the above "representations" is determined both by considerations of "medium"--a computer behaves like a computer and a crayon behaves like a crayon--and by considerations of purpose. The things we call "representations" are members of a class, not because all representations are blue, or because they all exist on flat surfaces, but because the ways in which they are made and used by humans

necessarily reflect both cognitive and cultural commonalities.

This is not to say that we all make or use representations with equal effectiveness. The art-sceptic's three-year-old daughter really can not do as well as Picasso. If the discussion were limited to painting here, it would no doubt be obvious that effective representation requires both knowledge of the means (e.g, paint chemistry, color, light and shade, perspective, anatomy, pictorial structure) and skill in manipulating those means. However, these requirements merely define technique. Maximal effectiveness requires also the possession of a theory of operation, a theory of how the means enable and/or constrain representation-building, given that the outcome of the activity is deposited in the real world and is used there by other people. Since the existence of such a theory of operation is what distinguishes technology from technique, we might reasonably talk of technologies of representation. Three-year-olds certainly do make representations; they simply lack a rich technology. The result is that all three-year-olds in any given culture make essentially the same representations.

The thrust of this paper is to characterise cognitive activity as internal representation-building and to do so in such a way as to show that differences in quality of cognitive performance among individuals may be described in technological terms, i.e, in terms of the varied possession of internal equivalents to the artist's means, skills, and theory of operation.

Representational Subclasses

Since this part considers some of the technological considerations which bear upon representation in its general classificatory sense, we might first identify three subsets of the class "representation" which figure significantly in the balance of the paper. Elucidating the ways in which these three relate to each other will provide interesting examples of these technological considerations in operation.

The three subsets are:

1. The subset "human-internal representations" or, because our concerns here are with human functions rather than with machine functions, simply "internal representations."

2. The subset with which this section opened: "machine-internal representations," carefully designed by the programmer to permit the manipulation of bodies of externally-acquired knowledge.

3. The subset "external representational objects:" material artifacts, noises, utterances, physical gestures and other items which are projected into the world by the individual. For present purposes this subset will be exemplified particularly by "visual representations"--paintings, drawings, diagrams and so on--which refer to the appearance of the world or refer to the world in terms of its appearance. Any observation made about these items should transpose easily onto other examples in the subset.

What a Program Models and What it Doesn't

With regard to the relationship of 1 and 2, if the first order of business for the programmer is the performance of knowledge-based tasks, as opposed to the examination of cognitive performance itself, he has little a priori reason for wanting to emulate human-internal representation schemes. The program's human counterpart--the expert who serves as a knowledge source for the design of an expert system--functions as the program's prototype only in the limited sense of "owning" the tasks and possessing the knowledge with which the program has to deal. It follows from the fact that machine-internal representations are designed to perform machine-specific manipulations that the mechanisms of prototype and model cannot correspond in detail. Any confusion about the status of a machine-internal representation in respect of its prototype creates real difficulty in characterising the prototype. It is difficult, for example, to use a hierarchically-organised model for any time without losing sight of the fact that hierarchical organisation is a property of the model rather than of the prototype. The danger of confusion is endemic here: a program does model those aspects of a prototype which are believed to be of consequence, and it is a representation--the only one available--of the prototype. It is inevitable that it should seem to model more than it actually does (see "the completeness illusion" below).

Representations and Transformations

With regard to the relationship of 2 and 3, it is a common confusion to regard the artist as an essentially passive component of

the representation process, through which information is channelled from the outside world onto the canvas. Obviously, this view is quite at odds with the view of representations and representation-building being advanced here. However, the confusion cannot simply be put aside. The fact that it is so widely-held tells us a good deal about the cultural pressures that act upon conceptualising: for example, it reflects a deeply-held belief that all processes can be adequately characterised as transformations. Thus, the artist as representation-builder is seen through the ubiquitousness of photographic imagery in the late twentieth century. The photographic process is a simple transformation of light-energy from the real world, through the lens of the camera, to grey levels on a sheet of film, independent of any association of that light-energy with things. It is true that several hundred years of chemistry-less "photography" preceded the invention of the photographic plate, and that, as a result, photography appears to have inherited the representation-building functions of art. It seems altogether reasonable to believe that the individual represents the outside world as the camera does. However, the camera's absolute indifference to what is before it lacks any counterpart in the human cognitive system. As a paradigm for what the artist does, it is thoroughly misleading.

The confusion of "transformation" with the more general "representation" becomes a major source of difficulty in understanding human representation-making, and it will become increasingly troublesome as we consider the workings of the internal representation-building processes in finer grain. A transformation in this sense--if we want to insist that all processes must be transformations of one sort or

another--is a black box; material goes into one slot, and reconfigured material pops out of another. Considered as a "processing unit" of mental activity, the problem arises that the transformation is fundamentally fixed by the "shape" of the input slot, as it were. Material has to be appropriately formatted to get into the transformation, while we will be seeking mechanisms capable of reaching out and grabbing material.

If we lay aside the view of the artist as a black box, passively transforming light from the outside world--but not the outside world itself--into patches of color on a canvas, it becomes obvious that the building of external representational objects must be continuous with internal representation-building and must be presumed to advance that internal process in some way. Cezanne, for example, drew not to "show" the world, but to further his understanding of its structure and its appearance.

What Do External Representational Objects Represent?

Elsewhere in this paper, I advance the view that all representations represent lower-level representations, that internal representation-building is an essentially free-running process in which memory data are reconstituted, given form, for purposes of checking against the external world. In this view, the making of an external representational object serves to "fix" the highest--but essentially transitory and incomplete--level of internal representation. In relation to visual representations, we might consider this "highest-level internal representation" to be a "mental image" viewed by the

"mind's eye." (The inevitable question, "what are memory data like?" is discussed in Part 4.)

Machine-internal representations and external representational objects--2 & 3--may now be seen as essentially equivalent items in relation to human-internal representations. In both cases, the forms taken by these representations answer to a number of considerations, among which technological considerations--considerations of "medium", for example--figure large. And whatever other practical application either of them may have, they both serve to advance, to fix, the internal representations of their owners, the program designer on the one hand and the artist on the other. What this appears to mean is that the purposive, task-specific computer program may be the result of creative activity on the part of the programmer, just as a painting may reflect creative activity on the part of the painter. But, like a painting, it cannot be expected to act creatively. Unless we can conceive of a different "kind" of program, one which models intellectual processes rather than seeking only to accomplish human intellectual tasks, the satisfaction of particular goals may keep computers from ever manifesting creative behavior.

Examples of Purpose and Constraint

Representational technologies are specialized in their usefulness, just as other technologies are, and not all purposes may be satisfied by the application of any particular technology. For example, perspective drawing is capable of dealing with the appearance of 3-dimensionality, but it cannot carry dimensionally-explicit information of the 3rd

dimension, for the reason that the making of the drawing from a single viewpoint precludes the acquisition of that information. Conversely, the architects's plan-and-elevation drawing of a structure carries explicit information concerning all three dimensions, but, since it does not involve a single viewpoint at all, it cannot convey anything about the structure's appearance.

Statistics is a technology which provides a representation of the large-scale behavior of a complex system, but without representing the behavior of any single unit within the system. Black-and-white photographs do not represent coloration, because the photographic emulsion is not color-sensitive. A still from a movie represents the positioning of actors within the frame at a given moment, but not their movement, in time, through the frame. A recipe in a cookbook represents the processes by which a gourmet dish may be prepared, but the language in which the recipe is written possesses too few taste-descriptors for the recipe to represent the taste of the dish. A plan for how to spend a day running errands does not represent how one <u>will</u> actually spend the day, for the reason that the plan cannot include <u>all</u> the determinants. And so on and so on.

