AN OUTLINE THEORY OF ART ON CYBERNETIC PRINCIPLES

An outline, based on the principles of cybernetics and effective procedures, of minimum conditions for authorship and its products, and for distinguishing those that may be denoted as art.

M.J. Rosenberg.

This thesis is submitted for the degree of Doctor of Philosophy, Brunel University, 1976.

PAGE
NUMBERS
CUT OFF
IN
ORIGINAL

ABSTRACT

The object of this study is to draw art into the common net of organization, along with those other enterprizes more commonly associated with the exercise of intelligence. The method chosen for this is based upon the idea of effective procedures, namely by setting out to construct a (notional) 'art machine'. The argument falls into two parts, the first dealing with the general concept of authorship and the second with its products. Part I offers a definition of an abstract, rudimentary productive process and describes its observers. There is an examination of the relation between structure and purpose, which moves towards a general definition of authorship made in terms of extracting order from a surrounding. Principles of order extraction are examined, with particular reference to the Law of Requisite Variety. Examination of extracted order, as structure, heuristics and the like, leads to discussion of the transmission of purposes between purposeful systems, as well as general problems of constraint, and of regulation and control. Part I ends with a proposal for a paradigm for a rudimentary mechanical author. Part II concentrates on the products of authorship, seeking characterizing features of those that may be classified as art. There is discussion of objective knowledge and its value and of the characteristics of experience as a form of objective knowledge. It is suggested that art is concerned with experience and that this dictates its method, which is to produce simulation procedures based on a language constituted by the synthetic structures discussed in Part I. Lines are suggested for realizing an 'art machine' and there is a review of prospects. A section of notes consisting of speculative ideas and empirical applications connected with the conclusions of the text follows Part II.

PREFACE

Almost no literature considers directly the relation between cybernetics and art. Philosophy, where it has touched on art, has concerned itself principally with aesthetics and questions of art's ontological status, both of which interests fall substantially outside the present subject. Psychologists' interests in art have shown similar inclinations. Even approaches such as Arnheim's, which attempts to combine notions of entropy with gestaltist views, are largely irrelevant to the present study. Closest, though still with a quite different outlook from the one adopted here, is Moles's (1966) which approaches art through information theory.

In the absence of previous investigations, it has been impossible to provide an orthodox historico-bibliographical background to the present study. Instead the attempt has been to incorporate the basic notions of cybernetics piece-meal, as and where they have been needed, supplementing these general concepts with more recent, more specific results where they were relevant.

Mostly - notably in chapter 2, which is in a sense at the bottom of the whole argument - presentation makes little attempt to cover the traces marking the paths of reasoning that have led to the conclusions reached. The object has been to display the scaffolding, not simply the finished structure, in a belief in the helpfulness of this towards offering a fuller grasp of the argument's aims. Also with this intention, sections and paragraphs as well as chapters have been furnished with brief titles, a general synopsis has been provided, each of the argument's two parts has been introduced by a paragraph outlining its aims, and each chapter has an introductory section of two paragraphs devoted respectively to detailing its aims and providing an abstract of its contents.

The main argument of Parts I and II - the text - restricts itself as far as possible to the consideration of abstract systems in a more or less rigorous way. But the entities involved, designated by the names 'A', 'W', &c. are not always referred to as 'it'; clarity often makes 'he' preferable. Usage follows no particular rule. The reader may understand 'it' whenever 'he' refers to an abstract entity.

Applications to the real world of art of the arguments and conclusions of the text, along with other empirical considerations and also the more speculative notions that arise are dealt with in the notes that follow Part II. These notes refer to individual chapters, and are written in a way that permits the whole work to be read either consecutively from page one, or with each chapter followed by the notes that refer to it.

It remains to thank my supervisor, Dr David Stewart, to whom I am indebted for his valuable and unobtrusive suggestions, and his patient criticism and guidance. Where my own efforts only dully reflect his, he is naturally in no way to blame.

Professor Frank George's helpfulness, encouragement and imperturbable forebearance throughout, from the moment of taking on trust my proposed subject, have greatly smoothed my way. To him, to Professor Pask, and to the other members of the Department and participants in seminars, for the leavening of their ideas and criticisms, to the Department's secretary, Mrs Kilbride, for her help and always friendly welcomes, and to Mrs E R Creecy, Miss Susan Pinn and Mrs Elaine Towsey, for their concern, care, understanding and hard work in typing a difficult manuscript, I am also most grateful.

References Presentation of references is according to the Harvard system. There are two lists of references, one for the text and a supplement to this that gives the references cited only in the notes. Where reference throughout is always to the same work of a particular author, the year of the work's publication follows only its first citation. Reference to citations re-cited is by paragraph number of the first citation.

Paragraph references are made by paragraph number; section references by chapter followed by section number. Thus for example section three of chapter seven is referred to as (7.3).

AN OUTLINE THEORY OF ART ON CYBERNETIC PRINCIPLES

TABLE OF CONTENTS AND SYNOPSIS

The argument divides into two parts preceded by a short introduction. The first part deals with authorship and the second with its products.

CHAPTER 0

ART AND SCIENCE

This introductory chapter is self-contained. Its object is to look at some attitudes to art, especially the claims and counterclaims of the worlds of art and science, rather than to furnish a background to art theory or to offer any specific preparation for the argument that follows. The wish that animates this aim is to break away from the somewhat narrow confines of traditional art theory, in order to construct an approach that will admit the theory's entry within a wider framework, not merely as an off-shoot of philosophy or psychology, but as an integral part of the study of organization. It is this that prompts the recognition from the outset of widely-canvassed, non-specialist opinion. But perhaps the ultimate justification for setting out to treat art theory within the realm of cybernetics must be the hope of benefitting cybernetics in some way, as well as art theory, and this is the spirit in which the enterprize is undertaken.

SECTION 0 ART'S PLACE IN THE MENTAL WORLD

0.0.0 The theory of art is of interest both in itself and more particularly, because it brings to light a number of questions in the broad study of intelligence. Traditionally treatments of these subjects have tended to associate with art such capacities as words like creativity, imagination and aesthetics denote, usually combining connotations of emotion in some degree; and with intelligence faculties of reason and deduction, and the pursuit of logical agrument. Ironically it has been left to the study of 'artificial' intelligence to display the arbitrariness of the distinctions that these various terms imply, and reveal in a practical way the inescapable intimacy that unites the different elements they signify in intelligent behaviour. 'Works of art'+ are part of a wider class of enterprizes commonly supposed to need intelligence. The difficulty of defining what is meant by art, which any art theory encounters at the outset, is simply an aspect of the wider problem of deciding what qualities characterize intelligence.

Art and intelligence

0.0.1 The traditional distinction between art and the capabilites needed for it, and intelligence and its characterizing capacities, has been facilitated by the view that both sets of faculties are characteristic but distinct aspects of the all-embracing faculty of mind, taken to be the defining principle of homo sapiens. But mind itself has shown peculiar resistance to attempts at capturing its quality by succinct definition, principally because such efforts have usually set out with tacit or even

Mind

+ Inverted commas are placed about this phrase to indicate that its meaning has not yet been properly defined. Hereafter it will be denoted simply by the letters WOA (pl. WOAs) as a reminder of this. As in the case of the elements of the 'productive procedure' introduced later (1.1.0), the main argument attaches no meanings or connotations to the symbols used, beyond those of their definitions. The object is to allow the various elements to develop and become familiar in their own right, like characters in a novel, free from pre-conceptions carried over from earlier acquaintance. Interpretations of the symbolic elements are provided in the notes.

explicit assumptions of its incorporeality. Such assumptions are not disinterested. They underpin the claim to a quality that sets men apart from the rest of the universe, which results from a conviction that is probably among the most deeply rooted in cultural history.