<u>Delimiting Constraint, and the Limits of Delimitation</u>

What all these examples show is that the ability of a technology to satisfy any particular purpose is quite tightly constrained by intrinsic properties of the technology itself, and no degree of insistent purposefulness on the part of the individual will loosen the constraints. On the other hand, the possession of a range of

technologies by the sophisticated individual allows flexibility, in representation as in other things. For example, as an alternative to both plan-and-elevation and perspective drawing, isometric projection allows some reference to appearance at the cost of some loss of dimensional information. To some degree, also (see "representation and culture", below), an existing technology may be modified by the individual who understands its structure, i.e., one who possesses a theory of operation. For example, we might consider the history of Western painting from 1400 on as a series of modifications to representational technologies--resulting in the pushing-back of the painting's "sky", in simple terms--that parallel the culture's changing view of man's relation to the physical world.

Thus, while representation-building is clearly technology-constrained it is also purpose-specific, in the sense that choice of technology may be dictated by purpose. For example, the differences between the internal representation which precedes a verbal utterance, the utterance itself, and the written form into which that utterance may eventually be cast are determined by differences in the purposes associated with the three forms. We should not leave this point, however, without stressing the extreme difficulty of conceiving of a purpose for which a technology does not already exist in the individual's repertoire. The possession of a technology--this is true of any technological domain--provides the individual with a handle on his world, and the more effective a handle it is, the more unlikely the individual is to relinquish his grasp on the handle, and hence on his world. Very few individuals ever do it. In short, the technology does

not only constrain the forms of representations, it constrains the individual's grasp of what is representable: i.e., his conceptual powers.

The Uses of External Representational Objects

All of the examples in the above paragraphs may be construed to mean that the use of a particular representational technology constrains the individual's ability to express his purposes to the outside world, i.e., to communicate. Of course, that is the case, but not the whole case. Beyond the habitual satisfaction of everyday requirements, most individuals enter an externalising mode only occasionally, and we all spend most of our mental lives in a wholly internal "self-communication" mode. Just as we only know what an individual "means" through communication, i.e., by the external representational objects he makes, the individual in this mode only knows what he "means" through the internal representations he is able to make.

All this suggests that externalising is an extremely important phase in representation-building, aside from its communicative aspects, because it affords the individual a more coherent view of what he "meant" than is given by the internal representation. It seems quite unlikely that the individual could much advance his understanding of appearances, for example, solely by the formation of "mental pictures," and neither he nor anyone else could know whether he had done so.

Representation and Culture

The forms we find in externalising technologies in some degree reflect internal mechanisms: i.e., cognitive propensities. However, the acquisition of these technologies reflects mechanisms of a different kind. Like the acquisition of any other "personal" characteristics, this one takes place in a cultural environment and is determined in large part by what the culture finds desirable.

The observations made above concerning the constraining influence of technologies and the modifying effect of purpose apply to cultures just as they apply to individuals. That is to say, we would hardly expect a culture to develop a technology which did not rest upon cognitive propensities, but, at the same time, we would expect different technologies in different cultures, subject to the same propensities but driven by different cultural considerations. And that is, in fact, the case. For example, although the ability to differentiate between fine gradations of light and shade is a fundamental property of the visual apparatus, we do not find that propensity playing anything like the same role in the development of technologies in different cultures. In most cultures it supports an essentially decorative patterning of the representation itself, while shading--a technology for dealing with the illumination of the surfaces of objects--is almost unique to Western European art, not surprisingly, the context within which photography developed.

The fact that the individual learns his technologies within a cultural context has important ramifications. The purposes of the culture, which bear upon the development of its technologies, are

directed primarily towards the unambiguous transmission of socially necessary, or desirable, material. The development of grammar rules, standardized spelling, common usage protocols, and so on, all serve this end to some extent. In any case they serve to reassure the individual that someone else's utterance means what he would mean if he made the same utterance. However, the remarks made above about the constraining of conceptual power by a representational technology also apply here. For as long as the individual functions within the range of meanings made available to him by an acquired technology, then the culture's purposes are served. But the culture's purposes may be quite at odds with the purposes of the individual who somehow manages to break through the constraints to more highly individualized and particularised meanings and, simultaneously, to modifications of available technology.

Like any other "nature-nurture" dichotomy, the fusion of these two fundamentally antagonistic purposes is vital to the culture. It is the constant rebuilding of technologies by individuals for their own needs which guarantees the renewal of meaningfulness to the culture. At the same time, the diffusion through the culture of technology-modifications, and the newly-formed meanings which accompany them, is a slow business, during most of which the initiating individual can expect very little "communication" with the culture at large. Whatever other limitations on communication arise through noise, error, and ineptitude, correctness in communication is primarily proportional to the familiarity of the meanings involved.

Evocation: the Stimulation of Meaning

"Communication" has been used in the above discussion in the commonly-understood sense that implies passage of explicit information between individuals. However, not everything that passes between individuals through the mediation of an external representation is "communication" in this sense. As we might anticipate, technologies that are strongly enculturated give the greatest likelihood of unambiguous, information-specific communication, while technologies that capitalize on inherent human capacities tend to stress their human origin. The objects of these latter technologies tend to be "evocative" rather than informative. They offer convincing evidence of cognitive activity, but do not provide the meanings which are normally anticipated from enculturated cognitive activity. In doing so, they apparently serve to stimulate the receiver's own representation-building proclivities to the point where he will provide "meanings" to support his own anticipations.

It is worth remarking in passing that both informative and evocative objects enjoy some level of autonomy once they are in the world. They are what the receiver receives, and he inevitably deals with them as if they convey what the maker "had in mind." What--if anything--the maker actually "had in mind" is quite irrelevant [2].

The Completeness Illusion

Since representational technologies are limited in what they are able to represent, it follows that representations are incomplete with respect to all that can be represented. They may also be fragmentary,

incoherent, and discontinuous.

This observation applies to the processing of both internal representations and external representational objects. The viewer is normally unaware that the face in a Rembrandt portrait has been represented by an accumulation of paint patches, just as he remains unaware how discontinuous are his own "mental images," and how little information they actually contain. We don't lose our place in the world when we blink, and the cognitive system creates for us an illusion of smooth, high-resolution visual experience out of the rapid discontinuous shuttling (saccading) of a miniscule patch of high-resolution receptors across the field of view.

By whatever mechanisms this perplexing "completeness illusion" is maintained, it is a fundamental attribute of the cognitive system and--in characterising the cognitive system in terms of representation-building--a fundamental property of representations. Thus, in this characterisation, the successive development of lower-level into higher-level representations means an increasing illusion of completeness, up to--but not beyond--the point where the representation is complete enough to function as a surrogate for the external world.

PART THREE: DRAWINGS OF A KNOW-NOTHING, ALMOST

People in AI know very well that the form chosen for the representation of a body of knowledge in a computer program is determined by functional considerations--the operations one hopes to perform--rather than by the knowledge itself. What remains elusive is the fact that all representation is similarly purpose-specific in its forms and in the choice of formal strategies. For example, a landscape drawing by Cezanne owes its character primarily to Cezanne's representational strategies, secondarily to the properties of the materials he uses, and only then to the "real" landscape. It is a central contention in this paper that the making of external representations of this sort exemplifies representation-building processes which are wholly internal, i.e., those processes by which we are able to "conjure up" images in the "mind's eye."

The Program's Aims

This imaginational function obviously implies both the existence of stored knowledge and some strategy for the recapitulation of that knowledge. ANIMS was written in order to explore the different parts played in the generation of a representation by a clearly-defined body of knowledge on the one hand and a clearly-defined representational strategy on the other and to permit the two to be modified independently. ANIMS draws animals [Figures 1-5]. As I mentioned in the Introduction, the program's strategies rest heavily upon concepts I developed in earlier work on image-generation, and, in particular, it

uses heuristics for the simulation of freehand-drawing derived from that source [2].

Artists draw in order to find out what the world is like. Representation-building is important to the degree that it provides a methodology for generating new possibilities of what the world may be like, not for its power to illustrate a pre-existing body of knowledge. The representer gets more out of the process than he puts in. In developing ANIMS, I hoped that its representations would be seen to contain more than could be accounted for by either the body of knowledge the program had at its disposal or the representational strategy it employed. For similar reasons, I felt that the program ought to exhibit marked economy, corresponding to the economy of the cognitive system itself, in its ability to generate a diverse set of representations from limited knowledge and simple representational strategy.

The Current Form of the Program

ANIMS' knowledge is essentially structural, i.e., what it knows about animals has mostly to do with the articulation and relative sizes of their parts, and prescriptive. ANIMS knows how to make "visual" representations of animals, not how to draw inferences about them. At the highest level of the program it knows that to "make" an animal, one first "imagines" a body (involving in turn a spine, a chest, a belly and a behind) then four legs, and so on and so on. Each of the animal's parts is represented to the program by a subroutine which generates a new body of data for this instance, not by a pre-existing data-structure. These parts are "imagined" as lines: the subroutines

determine where they are to start and stop, in relation to the space at their disposal and in relation to what has already been "imagined." The lengths of the various parts are all related to the length of the spine, which is initially derived from the extent of the space in which the figure is drawn. The angle of the spine determines the posture of the figure, and the angles of legs, neck and head are constrained to remain consistent with the posture.

Thus, generating the data results in the production of an intermediate representation. It is referred to here as "imagining" rather than as drawing, because what is produced is not the final representation but something roughly approximating a sort of stick-figure, a core figure, visible in most of the figures as the dotted line within the outline. It is drawn here only for demonstration purposes. The final representation is generated by "embodying" or fleshing-out this core-figure by quite literally drawing a line around it.

The Function of Meta-Data

This embodying procedure is indifferent to what is being embodied, i.e, whether it is a belly or a tail, and, in fact, it will embody anything that gets in its way [Figure 1]. However, its action is modulated by meta-data which originate in the data-generating subroutines and which are carried along into the intermediate representation. These meta-data are numerical determination factors which might be based in different cases upon any of various considerations. In this particular case, they are a measure of the level of care which ANIMS believes to be appropriate for any part, or

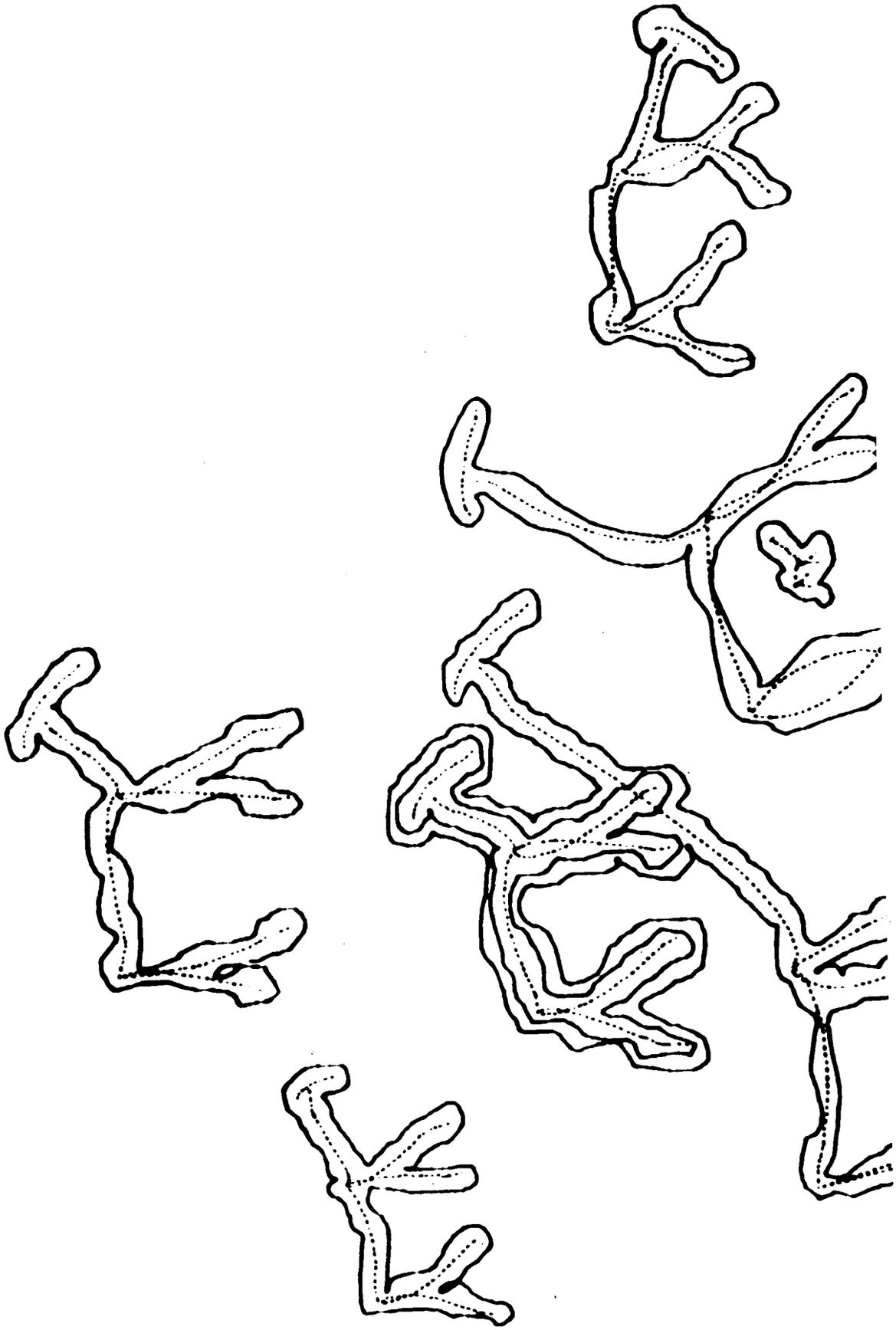

Fig. 1

alternatively as a measure of the amount of information it asserts it has for that part. (These latter assertions are spurious, of course. Since each part involves generating only the starting and ending conditions for a single line, all parts involve the same amount of information. The general issue is not exactly what the meta-data stand for, but their effectiveness as an economical way of modulating the embodying process.) If the animal's "tail" carries a high determination factor, the embodying process will respond by trying to follow the core-figure more-or-less precisely. If the "belly" signals a low determination factor, the embodying process fills in the lack by making something up. Figure 2 illustrates what happens when every part carries a high factor, and the embodying line stays everywhere close to the core-figure.

In its current implementation, the embodying program pays no attention to the lines of the core-figure. These lines are mapped onto a matrix, and once this mapping is complete, for practical purposes, the lines cease to exist. The matrix might be thought of as roughly analogous to the field upon which internal images are projected for viewing by the "mind's eye," in the sense that it carries a fairly crude, but serviceable and convincing, representation. No data are carried into the embodying program except for the contents of the individual matrix cells, which are limited to a flag which proclaims the cell to be occupied, and the determination factor left there by the occupying line.