Art the last stronghold of human uniqueness

0.0.2 To believers in such uniqueness, science appears inimical, compelling men to concede their own supposedly unique capacities to lesser creatures and making them common subjects of the same universal laws, from whose jurisdiction god-like pre-eminence had once unquestioningly been supposed to exempt them. Repeated scientific incursions have forced retreats from one after another seemingly impregnable stronghold of human exclusiveness. Successively science has driven men from the centre of the universe, demanded their acknowledgement of familial bonds with animals, deprived them of vital spirit and free will, and finally obliged them to contend with machines for their very character, intelligence. To protagonists of the belief in men's extra-natural pre-rogatives, art, and its connexions with imagination and creativity, intuition and the emotions, has seemed the last of those activities and accomplishments hitherto arrogated to the human preserve, not yet compelled to surrender its distinctively human character. It has appeared to offer the exclusively human faculties a last defensible position against the offensive of hostile science. 'The weakness of the advocate of inexplicable creativity', Minsky (1958) says, 'lies in the unsupported conviction that after <u>all</u> machines have been examined some items will still remain on the list' of 'creative performances'. (p.6) Art is looked to as one of the last possible sources of this view's needed support.

SECTION 1 SCIENTIFIC AND ARTISTIC REALITY

0.1.0 Science and art were not always so opposed. Earlier in the scientific age, in 1711, Pope could speak of them almost as if they had been synonymous:

Science and art

> 'One science only will one genius fit:
> So vast is art, so narrow human wit.'
> (An Essay on Criticism, lns. 60 & 61)

Similar attitudes characterized the Enlightenment. '... the genius of the arts can have no guide other than the torch of reason' Jacques Louis David wrote in 1793 (Honour, 1968). But by 1959 the two enterprizes had become The Two Cultures: 'Literary intellectuals at one pole - at the other scientists ... Between the two a gulf of mutual incomprehension - sometimes ... hostility and dislike, but most of all lack of understanding' (Snow, 1971, p.15).

0.1.1 The advent of 'analytical machines' quickly gave shape to two sets of views. One was proposed by Lady Lovelace:

Analysis v. originating

> 'The Analytical Engine has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform. It can follow analysis; but it has no power of anticipating any analytical relations or truths.' (Bowden, 1957, p.398)

These often-quoted words might still stand as an acceptable basis for an ennunciation of the idea of an 'effective procedure', provided the word 'originate' were properly understood. But Lady Lovelace's aim is to stress the limits of machines' capacities, not the extent of their powers. Her evaluation depends on the distinction it draws between analysis on the one hand and the ability to originate on the other. This distinction corresponds to the difference, already mentioned, that is commonly made between intelligence, and imagination and the emotions. In this form it is a view that provides a nucleus for a formulation of the evident opposition between science and art, and between their respective proponents, one that often shows, especially among scientists, considerable ambivalence.

0.1.2 Medawar (1972) distinguishes a kind of internal consistency as a minimum requirement for any 'structure of imaginative thought' (p.32), whether scientific or literary, which only the scientific extends to the more rigorous constraint that calls for a 'correspondence with reality' (p.29). 'All scientific theories ... are expected to conform to reality, to be empirically true.' But in imaginative thought there is 'a relaxation of this condition, or a failure to enforce it' (p.33). A moralizing tone, revealing a vein of puritanism pervades Medawar's strictures on literature, irritation at its waywardness and attitudinizing. At times his concern is literature itself, at others the 'literary syndrome in science', whose 'distinguishing marks' are:

Science's moral attitudes and ambivalence to art

> '<u>First</u> ... an open or implied claim to a higher insight than can be achieved by laboratory scientists or historians or philologists, or by philosophers of the traditional English kind, an insight which soars beyond the busy little world of test tubes and graphs and measuring instruments, or indeed of facts. <u>Second</u>, there is a combination of high imaginativeness with a relaxation of or a failure to enforce the critical process, so that the critical and inventive faculties no longer work together synergistically, but tend if anything to compete; and with this goes a whispering compaign about the importance to be attached to validation or justification and even, in extreme cases, now beyond remedy, against rational thought. <u>Third</u> (and this is what gives the syndrome a literary rather than a metaphysical character) is the <u>style</u> in which the high truths of the imagination are made known, a style which (among many other disfigurements) deliberately exploits the voluptuary and rhetorical uses of obscurity, a style which at first intrigues and dazzles, but in the end bewilders and disgusts.' (p.37)

Medawar may reserve his disgust for the abuses of literature, but it is difficult for him to avoid his steam scalding the wider literary world as well. The object of quoting so lengthily is to capture the tone of the invective with which science meets its detractors. Yet overall

Medawar's attitude is ambivalent. It is, he admits, 'mainly a love of science that prompted me to speak as I have done', though 'it could equally well have been a love of literature'. (p.38)

0.1.3.0 Other upholders of science are less discriminating than Medawar in their disparagements, with their views frequently acquiring the character of a fully-fledged moral crusade. Such a spirit animates Plato for example to banish art from his Republic, on the ground that, as mere imitation (cf. Lady Lovelace's complaint about machines), it offers 'no knowledge of value' (Lindsay, 1954, p.305) on the subjects it imitates, unlike philosophy, whose aim is to reveal truth by a process just the reverse of imitation, that of stripping away what lies between sensation and reality. Not that this is to be taken as detracting from art. On the contrary, art is to be prohibited precisely because of its 'magic charm', which gives it such a profound capacity to mislead and is the reason why, like a wicked seducer, 'poetry is not to be taken seriously'. (p.312)

Some attitudes to art and science: Plato

0.1.3.1 Distance from Plato both in time and interest does not reflect in the views about art and science that Freud held, which in important respects, are strikingly like Plato's. Freud acknowledges a persistent and continuing indebtedness to literature, resorting to it often before clinical evidence. Psychoanalysts are only repeating scientifically what writers have always said. Yet, paradoxically, such praise of writers' and artists' insights turns to dismissiveness when serious questions come up. 'Art does not seek to be anything else but an illusion ... it never dares to make an attack on the realm of reality' (Freud, 1933, p.205). Enjoyment of works of art is simply 'satisfaction through fantasy' (Freud, 1966, p.80); literature merely an intermediate stage of the mind on the

Freud

route to rationality. (Freud, 1966a)

0.1.3.2 Even allowing art - once again, especially literature - a degree of answerability to reality, poetic truth remains no more than a second best for science's defenders. To Von Mises (1951) for example, 'What the poet reports ... are experiences about vital interrelations between observable phenomena' (p.294). Nevertheless poetic truth is only to be turned to where science proves the laggard.

Positivism

> 'In areas of life that are not sufficiently explored by science, poetry expresses, by means of linguistic forms that have been created for that special purpose, experiences that are present in the consciousness of the poet, in the form of moods, feelings or inspirations, with the aim of communicating these states of consciousness to the reader or listener.' (p.300)

0.1.4 The other side of the argument, that of art's advocates, often goes a long way towards vindicating attacks on it. Jefferson (cited by Bowden, 1957) for example takes up the argument against machines where Lady Lovelace has left it:

Machines and men: a view from the arts

> 'Not until a machine can write a sonnet or compose a concerto because of thoughts or emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain - that is, not only write it but know that it had written it. No mechanism could feel (and not merely artifically signal, an easy contrivance) pleasure at its success, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants.' (Bowden, 1957, p.320)

Jefferson's tone is dogmatic, his concern ostentatiously with <u>human</u> prerogatives, as if this alone endowed him with a moral superiority sufficient in itself to obviate any need for rational argument. Commenting on the passage, Bowden recalls Turing's (1950) question as to how one would know what the machine felt, other than from the evidence of what it produced. But from Jefferson's lofty viewpoint, arguing at all concedes too much to his opponents' barrenness. He prefers deliberately to leave his assertions unsupported, ignoring such trivia as vulnerability to mere scientific attack.