The heart of the embodying program is a simple algorithm for traversing the boundary of any group of occupied cells. As each new

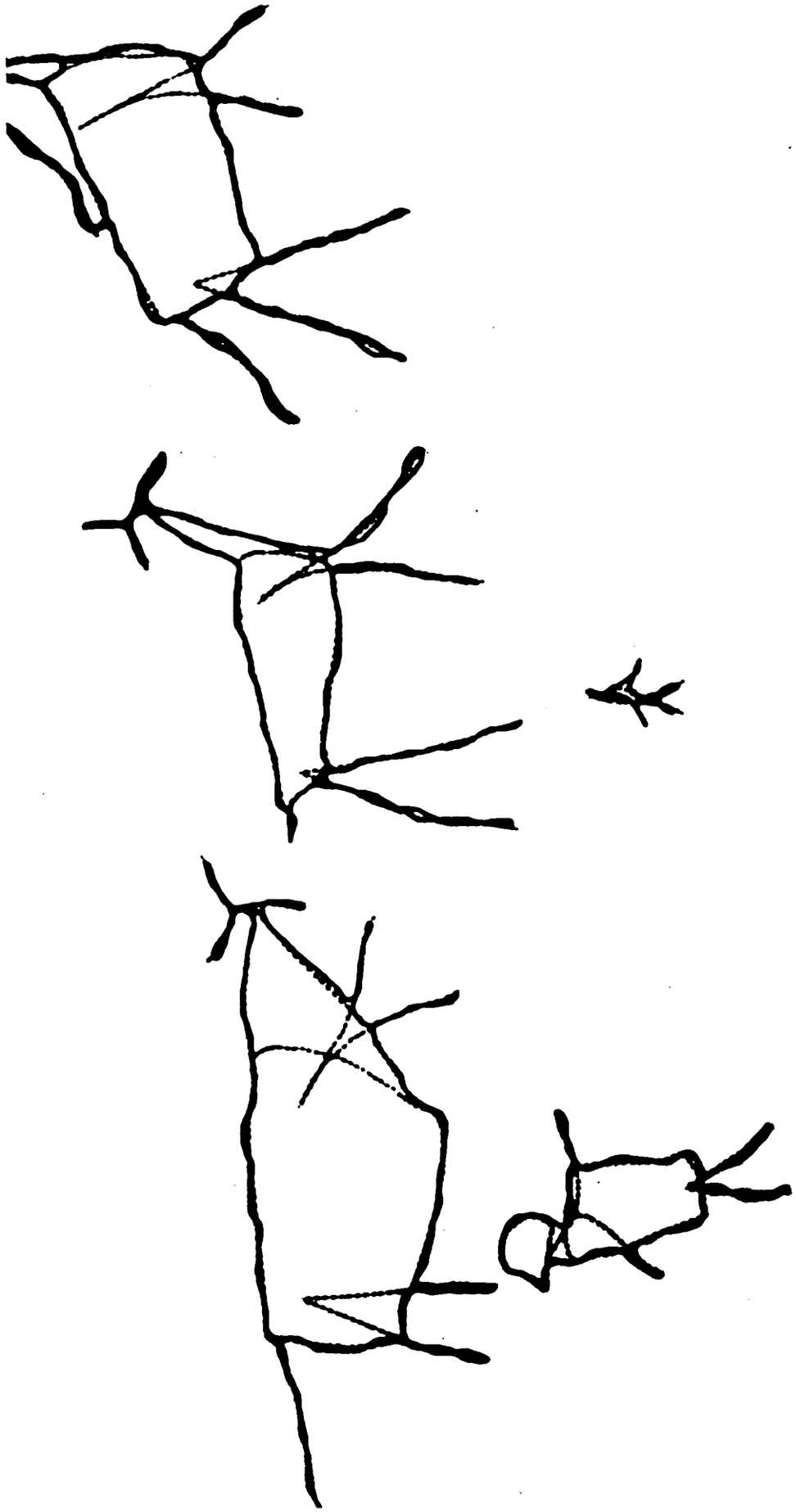

Fig. 2

cell in this boundary is located, it may be used to set a target towards which the advancing embodying line will veer. However, some proportion of these cells is skipped: the proportion being a function of the determination factor, the lower the factor, the further apart are the targets. Also, the targets are not the cells themselves, but points at some distance from the cells' centers and on the outside of the boundary. Again, the distance is a function of the determination factor. In short, high determination factors cause a slow, careful traversal of the boundary cells themselves, while low factors cause a loose, fast traversal of a set of points well outside the boundary. Figures 3 and 4 show the result of determination factors which vary rather strongly from part to part.

The Program's Development, and its Results

We are now in a position to examine the development of the program to its current state and to demonstrate some of its results. The data generated by the first version of the program [Figure 1] was more limited than we have described: the body consisted only of a spine, with neck and forelegs attaching to it at one end and rear legs at the other. The illustration is typical of the program's output at this stage. Figure 5 shows the output after the data generation has been amplified to give a "full" body. There were no changes to the action of the embodying program between these stages, and the differences between the two drawings simply reflect the change in the state of the program's knowledge.

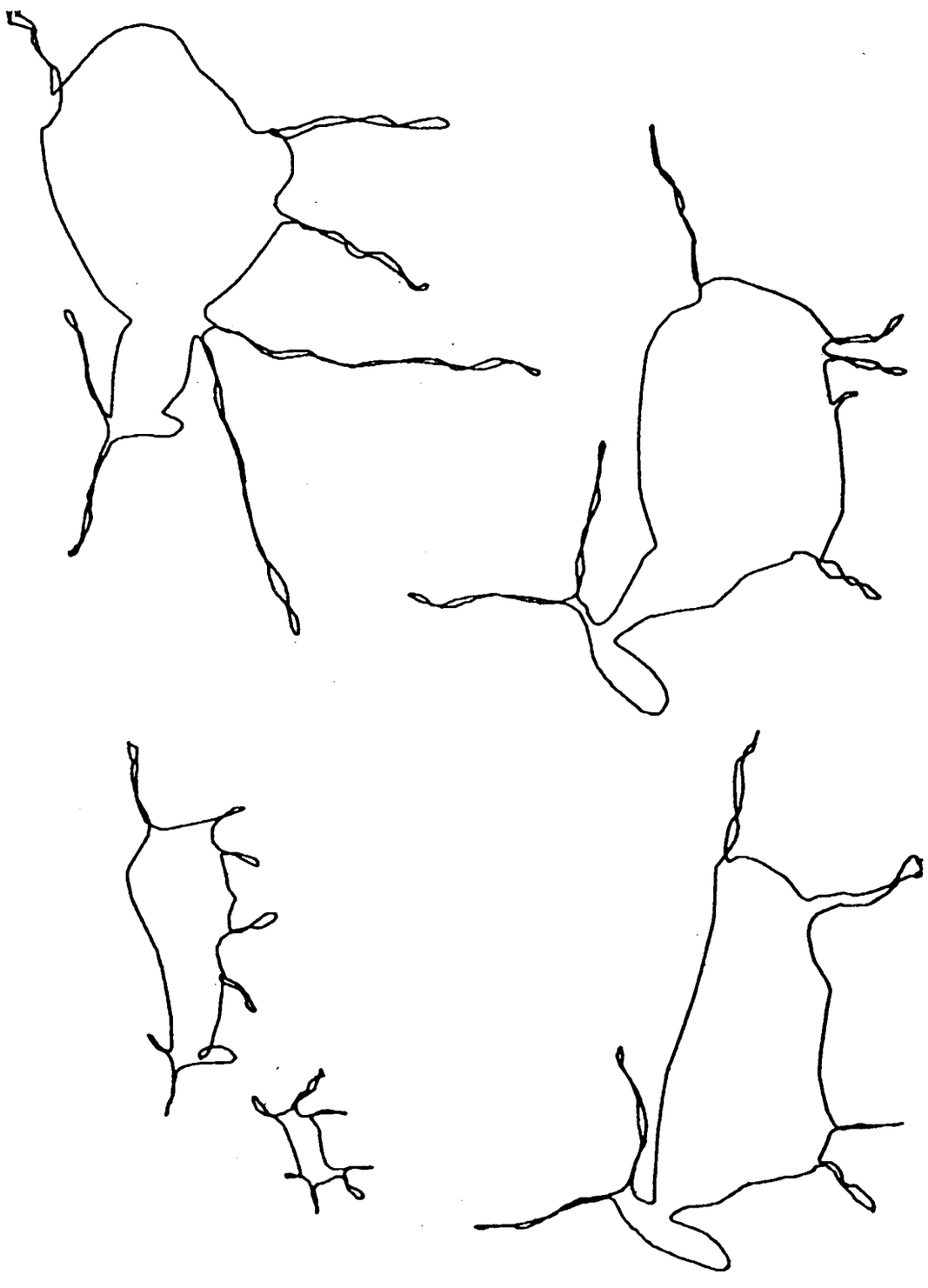

Fig. 3

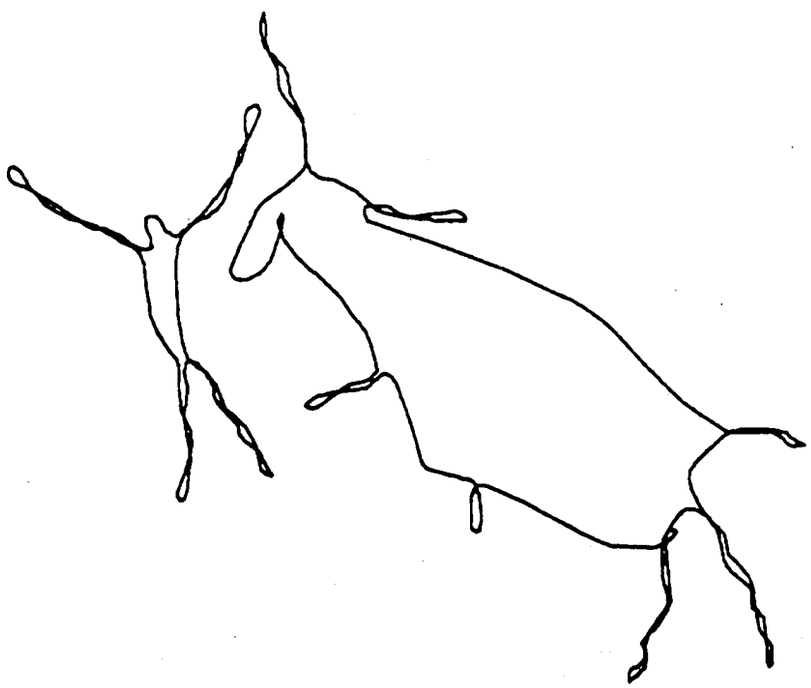

Fig.4

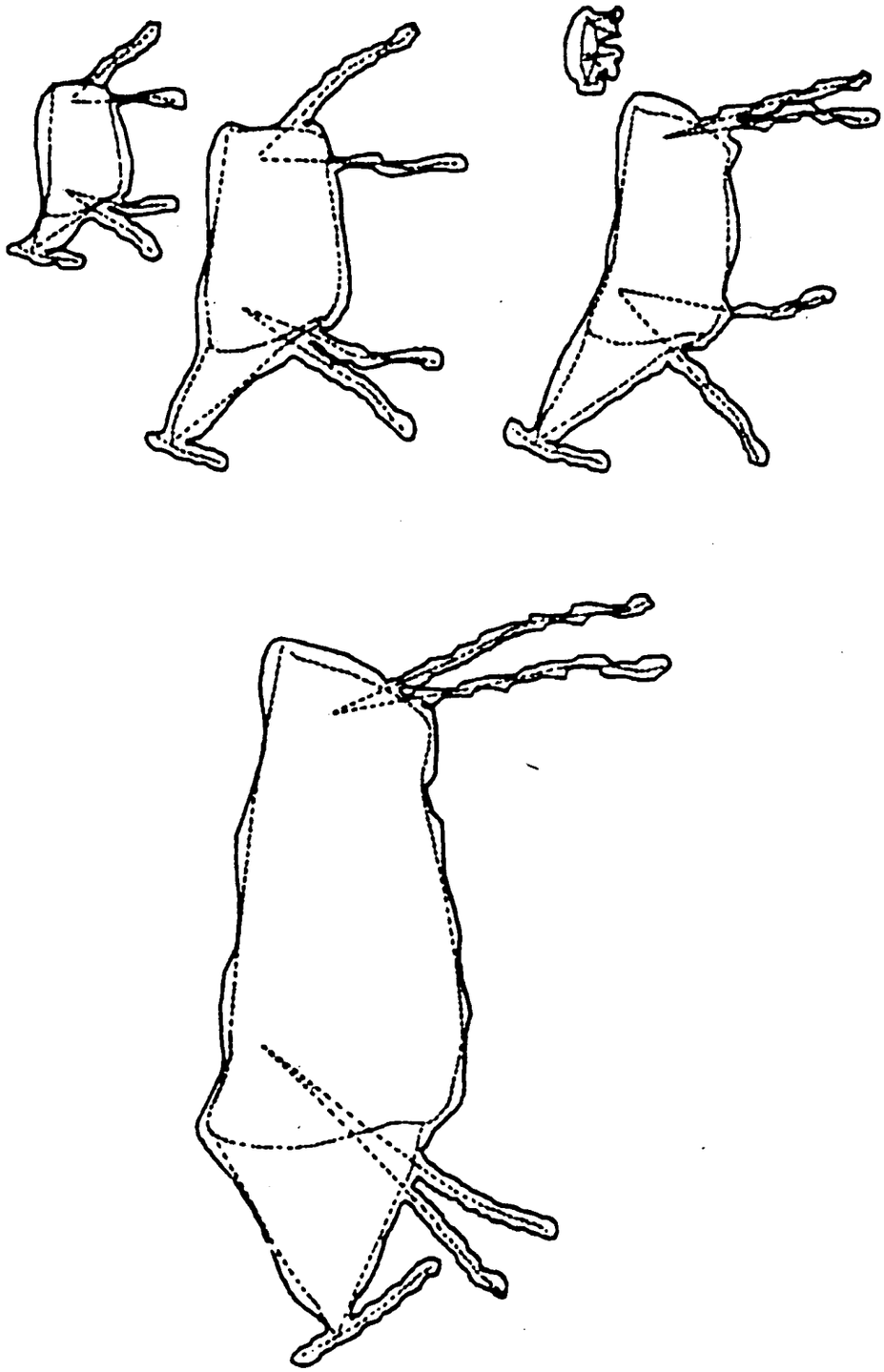

Fig. 5

Interestingly enough, the difference between the drawings made in these two stages appears to correspond quite closely to differences between certain strains in so-called "primitive" art. Figure 1 is identifiably similar to common examples of African Bushman art and Australian Aborigine art. Figure 5 is much more reminiscent of the Northern Paleolithic, exemplified in the caves of Altamira and Lascaux, with which it shares a preoccupation with bulk absent from Figure 1 and its Bushman cousins.

Medium-based Determinations to Representation.

Clearly, the representational strategy, i.e., the embodying program, has left its mark on each of the drawings. The curious jogging of the lines, particularly clear in the legs of the large animal of Figure 5, is a function of the fact that the matrix is quite coarse in relation to the size of the figure. This effect is essentially equivalent to the medium-based determinations to form found in Cezanne, Aboriginal art, or in any other artifact.

In later versions of the program the matrix was made much finer and the jogging, though it remained, became less evident [Figure 3]. A subsequent refinement in the drawings, which appears to suggest more sophisticated knowledge, in fact came about through a minor change to the boundary-finding algorithm used by the embodying program. Previously the algorithm had examined only edge-adjacent cells in its search around the figure. Including corner-adjacent cells in the search meant that a move towards the next target cell could involve a less radical change in direction, and the animal's contours became a good

deal smoother as a result [Figure 2]. No changes to the data-generating subroutines were made during this modification.