0.1.5 Yet implied by Jefferson's words is the same claim for art that is made for science, that it is in some quite distinctive way concerned with reality. Vehemence and indignation is common to both sides, the indignation of each at the presumption of the other in claiming to deal with reality at all, accompanied by the vehement assertion of the unique entitlement to do so itself. Scientific reality is based on measurement and abstraction; its aims are discovering regularities, generalization and formalization. Artistic reality is concerned with human experience, with what Beckett (1970) likens to Suffering, which he contrasts with Habit.

Artistic reality

> 'The fundamental duty of Habit ... consists in a perpetual adjustment and re-adjustment of our organic sensibility to the conditions of its worlds. Suffering represents the omission of that duty, whether through negligence or inefficiency, and boredom its adequate performance. The pendulum oscillates between these two terms: Suffering - that opens a window on the real and is the main condition of the artistic experience, and Boredom - with its host of top-hatted and hygienic ministers, Boredom that must be considered the most tolerable because the most durable of human evils.' (p.28)

Beckett's definition of Habit agrees nicely with descriptions of the regulating function of the brain such as those advanced by Sommerhof (1950) or Ashby (1965, 1967). Little of what Beckett says about the omission of Habit's duty opening 'a window on the world' should disturb anyone who defines regulation as a means of absorbing variety. It is in this enterprize in the physical world, and somewhat less in the mental that science has shown itself so successfully as technology's hand-maiden, errecting barriers of growing impenetrability between human sensibility and the 'condition of its worlds'. If, as Beckett suggests art's purpose is to bring about the destruction of this regulating shock-absorber and its conniving science to reveal the 'reality' of undiminished variety and undamped disturbance, the hostility between science and art is more readily understandable. It is due to a conflict of aims. Habit, unselfconsciously forging heuristics for finding paths to stability and equilibrium, to resting states whose attainment signals an end to further search, easily comes to be associated with science, which has so much abetted Habit's imitator, technology. And it is easy to see how one attracted to the negligence or inefficiency that leads to the omission of the duty of Habit, might come to find science not merely an inadequate way to represent reality but, like Habit itself, stupefying. Like the advocates of science, those of art argue from their own premises, almost in seeming-invitation of their adversaries' insults and assaults. But in time the 'old terms move from within technical discussions to the outside, taking successively the forms of chapter headings, book titles, and names of professions' (Minsky, 1972, p.4). It becomes harder to maintain arguments like Jefferson's, as technical advances steadily undermine their substance, though Beckett's words are less vulnerable. For Beckett expresses as much a preference of taste as an argument of advocacy, and probably comes closer to saying what determines a choice of sides between art and science than tendentious disagreements about whose wares are best.

SECTION 2 A TASTE FOR ART OR SCIENCE

0.2.0 A preference for artistic reality does not of itself necessitate an insistence on special human prerogatives or the incorporeality of mind, though such beliefs are often its associates. To one seeking the kind or reality that art offers, the efforts of science may easily appear in-

Artistic reality not dependent on human uniqueness

relevant or worse. The relatively restricted pre-occupations of traditional science; scientific aims such as measurement, abstraction, formalization, which have nourished the prejudice that makes 'hard' science seem inherently more scientific that 'soft', and both more than 'social', contribute to this sense among the artistic of science's emptiness. And by contrast fills out the arguments in favour of human uniqueness with an illusory substance that inclines artistic judgement towards them, not because it finds them necessary, but primarily because they satisfy artistic prejudice through their claims to a concern with human qualities. (N.0.4)

0.2.1 Scientific reality on the other hand allows no leanings one way or the other. It commits its adherents against dualistic principles. Cybernetics, particularly the study of artificial intelligence, has furnished specific arguments to show the sufficiency of the physical world alone to account for the existence of mind. Craik (1943) provides a detailed justification for 'physical explanations' as being those that cover 'the most facts by the fewest postulates and leaves the fewest anomalies outstanding' (p.49). McCulloch (1965) made it his self-appointed life's work to delineate the 'Embodiments of Mind' by showing at least the theoretical possibility of making physical components that reproduce its attributes. The classical expositions both of Turing (1950) and Ashby (1965) have demonstrated the theoretical adequacy of physical machinery for capturing mind-like qualities. And pattern-recognition, problem-solving, inductive and heuristic programming, together with all the other enterprizes of artificial intelligence have been at least primitive practical demonstrations of the theory.

Science committed against dualism

0.2.2.0 But the question why the notion of dualism should occur at all, or where

Men's models of the world and of themselves

the sense of mind-body separation springs from, which is intuitively so strongly felt and richly represented in thought and language, has attracted less attention than its refutation, even among psychologists. What basis is there for the claim for the existence of uniquely human qualities? What leads to the belief in any quality at all of this mystic order, able to evade natural laws? Minsky (1965) offers an hypothesis to account for this division in terms of men's self-models. He uses the term 'model' in the sense of a ternary relation: 'To an observer B an object A* is a model of an object A to the extent that B can use A* to answer questions that interest him about A' (p.45). We precis Minsky's argument as follows: taking ourselves the role of B, we attribute the ability of a man M to answer questions about the world W to some 'internal mechanism W* inside of M'. Questions about 'things in the world' may be answered by reference to W*, in ways such as those suggested by Craik, consisting of machinery doing symbolic calculation with analogue characteristics. But questions about the 'nature of the world' are really questions, not about W, but about W*. If W* includes a model M* of M, it follows that M* can contain a model W** of W*; and, at the next step, W** may contain a model M** of M*. On the definition Minsky proposes of a model, this must be the case if M is to answer 'general' questions about himself. M* answers questions about himself like how tall he is, but very broad questions about his nature, 'what kind of a thing he is, etc,' require descriptive statements by M** about M*, when answers are possible at all. Minsky notes the inadequacy of his intuitive description of 'model'. Difficulties arise for instance immediately one questions the internal relationships of the substructures it postulates. For example there is a sense in which W** 'contains' W*, since W** must include an interprative mechanism that refers to it. In another more usual sense W* contains W**. 'An adequate analysis will need much more advanced ideas about symbol representations of information-processing structures'. 1C

0.2.2.1 But Minsky's concern is to indicate how the dimorphism of our world models results in notions of dualism. This, he suggests, is due to the 'distinctly bipartite structure' of these models

Dimorphism of world models

> 'One part is concerned with matters of mechanical, geometrical, physical character. The other part is associated with things like goals, meanings, social matters and the like. This division of W* carries through the representations of many things in W*, especially to M itself. Hence, a man's model of himself is bipartite, one part concerning his body as a physical object and the other accounting for his social and psychological experience.' (p.46)

The man's belief that he has a mind as well as a body is the conventional way he has of expressing this to himself. The reason for the structural divisions in our models are probably connected, according to Minsky, with the heuristic value they have for us, a value which, he indicates, should be thought of in terms of the stratification of computer programs into different levels, in which there may be executive programs, sub-routines and the like.