Economy

All the data-generating subroutines, together with the embodying procedure, occupy about 11,000 words on an 11/45; the entire run-time package, complete with all the space-finding, freehand drawing, and data-management stuff, stands at about 23,000 words, of which a large part is involved in "paging" rapidly growing data-structures to disk. By contrast, the line-segments generated in the drawing of a single animal--leaving aside the diversity of output of the combination--may take as much as 5,000 words. It seems unlikely that a more conventional approach to the production of so diverse an output, e.g., storing prototypes, and performing transformations upon them, could match ANIMS' economy in its use of resources. Nor is it clear that one could define transformations of adequate flexibility, or any coming close, that would not take longer to perform than the mapping of the intermediate representation and the action of the embodying program. Although I am not suggesting that these results "prove" anything about the way human cognition functions, they appear to lend substance to Marr's views [3] on the storing of visual knowledge and its recapitulation by the cognitive system.

The Program as Representation-Generator

As an analogue for anything as ill-defined and uncharted as mental imaging, ANIMS could not hope to be other than a suggestive sketch. An

interesting feature of this sketch is that knowledge is not represented anywhere in the program as static data that can be read from a conventional data-structure. Data are generated by invoking procedures and are already partial representations when they first appear. Legs, spine and tail never actually appear in the program, that is to say, only the starting and ending conditions for lines. These lines are not parts of an animal in 3-space; they are parts of the core-figure, the 2-dimensional intermediate representation. The final representation produced by the embodying program is generated from--it is by no means a simple transformation of--this core-figure.

In short, ANIMS is a program which makes representations of representations of representations. It does not proceed from the external world in, as most of us believe visual imaging to proceed, but from the inside out, while the "inside" is just a few procedures which could never be claimed to possess more than the most rudimentary knowledge.

The Free Lunch

ANIMS' drawings give the illusion of vastly more knowledge than the program actually has, however. In particular they appear to arise from an intimate visual knowledge of animals in movement, whereas the program has no data at all concerned with appearances as opposed to structure. How does the program produce jutting shoulders and bent knees? How does it cause the animals to turn their heads away, to thrust their legs towards the viewer, to rear and prance? What is the source of this "extra" knowledge? Of course, the answer is that the program does not

do those things; it makes drawings which create the illusion of those things. There is no "extra" knowledge, either in the program or in the mind of the viewer. ANIMS is an evocation machine, not a documentarist.

Mental images are not high-resolution holographs. They are certainly not more complete or coherent than external representations, and they are so elusive that it takes careful introspection to see how unstable they are and how little knowledge they actually contain. The illusion of completeness and coherence which they generate must certainly be one of their most important--and puzzling--characteristics. It is the coincidence of this characteristic with the illusion of "extra" knowledge in ANIMS' drawings which suggests that mental images may be members of a broader class of evocative images, all of which function by persuading the viewer to accept a minimum of data in place of all of the data. Cognitive economy may rest upon "satisficing" behavior [4].

PART FOUR: METAPHORS AND MODELS

The previous sections have discussed some of the factors that must necessarily arise in any theory of intellect required to account for creative behavior, or, as I have called it, enhanced intellectual performance. At this point further development of a theory must wait upon a demonstration of the plausibility of these factors by an operational program, and this program has yet to be written. (Rather than refer to "the program" constantly, it will be called on by its family name, AARON 2, from this point on.)

4.1 The Limits of Metaphors.

Any program designed to elucidate some aspect of intellect is an operationalised metaphor, and if inappropriate technological limitations are to be avoided its structure should arise from the metaphor, not from prior programming practise. That would be the case, for example, with respect to a program conceived in transformational terms--the program as "algorithm acting upon data"--in the sense of the word previously discussed. Even in a simple program like ANIMS, as indeed with so many other programs aimed at modelling intellectual performance, the distinction between data and control structures began to blur, as it was bound to do, if only because so much of the knowledge held by the human is knowledge of how to do things. The aim in this section is to draw some inferences from the preceding parts, in broad terms, about what AARON 2 might be like, and to develop a fuller metaphor for the operation of intellect as a basis for subsequent considerations of program design.

The Case to Date

To summarize the position so far: intellectual performance has been characterised as a free-running and continuous process of representation-building, internally-driven in the sense that each representation represents some lower-order representation and not some aspect of the external world. The external world is involved in this generative process, however, and the process is purposeful, in that representation-building enables a coming to terms with the external world. Representations are developed internally until they generate an adequate illusion of completeness to allow their "direct" comparison with external-world data. The building of external representational objects is continuous with the wholly-internal phases of the process, and it serves to fix the fragments of the shifting internal representations into a more substantial and less transitory object. (This view offers a powerful explanation of the apparent prodigiousness of the mind as a data-processor: the mind does not process more than the minimum amount of data required to construct the normally sketchy representation.) Since representation-building is so constrained by representational technologies, the view also implies that differences of representational technology from individual to individual may account for differences in intellectual performance.

The World Model as External World

The "external world" in the above paragraph should not be understood to imply a complete dichotomy of internal and external events. The "real" world outside may be the individual's primary source

of data--primary, that is, in terms of his experience as a whole--but it is not easily separable, in terms of his day-to-day awareness, from existing internal representations of previously acquired data. Nor does the acquisition of new external data make heavy claims upon the individual's resources. Thus, the outside world should not be seen to constitute a source of special considerations; what is central is a matching process in which representations match something, whether that "something" is a car driving towards one on the freeway or an unsolicited recollection of childhood.

Making the Best of Things

Individuals reason, plan, search for items of memory, and strive to conjure up mental images. When they are not so deliberately engaged they are ruminating, free-associating, and spontaneously generating mental images. In terms of the matching of representation against data, the primary (perhaps only) distinction between these two kinds of activity is purpose. The purposes of the former may be deliberate and overt while the purposes of the latter are not. As in the case of theory-building in the sciences, within particular disciplines, the deliberate purposes served by representation- building may require stripping off the products of spontaneous activity as irrelevant. This elimination of "irrelevancies" occurs at a late stage of representation-building, but the illusion is created that the complete suppression of spontaneous processes, so readily achieved in computer programs, is both possible and desirable in the individual. It is neither possible nor desirable, and in relation to the modelling of

creative behavior, it is no more desirable in computer programs.

Consider as a crude model of overtly purposive internal activity, say planning, a brainstorming session among a group of experts. Each expert represents a single "voice" in the individual's internal dialogue. One expert makes an assertion; a second responds to what he believes the first to have said; a third adds his own comment. Self-watchful participation in situations of this kind quickly reveals that at least three factors prevent "complete" and undistorted communication. First, utterances are themselves external representations through which individuals seek to grasp their own intentions: they are not synonymous with these intentions. Second, other individuals cannot apprehend these utterances other than through the mediation of their own prior mental states, which will be different from each other's and different from the speaker's. Third, no individual apprehends all that is being said, because he spends part of his time and his mental resources doing something else: daydreaming, pondering a previous utterance, thinking ahead, running his own spontaneous representation-building processes. Even allowing for redundancy, it cannot be believed that unambiguous transfer of data is a prerequisite for productivity or even that it is possible.

It is not being claimed that tower-of-babel conditions are the source of enhanced performance. It is clearly desirable that each expert represent his position as correctly as he can. But what is at stake is not that position "itself," but the representations of (the representation of...) that position which the other experts form. Thus, the brainstorming session would not necessarily be more productive if

the participants could stick precisely to the topic, could express themselves entirely without ambiguity, and could discard all prior assumptions.

Since the "absolutely correct representation" of a position is, like Thurber's unicorn, a mythical animal, I may seem to be arguing that tower-of-babel conditions are inevitable. They probably are, in an absolute sense. But it doesn't matter. What is actually being argued is that up to a certain point, the incompleteness of representations, and the lack of perfect communication, are advantages, not disadvantages. It is through them that the domain of discourse is enlarged beyond the limits which would exist if all communication were impeccable.