> 'When intelligent machines are constructed, we should not be surprised to find them as confused and stubborn as men on their convictions about mind-matter, consciousness, free will and the like. For all such questions are pointed at explaining the complicated interactions between parts of the self-model. A man or a machine's strength of conviction about such things tells us nothing about the world, or about, the man, except what it tells us about his model of himself.' (p.48)

0.2.3 The particular value of this hypothesis of Minsky's lies in the help it gives towards drawing together into a single thread the several loose ends of conflicting claims and conjectures advanced in the futherance of the different interests we have mentioned so far. These conjectures have taken the form of a number of antinomies, between reason and imagination, body and mind, scientific and artistic reality, all running to some extent parallel with one another. The hypothesis suggests that the antinomies reflect a valuable world modelling heuristic, which calls for a division of models into mutually referring parts, corresponding roughly

The value of Minsky's hypothesis

with distinguishable aspects of the world, broadly the material and the organizational. If the hypothesis were correct and the use of some kind of stratifying heuristic were the case, it would be as much part of reality as anything else, and its existence would therefore go some way toward resolving the conflicting claims of science and art, each to being reality's sole representative. On the one hand, according to the hypothesis, an insistence on the oneness of the world would result in science and the scientific reality that puts this oneness among its chief premises; on the other, a concern with what is distinctively human and to do with human experience would be synonymous to a concern with the human self-model, and so easily lead to art and artistic reality, which, because it centres on what M** models, emphasizies M**'s qualities, namely the organizational rather than the material aspects of the world.

Cybernetics enables science and art to share 'raw material'

0.2.4 Only relatively recently have things like goals, meanings and social matters become the concern of science at all. Psychology in its psychoanalytic avatar has, it is true, retained its early interest in literature and has also aroused considerable literary interest in itself. But 'hard' science is disinclined to take at all seriously the claims of psychoanalysis to be a science at all. Behaviourism, on the other hand, and those branches of psychology that have made special efforts to imitate the 'hard' sciences, largely by deliberately restricting their scope to domains accessible to science's traditional tools, are barren places to search for representations of the richness of human experience. And art, aesthetic pleasure, 'resensualization' as Moles (1966) calls it, 'is proper to a psychology other than behaviourism' (p.167). Cybernetics though is uniquely equipped to deal with richness, while easily satisfying demands for scientific hardness. Its growth has been largly synonymous with the advance of concern with machine intelligence and its ramifications. Thus for the first time only recently has it become possible for science and art to share a common 'raw material' and so open a new field for investigation, in which the 'two cultures' may hopefully prove mutually illuminating.

SECTION 3 PROBLEMS FOR AN ART THEORY

0.3.0 A satisfactory art theory should aim at resolving conflicts that result in artistic claims on truth and reality contending for precedence against the claims of other modes of expression, without destroying by reconciliation the distinctiveness of the different modes or their centres of interest. It would almost certainly be impossible to distinguish art from science by assessing, if it were possible, the different amounts of intellectual and imaginative energy that needed to be expended in the pursuit of each, even supposing these faculties were separable. It is therefore correspondingly futile to look for characterizing features in art that rely on such distinctions. Ideally one would hope to find a common set of criteria for judging both scientific and artistic work, as well as other kinds of human expression, at least for 'deep level' evaluations, such as those that might begin from the common starting point of the raw material say. More superficially one would like to suggest answers to some of the more traditional questions of art theory.

Aims of an art theory

0.3.1.0 Osborne (1972) gives a brief outline of art theory's recent history. Currently, theories of art, in the sense of complete, self-contained systems, are out of favour.

Recent art theories

0.3.1.1 Osborne (1952) himself, whose concern is principally within the narrower framework of aesthetics, following earlier writers such as Collingwood (1963), suggested the need for a set of necessary and sufficient conditions for deciding whether or not a thing was a work of art. His contention was repudiated on the grounds that variety among possible candidates made agreement of conditions too unlikely to warrant the effort of the search.

Necessary and sufficient conditions

0.3.1.2 An approach to the problem of defining a 'work of art' offered as an alternative to Osborne's, appealed to the idea of 'family resemblance groups'. This concept due to Wittgenstein (1966), is of an assemblage in which each item has some resemblance to some other item, but in which no property is common to all members of the group. Such a group lacks natural boundaries other than those implied by the word 'family', so that 'work of art' becomes an 'open' concept. On this basis however it is possible to conceive of almost any assemblage being linked, by a chain of suitably chosen resemblances, to any other, leaving the problem of finding defining criteria unresolved.

'Family resemblance groups'

0.3.1.3 Functional definitions have been proposed, which distinguish WOAs by what they do: express, inform, evoke, amuse, puzzle, advertize, exhort, etc. The objection to this approach is that it prescribes in advance of a definition what does what. Defending such 'definitions' against this objection, Osborne points out that they are the more acceptable because they may be taken as mnemonic devices for 'shorthand summaries of conclusions reached after protracted enquiry' (p.20).

Functional definitions

0.3.1.4 A definition of a kind related to the 'functional' one makes a WOA simply an artefact, though one primarily intended for 'aesthetic consideration'. Osborne calls this a 'formula definition'. Little results from the change

'Formula definition' in approach that it entails beyond a shift of emphasis from the WOA to the question as to what constitutes aesthetic consideration.

0.3.1.5 In answer to this question it is suggested that the material for such consideration is 'aesthetic properties', as distinct from 'non-aesthetic'. Aesthetic properties are known by acquaintance and are not to be inferred from the presence of non-aesthetic properties. They are analagous to Moles's concept of 'aesthetic information', as distinct from semantic and other categories of information. Osborne lists 'descriptive "tertiary" or Gestalt properties ("dainty", "dumpy", "gawky"); "emotional" qualities (the serenity of the landscape, the gaiety of the music, the sombre colours of a picture); emotive or evocative qualities ("moving", "stimulating", "depressing"); structural properties ("well or ill balanced", "formless"); a class of properties which are not, but which are dependent upon, structural properties (the difference between "pretentious" and "unassuming", "eloquent" and "bombastic", "grandeur" and "grandiosity", etc.)' (p.9). The difficulty with this definition is that most of the qualities suggested are ubiquitous and the presence of such 'tertiary' properties is therefore insufficient for defining a WOA.

'Aesthetic properties'

0.3.1.6 The notion of aesthetic consideration raises questions, on the one hand as to what are the objects of aesthetic attention, the grounds of aesthetic judgement, and the differences between aesthetic and other sorts of appraisal. Or, more broadly, what is aesthetic experience, or experience generally for that matter, and what distinguishes aesthetic experience from other sorts. On the other hand, questions arise concerning aesthetic consideration and WOAs: Is art revelatory and what is the connexion between revelatory and aesthetic qualities? Is there such a thing as an 'imaginary reality', obeying only its own rules of internal consistency? (cf. Medawar,

Further questions for art theory

0.1.2) What is the relationship between revelatory reality, imaginary reality and 'creativity'?

0.3.2 Implicit in these questions is the notion of a WOA in relation to an audience. Indeed any attempt to understand the concept WOA in isolation is almost certain to prove vacuous, (although theoretically the audience need number no-one besides the artist himself). Explicitly acknowledging an audience raises questions relating to the interpretation and 'meaning' of WOAs, the relationship of art to truth or reality (0.2); and this leads to still further questions, especially those regarding the artist's intention, what it is, how it is manifested, and what its relevance is in aesthetic and other judgements to the WOA.

WOAs

0.3.3 The word 'art' has itself been used differently at different periods, and a number of varying meanings have persisted side by side to the present. Thus for example the word may denote a set of skills, as in 'the art of government' or 'the art of war', or intuitiveness, as in 'the art of politics', inviting contrast with military or political science. (Fowler (1926) asks: 'the art of self-defence, and the boxer's s[cience] - are they the same or different?') Or the word may appear on its own, often indicating some kind of general concern with the attainment of the beautiful. Even within this relatively restricted meaning it occurs in several different senses, often without explicit distinctions being made between them. According to the Oxford English Dictionary, 'The most usual modern sense of art, when it is used without any qualification' is 'The application of skill to the arts of imitation and design, Painting, Engraving, Sculpture, Architecture; the cultivation of those in its princ-

Meaning of the word 'art'

iples, practice, and results; the skilful production of the beautiful in visible forms.' Commonly the subject matter that the word denotes within this usage is often even wider, comprehending besides visible forms aesthetics (0.3.1) and the artistic or 'creative' process itself, and so, by the extension this implies, literature, music, and the performing arts, as well as painting, sculpture, and the like. Unless otherwise indicated, in what follows the word will be understood to have this broad sense. Notice that, within this meaning, reference may be to (1) the artist's output; (2) WOAs, a subclass of this, selected according to certain more or less explicit criteria; (3) the topic whose subject matter includes artists, WOAs, and the criteria of artistic judgement; (4) the abstract quality of WOAs that distinguishes them from other objects.

SECTION 4 EFFECTIVE PROCEDURES FOR AN ART THEORY

0.4.0 Clearly a firm prescriptive definition of art is lacking. Moreover it has eluded protracted search. Often when problems of definition prove intractable it is because the questions they raise are meaningless, unanswerable or at least improperly formulated, or because an adequate analytical apparatus is lacking. In the case of art, to these difficulties, we might add, as we have suggested (0.0.1), an actual interest in vagueness, or at any rate a resistance to approaches that might lead to precision. Intimations so far suggest that, uniquely in science, cybernetics may be sufficiently rich in helpful formulations to support hypotheses for dealing with art's and its associated problems. Particularly in the use it makes of the concept of 'effective procedures', cybernetics appears also to offer an appropriate methodology.

Cybernetics provides methodolody for study of art

0.4.1 An example, in a realm associated with art, of the use of such a method is the approach of Newell, Shaw & Simon (1958) to the problem of 'creative thinking'. Following reasoning like Turing's (1950), that a machine which could counterfeit human intelligence could claim to share it, they set out to make a computer imitate human creativity.

Effective procedures for the study of 'creativity'

> ' ... we would have a satisfactory theory of creative thought if we could design and build some mechanisms that could think creatively (exhibit behaviour just like that of a human carrying on creative activity), and if we could state the general principles on which the mechanisms were built and operated.' (p.2)

They offer a description of 'creative thinking' depending partly on behaviour and partly on kinds of problems. Thinking is to be classified as 'creative' according as these take the forms they do when associated with creative thinking in humans. Such a restrictive approach precludes any chance of finding a more general, non-human-based way of defining creative thinking - analagous say to a definition of intelligence not linked to purely human capacities. A relaxation of the imitative restrictiveness of the approach of Newell & others leads to a wider approach, such as that implied by a remark of Minsky & Papert (1972): ' ... if a theory of Vision is to be taken seriously, one should be able to use it to make a Seeing Machine!' (paragraph 1.0). For our purposes, reversing this argument, we should say that the ability to make and described an art machine would enable the development of an art theory powerful enough to remove many of the difficulties experienced hitherto.

0.4.2 But such a strategy presents a seemingly inescapable difficulty. For the chief problem one hopes to overcome by building an art machine, that of defining art, immediately presents a barrier to ones building it. Lacking any agreed definition or prescriptive formula for what art is, it is not possible to say in general what an art machine ought to do; how would

Problem of defining the task

one know for example whether or not such a machine were working properly? Nor would it help to follow the Turing and Newell & others restrictive approach, by calling for the machine to imitate a human artist, since conceivably almost any output might be judged by some observer (participant in an 'interrogation game', for instance) as a work of art emanating from a human artist; though the converse, an output not taken to be a work of art, would obviously not, on that account, necessarily be judged to be a machine's rather than a human's. In other words, artists are not bound to produce only works of art, and judgements as to whether or not some object is a work of art, irrespective of how it was produced, depend finally upon the judge's decision alone. No object can enjoy more than the likelihodof a given judge classifying it one way or another. At least theoretically and possibly also in practice, no work can rely on the favour of every judge, while on the other hand practically any object might be classified as a work of art by someone.+

SECTION 5 AN ART MACHINE

0.5.0 As opinions about whether different object are works of art will differ from observer to observer, we should at least require that an art machine itself judged that its own products, more or less at any rate, met art's demands. This would necessitate its having some 'idea' of what it 'wanted' to achieve, which is to say, an intention of producing a work of art. In other words, although one might accept an art machine that did not always - or possibly even ever - produce what qualified in one's own judgement as a WOA, one would insist as a minimum demand, that the machine produced its object 'on purpose'.

An art machine must make products intentionally

+ Among possible illustrations of this are the celebrated 'ready-mades' of Marcel Duchamp, everyday, usually mass-produced, objects, arbitrarily displayed as aesthetically significant structures.

0.5.1 To do this the machine would itself need the capability of playing the part of observer, in order to be able to assess what it produced. The criteria it applied in reaching its own evaluations would determine how far they agreed with judgements of other observers, but would not themselves be of vital importance in the construction of the machine. It follows that the machine would require, in addition to the facility that set its output, a facility for observing judging it - feedback, enabling a control to fit output to intention. The machine sketched would therefore need at least the ability to entertain an intention, set an output, and then evaluate output against intention. One way of representing this process might be as a set of ordered triplets giving intention, output and evaluation. For, in general, one might expect modifications to occur to the intention too, as the product took shape. A convergent series of 'differences' between intention and product would ensure that the machine stopped, having produced, according to its own judgement at any rate, a WOA. Divergence or oscillation of differences would result in failures.

Machine needs self-observing facility

SECTION 6 DEFINITION OF ART

0.6.0 Such an outline of minimum requirements for an art machine still says nothing about the problem of defining art. Answers to questions about what art is such as an art machine might be able to provide would have to hinge on what it produced. Discussion of this will form the subject of Part I of what follows, while Part II will seek criteria for narrowing the class of its products generally, to those classifiable as art. Restrictions of generality will be avoided as far as possible. In

Human connotations for art avoided

particular, the meaning of 'art' will not be confined to denoting only what is accessible to humans. There may be systems, humans among them, for which a machine's output is not interpretable or even observable, as for example in the case of direct communication between two machines. Even if such communications were made observable by third systems, by provision of appropriate facilities, they would still remain uninterpretable, at least in the form in which they were intended, unless the third system's coding procedures were matched to the other two. In view of this and the common practical inconsistencies of evaluation of WOAs, it will be necessary at the outset to look for criteria such as will permit more stable results in WOAs' classification.

0.6.1 Making such wide generality an intial aim might seem greatly to diminish any hope of success. But there is still some reason for optimism. For it is widely accepted that such things as WOAs exist, and there is even a measure of agreement about some examples of them. So, proceeding by degrees, it may be possible to arrive eventually at a design for a machine with the desired properties. We shall begin by considering a general process, for which as far as possible we shall derive the properties considered essential or desirable for an art machine. At the outset, the machine is left undefined. We regard it simply as an entity with as yet no known attributes, save that it should be an 'author'.

Art machine to be an author

PART I

AUTHORSHIP

The argument that follows depends neither on the introduction that precedes it nor on the notes that follow. It divides into two parts. The first deals with authorship and the second with its products. Both parts propose their subjects in an abstract, general way, which restrictions gradually narrow and make more concrete. For the most part these restrictions are shown to follow as consequences of the characteristics inhering in the kinds of systems being considered, though there is no hesitation about introducing arbitrary limits where they are needed to achieve empirical plausibility. Such limits may be viewed as temporary supports to shore up theoretical weaknesses that the test of empirical reality has revealed. Applications of conclusions in the less abstract, more familiar world of art, together with certain other empirical considerations are held out of this, the main body of argument, and developed separately in the notes.