As I have said repeatedly, a theory of creativity should be simply a theory of intellect which accounts for creative behavior in terms of normal resources. If the observations arising from the "brainstorming" model are now re-applied to the modelling of the individual intellect, it may be concluded that the conventional task-specific computer program lacking free-running capabilities is actually modelling an entirely mythical beast: it cannot account adequately either for normal, or for enhanced, intellectual performance. (Nor is it necessarily claimed, within the AI community, that it does: witness the shift away from the earlier preoccupation with human intelligence to a more catholic view of intelligence as such, referred to earlier. The conclusion that programs do not have to be intelligent the way humans are intelligent seems very reasonable, but it may make the designing of programs more difficult rather than less so. This mythical beast has no prototype.)

Spontaneous Processes and Deliberate Processes

It follows from all this that AARON 2 should be free-running, but we should be quite clear about what that means. We are not distinguishing between deliberate processes on the one hand and spontaneous, free-running processes on the other, we are distinguishing between overtly purposeful free-running processes on the one hand and not-obviously purposeful, spontaneous, free-running processes on the other. Free-running means, essentially, that things are not under the control of a homunculus.

It also follows, then, that AARON 2 should exhibit both deliberate and spontaneous processes, though pragmatic considerations of data-acquisition would seem to demand that the stress should be upon the spontaneous in the first instance. When it is subsequently developed to further model deliberate, task-oriented performance, it will be extremely important to maintain the tightly interwoven structure of deliberate and spontaneous processes. The significance of the "brainstorming" model is that some of the elements brought forward into the higher levels of representation are "counter-purposive" in terms of deliberate purposes: they are, quite literally, irrelevant to the particular task in hand, but serve more generally in coming to terms with the external world, and thus to extend the individual's conceptual domain. This will occur through two features. The first is the close binding of internal representational elements by virtue of their close acquisitional association and by virtue of their close association in subsequent representations. Secondly, there will be elements that result from the exercise of representational technologies which are,

inevitably, imperfectly matched to the purposes they serve.

These considerations might be summed up as a strong emphasis on the "associative" character of memory. But "association" is an ill-defined term. Before the particular use of the term here can be described with sufficient precision for it to be operationalised, its context--mental activity as a whole--will need to be established in greater detail.

Experience and the Anatomy of Memory

If all internal representations represent lower-order representations, a number of questions appear to require answering. What precedes representations? Are there "memory primitives" from which representations are reconstituted? What is memory "like?" My position is that these questions are, in practical terms, meaningless. Without a great deal more fine-grained knowledge of the brain and its neurological functioning, there is no choice but to discuss the mind metaphorically, if it is to be discussed at all--although no metaphor sustains itself across the boundaries of a particular domain of enquiry. (An excellent example of the purpose-specificity of representations!) Memory isn't really "like" anything, and we can only know, more or less, what it does.

The acquisition of data, and its presumably permanent storage, begins in the individual with sentience. In the sense that, in gross terms, the medium of memory has to be the brain itself, and that experience has to involve changes in the physical state of the brain, at least some levels of storage mechanisms have to be innate. It is assumed, however, that storage structures are not innate, but develop in

an ad-hoc fashion as data acquisition proceeds, responding both to new data from the external world and to internal data arising in the building of representations. These structures are unlikely to be conveniently uniform, and we should expect any machine-dependent equivalent to be baroque: they will be representations already reconstituted to varied levels of incompleteness, and they will provide for their own binding into more developed representations in a number of different ways.

Consciousness as a System Characteristic

To risk a spatial--and dangerously visual--metaphor, the state of an individual's memory at any point might be likened to a cross-section of a densely-branching bundle of fibres, each fibre having been initiated at some arbitrary distance back from the cross-section. The grouping of those fibres into smaller bundles, the representations, is what we mean by representation-building. More precisely, the sum total of the strategies that determine the changing grouping of the fibres along the bundle is what we mean by representational technology. There, the value of this particularly visual metaphor ends. It is useful in showing memory as the history of the individual's states of consciousness, but if it appears to imply that the cross section is seen from some other place, its danger is that it vests the central function of consciousness in that old homunculus, the see-er. The intent of the metaphor, on the contrary, is that the cross-section IS the state of consciousness. It "is" a double-sided screen whose function is to resolve the discrepancies between the states of its two sides: the

newly-generated representation on the one side and the incoming world data, or internally-generated data, on the other. In more orthodox cognitive terms, the "distortion" of external world data in the direction of what we "know", i.e., in the direction of the internal world model, and the modification of that model by the introduction of new data occur simultaneously. The resolution of discrepancies is a single function.

PART FIVE: CONSIDERATIONS OF PROGRAM DESIGN

We are now in a better position to identify a number of target areas with which AARON 2 will have to deal. No attempt will be made to consider its design in detail, and the following discussion should be viewed as no more than a loose program specification.

States of Mind, and Program States

It is a pre-eminent consideration that the acquisition and storing of data in the individual is essentially experiential: a lifetime of experience has contributed to the state of an adult mind. Obviously, a tabula rasa program that builds its own storage structures from scratch is not a practical possibility. AARON 2 will be brought into existence in some arbitrary state, and the experiential quality of its prototype will be reflected if free-running representation-building is seen to result in changes of state.

What is intended by "state" is not the state of the program's memory, however, and "changes of state" should not be understood, in conventional terms, as the writing of new data into a pre-existing data structure. It has been noted that the individual's state of mind is not likely to be adequately represented in a program either by a data-structure or by a collection of algorithms, but by something that is neither or both. While it is unclear what that something should be, in AARON 2 the technological functions which generate new representations will live in the same space as the representations themselves; or, to get a little closer, they will be elements of the representations. That

representation space is what is meant by the "state" of the program. What we might in a more orthodox frame think of as "accessing data" is equivalent to entering that space and initiating the generation of a new, higher-level representation. The path that representation will take will be a function of the purpose at which it is aimed.

Creativity as a Function of Technological Diversity

It will have to be recognised that the state of mind of any individual is characterised by far greater diversity in its purposes, and in the forms of representations and representational technologies, than can be approached by any computer program. That doesn't stop one writing computer programs: representations have no claim to completeness. In this particular case, however--in parallel with so many cases involving the modelling of higher intellectual functions--it might reasonably be argued that creative behavior is a function of this diversity itself, and consequently that creativity is absolutely beyond the reach of a program. It would then have to be conceded that AARON 2 will not behave creatively, but that it should, nevertheless, exhibit some of the elements of creative behavior in individuals. The argument might be academic at this stage: it is clear, in any case, that the selection of a domain of mental activity as a working environment is unusually critical. Not only must it be constrained enough to provide tractability, but it must also be sufficiently central to mental activity as a whole, sufficiently characteristic a part of that activity, to make the exercise significant.

It is also clear that the domain must contain more than a single technology, or at least that a predominant technology should be augmented by more "general" material. If memory is simply the history of the individual's representations, and if the building of representations is specific to a number of different purposes, then it follows that the building of a new representation is carried forward from a base of partial representations which were not necessarily specific to the present purpose. At least some diversification of the program's purposes, expressed in its representational technologies, is important with regard to "association" and the part it plays in enhanced performance. (The foregoing assertion that technologies exist as elements of representations, and in the representational space, allows us now to redefine "association": it may come about either through the carrying forward of inappropriate material or through the selection of the "wrong" technology.) As ANIMS has already shown, some of the necessary diversity is given by the fact that lower-order representations are, by definition, less purpose-specific. Representations dealing with appearances can be generated out of "general-purpose" structural representations, provided that these lower-order representations can appropriately modulate the action of the technologies which are brought into play.

The Cognitive Basis of Visual Imagination

Building on what was learned in ANIMS, AARON 2 will model what we might loosely call "visual imagination" as its primary domain: that part of mental activity which results in the illusion that we "see things"

inside our heads. The choice has a number of advantages. It is a function that all people apparently share to some extent or another and one that figures to some extent in most mental activities. Artists are prone to talk about "visual thinking," reflecting an unusually heavy stress upon visual data, but any domain involving morphological considerations exhibits a similar stress: the double helix [5], for example, was not conceived without the participation of visual imagination. Thus the domain appears to have the necessary quality of centrality. There is further advantage in the fact that the exercise of visual imagination in some significant part of human populations has always given rise to tangible and relatively unambiguous forms of externalising. Consequently, there exists substantial evidence of the technologies individuals have employed in making their external representational objects to elucidate the internal processes. At the same time the opportunity is provided to model the externalising phase itself, to go beyond the wholly internal and have AARON 2 generate its own external representational objects. In all these regards, it is to be hoped that my experience as a "visual imagination specialist," as an artist, will provide useful direction.

Cognitive Functions

This selection of "visual imagination" then determines the major part of the technologies with which AARON 2 will be provided: they will emulate the operation of the human visual cognitive system. Thus, in addition to the cognitive primitives already exploited in the earlier AARON (figure-ground discrimination, open/closed discrimination and

insideness discrimination), this program will use three important cognitive features inferred from common technologies in the making of external representations. I describe these now in greater detail.

1. Line Versus Value.

First, the equivalence of line, as an element of the representation, for tonal (value) discontinuity, as an element of the visual field: the substitution of the one for the other must certainly reflect the dual function of the optical system as tonal discriminator and as contrast amplifier. The geometry of this substitution, for example, the ellipse-as-circle-in-perspective being fatter than the retinal image of the circle, provides a large measure of compensation, in the cognitive system, for the loss of the third dimension.

2. Occlusion.

Secondly, occlusion: almost everything in the visual world is partially overlapped by something else, and the individual is provided with powerful clues as to the nature of the physical world through familiarity with the configurations of occlusion. These configurations are invariably represented in external representational objects--in those cultures which deal with them--by 3-line junctions of various characteristic forms. They carry more information than the lines which join them, and allow those lines to function connectively for very little processing cost.

3. Spatial Distribution.

Thirdly, what we might call the "attention-zoom" effect: the cognitive system is able to force objects to take up a variably large portion of the attention field, even though the space occupied on the retina is fixed. Presumably this occurs as a function of the scanning feature of vision. The effect is present also in the imaginational field, as one may test by generating a vivid mental image and then asking oneself questions about its parts.

One important function of this effect is to permit the cognitive system to maintain a fairly constant continuum of scale relationships (relationships of small things to large things, small spaces to large spaces) by adjusting a threshold: the scale of tiny flowers to small pebbles to large boulders on the floor of the desert may be quite like the scale of people to cars to houses in a cityscape. The existence of a spatial-distribution constant against which external distribution data may be reconciled suggests itself as a powerful mechanism in coming to terms with the world.

These three technological features (the last is advanced quite speculatively) have in common their ability to generate the "completeness-illusion" discussed in part 2. In each of them, the image is purposefully developed to permit matching against particular aspects of the external data. The line-drawing, whether on the paper or in the mind, stands for a physical object, and the distribution constant stands

for the visual field. And they do so adequately, until more precision is demanded with respect to particular aspects of the external world. We are not normally aware of the information-bearing limitations of line drawing, for example, until some pressing demand arises in relation to the reflectivity of the surfaces of objects.

Binding and Loosening

These features belong to one broad aspect of the representational technology, that part which is involved in the carrying forward of mental states, the progressive development of higher-order, i.e., more purpose-specific, representations out of lower-order representations. Its counterpart is what provides the binding of the elements within representations, and without which the building of higher-order representations would be forced always to begin with the reconstitution of tiny fragments. ("Binding" is used in the sense of the elements "fitting" each other, having come into being at the same time and in the same space.) In any given individual, these two sides of the technology may be anywhere between complementary and mutually antipathetic, and it may be guessed that at least one central element of creativity rests upon the nature of the balance in which they are held. The binding of existing representations does have to be broken if the individual is to be capable of re-building his internal world, while, on the other hand, the binding has to be firm enough to maintain a coherent world and to permit the carrying forward of "inappropriate" material. It is to be presumed that many different technologically symbiotic arrangements occur in the human prototype. The interaction, in ANIMS, between the

data-generating procedures and the embodying program through the agency of the "information-level" feature is an example of such a symbiosis in an existing computer program, but certainly not the only example possible.

Feedback and Adaptation

Nothing will be said here about the mechanisms represented by our two-sided screen seeking to reconcile the discrepancies between the states of its two sides. The screen must obviously move freely in the representation space and only occasionally present one of its faces to the external world. The implication is that it is always somewhere. It is, in fact, the representational plane, and it is driven forward as a function of reconciliation.

As an issue of program design, there would appear to be a face-value requirement for the feedback provided by this mechanism. If the program generates a representation, whether external or wholly internal, it in turn has to be reconciled. How else could the state of the program change as a result of its own action? It is not clear how a feedback path of this sort could be modelled currently, unless with the assistance of a human collaborator: the programming task involved would be simply intractable in terms of current computing resources and programming technology. However, the thrust of the program would suggest that major emphasis should be placed on the state-modifying action of representation-building, rather than upon post-hoc consideration of the newly generated representation. Otherwise, the inevitable question remains: who is doing the considering?

Epilogue

The reader will be aware that we have now reached the end of a paper on the modelling of creative behavior without ever having explicitly defined what is meant by "creative." There seemed to be no choice but to develop a general model of mind, and to show that it would not deny the possibility of enhanced intellectual performance, rather than to build a specialized model based upon ill-defined concepts. It is, after all, exactly as difficult to ask what creativity means in a model as to ask what it means in the individual. Having said that, it would seem necessary to acknowledge a question which remains, so far, unanswered. How will we know whether AARON 2 is behaving "creatively"?

The answer is that this paper, and the program which follows it, are offered as a part definition of creativity, not as the recapitulation of evidence we recognise unequivocally when we see it. AARON 2 will behave the way it behaves, and it only requires the belief that creativity is a fundamentally human characteristic to deny that AARON 2's performance has anything to do with that characteristic. But it should behave differently from the way more orthodox computer programs behave. It should generate output which is unpredictable, not in the sense that it could not have been predicted from the state of the program, but in the sense that it did not arise from the initial state of the program. It should exhibit adaptive behavior. And, to the degree that it is involved with the manipulation of technologies arising from the human cognitive system, it should generate representational objects which are compellingly "visual." Perhaps more important, in the long term, it should produce persuasive evidence of the possibility of

addressing computer programs to a range of tasks which, like creativity itself, now appear to be fundamentally beyond their scope.

BIBLIOGRAPHY

1. Paraphrasing Herbert Simon, in the panel on the history of AI at IJCAI 6 in Tokyo: "there is no reason why machines should have to be intelligent the same way humans are intelligent."

2. "What is an Image?" Harold Cohen: Proceedings of the 6th International Joint Conference on Artificial Intelligence, 1979.

3. "Visual Information Processing: the structure and creation of visual representations." David Marr: Proceedings of the 6th International Joint Conference on Artificial Intelligence, 1979.

ACKNOWLEDGMENTS

My wife and colleague, Becky Cohen, has devoted years to curing my addiction to semicolons and a number of more trying intellectual habits. My communications analyst, Joyce Peterson, has been valiant; without her patience, I would have fallen easy prey to more varieties of exotic syllogisms than I knew existed. Mary Shannon has filled my innocent text with all manner of incantatory signs, as a result of which it ended up looking as if I actually knew what a text should look like. Doris McClure never failed to smile when I told her for the umpteenth time that I couldn't handle more than three screen editors simultaneously.

To all of these I owe that gratitude which is due for the easing of problems and soothing of frayed nerves. To Frederick Hayes-Roth I owe a special debt. It is literally true that this paper would have never seen the light of day without his constant encouragement, and his belief that I was On to Something, maintained through many months despite all the garbled evidence, up to but hopefully not including the final draft, that I could produce to the contrary.

My thanks to The Rand Corporation for providing the wherewithal to make this study possible.