Part I begins by defining a rudimentary, abstract, productive procedure and goes on to determine some of the principal characteristics it will require if it is to furnish a basis upon which to build an art machine. The argument begins with a definition of authorship and proceeds by uncovering its qualities in some detail, to reveal the minimal requirements that they will impose upon any system wishing to incorporate them. Broadly the course of argument moves towards discovering a purposeful process that depends on an ability to extract order from the surroundings in which it occurs, and to transmit this order to its products.

CHAPTER 1

PRODUCING AND CLASSIFYING

SECTION 0 THE RELATIVITY OF CRITERIA

1.0.0 No pre-conceptions at the outset restrict an art machine to resemble any familiar object, and its products may naturally share its exoticism. It would beg the question to insist that, as its defining characteristic, it produce WOAs, since one of the objects of making the machine is to define what a WOA is. Expecting WOAs from the machine would imply a belief in what is often called the universality of art, the notion that WOAs, however unusual, share some recognizable defining attribute. On the face of it no theoretical ground justifies any such idea, so the initial problem is to find criteria for deciding what if anything that an art machine might produce could be classified as a WOA - at least as far as the machine is concerned. In the context this is the weakest demand it is possible to make. But abandoning notions of art's universality, while it has the virtue of directing enquiry away from areas that have proved fairly barren in the past, raises new difficulties by introducing problems about the relation between WOAs and their audiences, the relativity of art's criteria.

No universal criteria

1.0.1 The chapter begins by outlining a paradigm for a productive process AW with a controller A and product W, and then goes on to consider the further requirements for distinguishing those Ws that are WOAs from the rest. As in general AW is taken to be non-human, the usual criteria for classifying WOAs do not apply, since W may not be interpretable by humans. Never-

WOA and audience

theless, as we wish to know whether the process is producing WOAs, according at least to its own interpretation, we must seek other criteria. This turns out to be possible if we assume the role of meta-observers, observing, not W, but the relation between W and its observers. If this relation has characteristics in common with relations known to be between observers and WOAs, then we classify W as a WOA; if not, not. It is shown that every day usage easily blurs the distinction between observers and meta-observers and that this leads to ambiguities, which sharpening the distinction helps resolve .

SECTION 1 THE PRODUCTIVE PROCESS

1.1.0 Dictionary definition of 'author'

The Oxford English Dictionary defines an author as an 'originator (of a condition of things, event, etc.)'. What kind of system will have the properties necessary to satisfy this definition?

1.1.1 Paradigm for a productive process

Consider an abstract process AW, consisting of a set of procedures A say, executed in a processor, and coupled to a transducer M, with outputs that act on an environment E, in such a way as to cause some alteration, call it W, in E. The process and its environment together make up a universe U. (Figure 1.1.1)

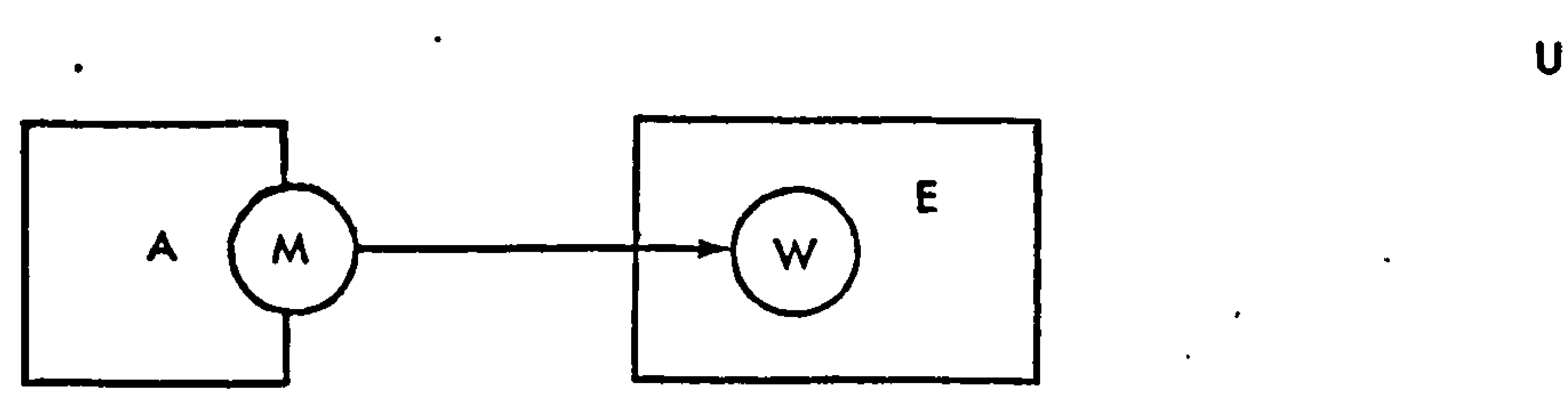


Figure 1.1.1

Let AW serve as a naive paradigm for a controlled, productive activity in which A is the controller that produces W. If A were not the controller we should not refer to W as its product; and if W were not A's product we should not call A controller of AW. By definition then, A is a controller if and only if W is its product. Conversely, W is A's product if and only if A controls AW. What this definition entails for A we shall see in due course (2.1).

1.1.2 If A is an autonomous, - lacking inputs - dynamic entity, W will reflect A's output states.+ In particular, suppose A and E are determinate systems and, following the form of Ashby's (1962) formulation, suppose A has the set of internal states Z_A say, and E has the set of internal states Z_E, then A is defined as the mapping of Z_A into Z_A, under some function α say, and E as the mapping of the product set $U_E X Z_E$ into Z_E, under the mapping ϵ say, where U_E is the set of E's input states. In so far as A's outputs V_A determine U_E, the operation of AW will have the effect of mapping A into E, and to that extent W will represent a mapping of A. The advantages of this kind of representation will be obvious. It at once satisfies the requirements that intuition demands as a minimal basis for a machine that is to produce what may ultimately be made art, namely the need for a process with a product, while it also is able to furnish a vehicle for the analytical tools that are available, such as those based on set theory and logic, in the way for instance that Ashby (1967) expounds.

W is a mapping of A

+ At this stage we disregard M, which is equivalent to assuming one-to-one correspondence between A's outputs and M's. It will become evident as we go on that M's role in the artistic process is secondary. But, as was seen in 0.5.1 and will appear in more detail subsequently, the part M plays in practice is in modifying artistic goals. Its imperfections are therefore an intrinsic part of finished WOAs.

1.1.3 In general the relationship between A, M and W in such cases is well understood and is too broad to be of particular interest for the present purpose. What is wanted is to uncover whatever may be of peculiar

Need for an observer

importance to the artistic process, Beyond delineating these minimal structural requirements therefore, we shall seek ways of understanding more about the characteristics of the entities involved. The situation becomes more interesting if we introduce a further element, an observer O, into the paradigm for AW. (Figure 1.1.3)

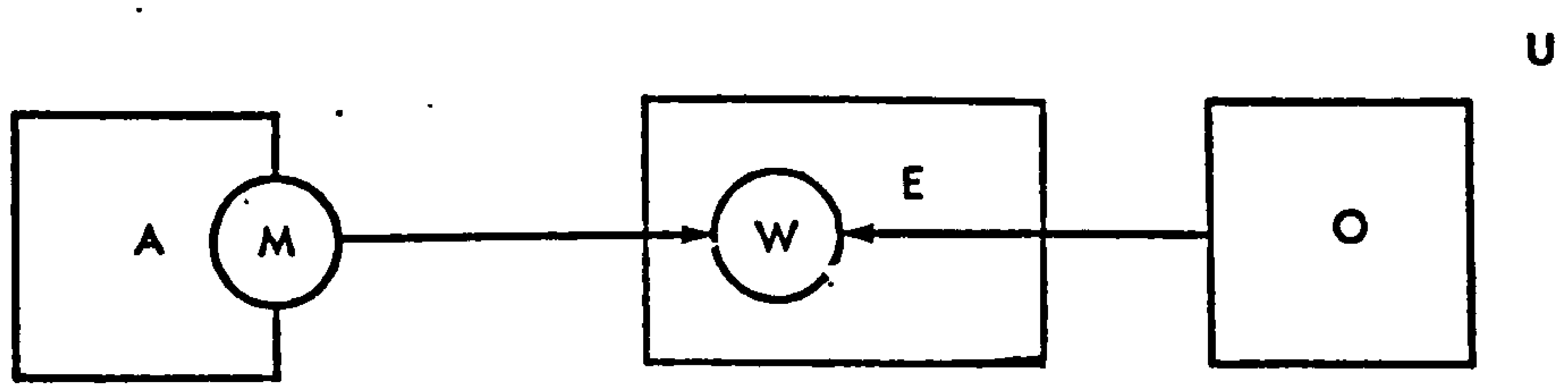


Figure 1.1.3

Notice that now, instead of ourselves having O's role as in the previous formulation, we are one step further removed. We are free to look, not merely at W, AW's product, but also at the relationship between W and O. This, according to Moles (1966), 'is philosophically necessary' if we are to be able to make statements about W's artistic qualities or, in Moles's terminology, about the 'aesthetic', distinguished from the 'semantic' information that W carries. 'In addition to the normal source-receptor channel, the <u>observer</u> who examines the signals received from the source constitues an auxilliary channel. This observer, considering the signals discrete and free from noise, describes them in a universally intelligible meta-language' (p.130). We shall show that such a meta-observer is a necessary part of any satisfactory definition of WOA.

SECTION 2 THE OBSERVER

1.2.0 What are the observer's characteristics, and what new characteristics does its presence induce in AW? We may take O to be an entity of the same kind as A, M and E, namely a determinate, dynamic system with internal states Z_0, receiving inputs from E - but not from A, since initially we suppose that A may be unobservable by O - and in particular from W. Clearly in general not all the Ws AW produces will be WOAs, and it is therefore necessary to find ways of distinguishing those that are from those that art not. Some procedure is needed for judging W and some entity to execute the procedure. Since - at this point at any rate - A is autonomous, - has no input - it cannot assume this role. Neither can W, since it cannot make decisions about itself - in the sense of being truly self-organizing (see 2.4.3). It follows that O must have the judge's role.

Evaluation of W belongs to O

1.2.1 It may seem that making O W's observer should alone have been enough to imply this. But notice that what this argument shows is the sufficiency of W and O for 'producing' a WOA. Anticipating chapter 8 where O is shown to be a necessary feature of an AW that is to produce Ws of any kind, it follows that O and W together furnish the necessary and sufficient elements

O and W are necessary and sufficient elements for WOA for defining a WOA.+

1.2.2 The reason for O's importance is that 'the real world gives the subset of what is; the product space represents the uncertainty of the observer'. (Ashby, 1962, p.262) The product space will therefore depend on the particular observer concerned, and different product spaces due to different observers may record the same actual events of the physical world. What O observes and so how he evaluates W depends upon the constraints that the communication between W and O implies. His observation will therefore have the form of a relation between W and O. Consequently the properties of the constraint and resulting judgement of W by O will depend both on W and O. Because judgement of W cannot be made except in this way, any theory as to what makes W a WOA will have to be concerned with properties that inhere, not simply in W but in the relation between W and O.

Reason for O's importance

1.2.3 This argument's consequences may be more readily grasped when applied to intelligence, where it admits to the possible existence of intelligent systems in which the exercize of this faculty may be quite unlike and unrecognizable to other intelligent systems, ourselves for instance. 'There is no difficulty in principle in developing <u>synthetic organisms as complex and as intelligent</u> as we please' (Ashby, 1962, p.268), as long as we understand that their intelligence will reflect an adaptation to 'keeping their own essential variables within limits' (p.268) in their own environments, without implying a wider applicability of the adaptation to any other environments or variables. Pask (1959) observes that we will recog-

Generalized concept of intelligence

+ Chapter 8 will incorporate O as part of A.

nize such possibly unfamiliar forms of intelligence, if and only if there is a field of activity common to them and ourselves. If for instance I say that 'a chimpanzee has "grasped a concept", it is because I can imagine myself having learned from experience in somewhat the same way' (p.881). But, in the absence of this common 'region of knowledge or ... region of connected and tentatively confirmed hypotheses' (p.884), we shall have to rely on observations of the relation between the organism and its environment. Specifically for example, suppose, as Sommerhof (1969) suggests, we set about defining intelligence by means of analyzing the fundamental functions of the brain as they occur throughout the animal kingdom, those of adapting, regulating, co-ordinating and integrating, in terms of 'directive correlations and hierarchies thereof' (p.200), with the object of using the equations that these mathematical definitions yield, as criteria for discovering brain-like mechanisms, then the definition of intelligence we arrive at will have assumed the existence of a meta-observer to derive and use the criteria. Broadly, the meta-observer must infer that the organism's behaviour reflects intelligence, from the similiarity he recognizes between the relation of this organism's behaviour to its environment, and the relation previously classified as intelligent by him, between other organisms and their environments.

1.2.4 Our present object is to find ways of classifying Ws and WOAs that do not exclude a priori any possible W or O. It follows from the two preceding paragraphs that the criteria for achieving this will be concerned with the relation between W and O and not simply with W itself. Such criteria will have to be employed by a meta-observer, call it $\bar{O}$. $\bar{O}$ will then be able to come to the decision, 'O judges that W is a WOA', from an observation of the relationship between W and O alone, without needing

Classifying WOAs

to apply or even to know the procedures that have led O to a decision about W.

SECTION 3 THE META-OBSERVER

1.3.0 The distinction between statements by O and those by $\bar{O}$ is not always clear however. The reason for confusion is connected with the correspondence that exists between these two categories of statements and the distinction that philosophy makes respectively between knowledge by acquaintance and knowledge by description; the former being knowledge that exists by direct awareness, without the need for the intercession of inference or the knowledge of truths; and the latter knowledge that is not directly apprehended. Russell (1948) relates the distinction to what he calls 'egocentric particulars'. In his formulation of the distinction, these are objects that we know by experience. Suppose one of these is an object a. Other objects we know by their relationships to these experienced objects. Suppose b is (known somehow to be) the only object to which a has a known relation R. b is not an object of our experience, but knowing that it is the object to which a has the relation R, we can give it the name 'b' say.

Knowledge by acquaintance and by description

> 'It then becomes easy to forget that b is unknown to you although you may know multitudes of true sentences about b. But in fact, to speak correctly, you do not know sentences about b; you know sentences in which the name 'b' is replaced by the phrase 'the object to which a has relation R''. (p.103)

What we know about the object to which a has the relation R may be expressed in sentences verbally identical with sentences about the actual object of b. But being able to describe these sentences or even knowing up to a point whether they are true of false, we still 'do not know the sentences themselves' (p.103). 'For in fact ... everything except myself is ... only known to me by description, not by acquaintance. And the description has to be in terms of my own experience' (pp.103.104). O knows W by acquaintance. $\bar{O}$ knows W by description. But $\bar{O}$ readily assumes O's role, and statements made in the two different capacities are confused as Russell's object b is confused with its name.

1.3.1 $\bar{O}$'s ability to make judgements about the relationship between W and O leads naturally to the extension of the concept of a WOA, by disengaging the judge from the role of audience.+ The extension is achieved by means of analogy. 'An analogy relation is a relation between relations described, in extension, as a morphism (for example, a one-to-one correspondence or isomorphism or a homomorphism)' (Pask, 1974, p.260). Minsky & Papert (1972) explain analogies in terms of descriptions. Suppose O_1 and O_2 are observers of W_1 and W_2 respectively; and suppose $\bar{O}$ observes and, from his observations, makes descriptions D_1 and D_2 respecively of the two relationships between O_1 and W_1 and between O_2 and W_2. If $\bar{O}$ knows (no matter how) that W_1 is a WOA in O's interpretation, he will know that D_1 describes the relation between a WOA and its audience. If he then notes a similarity between D_2 and D_1 it may lead him to suppose that W_2 is a WOA in O_2's interpretation.

Definition of WOA by analogy

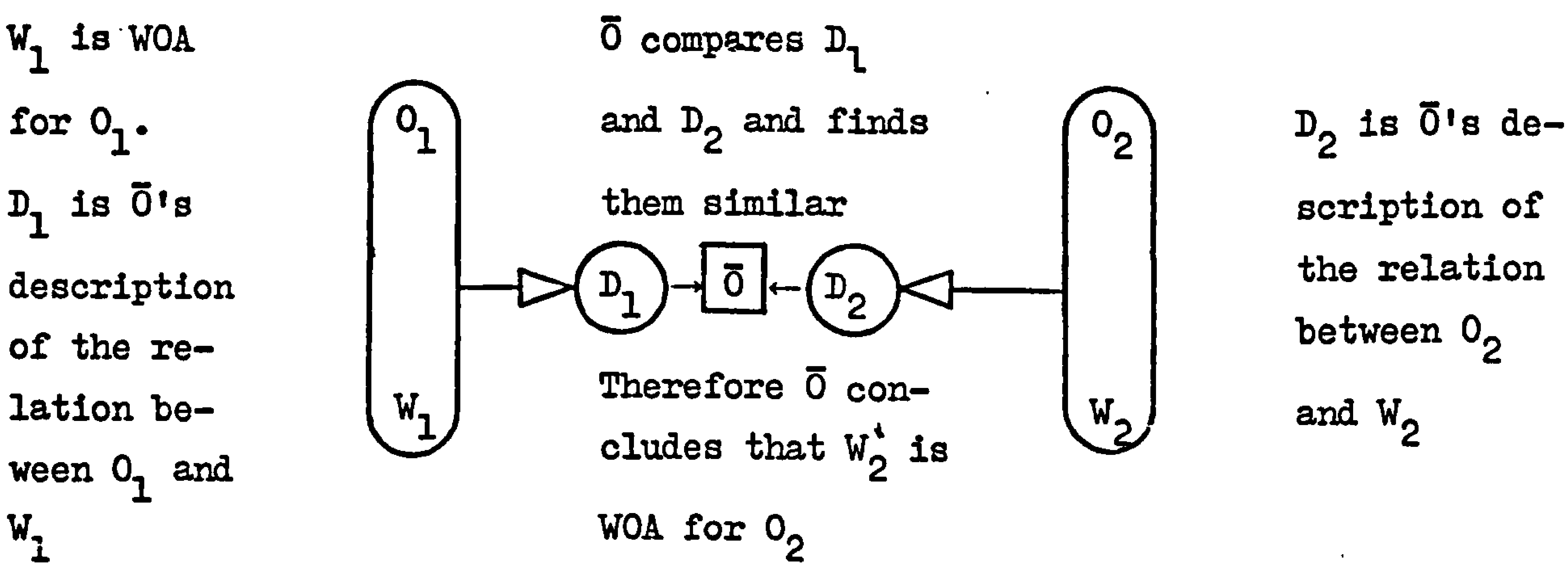


Figure 1.3.1

+ The word 'audience' is used throughout to denote whom- (or what-) ever enjoys or appreciates a WOA, whether strictly in the role of audience, as in a theatre or concert hall for example, or more widely as readers, viewers of paintings and the like.

1.3.2 Thus the same system S, in the role of $\bar{O}$ might classify behaviour as intelligent which in O's role it would not necessarily classify similarly. This might be for instance because the behaviour was quite unfamiliar to O and therefore not open to its direct (by acquaintance) classification. Applied to art, we might imagine two widely differing kinds of entity, O_1 and O_2 observing the same W. O_1 might classify W as a WOA while O_2 did not. Such a situation might come about if for instance O_1 were an intelligent being that had evolved in the ocean deeps and O_2 were a human, or in less bizarre circumstances. Yet by assuming the role of $\bar{O}_2$, O_2 would be able to recognize the relationship between W and O_1 as a relationship between a WOA and its audience, and thereby recognize that O_1 interpreted W as a WOA. (N.1.0)

Conflicting judgements of O and $\bar{O}$

1.3.3 In effect, procedures for O and procedures for $\bar{O}$ may not differ greatly. One might regard them as overlapping subsets of the total set of procedures of the system concerned. Because of this, the distinction between the two different kinds of evaluation of W, those emanating respectively from O and $\bar{O}$, is not always sharply maintained. A single entity alternates easily between the two sets of procedures that the respective judgements entail, assuming O's role at one moment and $\bar{O}$'s the next, without necessarily distinguishing between judgements reached in each of its distinct capacities. The faculty, frequently encountered, for an entity to act as a meta-observer of itself for instance, by commenting on its own evaluations and how or why it has reached them, adds to the confusion of the roles. Presumably a man's M** model of himself (0.2.2.0), which answers questions about what kind of thing he is, must be furnished to some extent by meta-observations. Part II deals with the question of what are $\bar{O}$'s criteria in the first place for labelling relationships between W and O. But notice here that the criteria need not at all depend upon $\bar{O}$'s ability to assume the role of O.

Confusion of the roles of O and $\bar{O}$

All that is required for deciding in individual cases whether a given O interprets a given W as a WOA is the existence of <u>some</u> criterion. The relative generality of what criteria $\bar{O}$ adopts will determine how frequently this will occur. So the criteria for classifying relationships between W and O will be of critical importance.

CHAPTER 2

STRUCTURE, PURPOSE AND EXPLANATION

SECTION 0 STRUCTURE AND PURPOSE

Parallel growth of structure and purpose

2.0.0 The argument of this chapter is complicated and difficult for a reason that is partly the chapter's subject, insufficiently evolved explanatory apparatus. The literature on the various topics that the chapter comprises does not in general deal centrally with the intimate relationship between structure and purpose. A number of authors give secondary consideration to the connection in the context of other questions which are their specific concern. Of those the chapter cites, Lofgren (1972) comes closest to showing direct interest in the relationship in his formulation of explication and self-reproduction. Partly this is because the investigation of other topics has logically commanded priority, partly because of the topic's inherent difficulty. Yet purpose and structure develop side by side. A complicated interdependence brings purpose out of structure and still wider structure out of purpose. As structure becomes more complex, so purpose grows more important. What begins as the more or less 'mechanical' growth of crystals and the stereospecific production of the more complex molecules that they build, has become an active 'teleonomic' - Monod's (1974) word+ - performance in the still more complexly structured bacteria into which the crystals form. The reason for undertaking here what others may have judged as inauspicious or uninteresting is the belief that a fuller appreciation of the qualities commonly associated with the notion of intelligence demands a closer under-

+ ' ... we shall define the essential teleonomic project [a project is a goal, viewed within the context of some wider structure of interdependent goals] as consisting of the transmission from generation to generation of the invariance content characteristic of the species. All structures, performances and activities contributing to the sequence of the essential project will hence be called teleonomic .' (Monod, p.24) standing of the relationship between purpose and structure, or more specifically for the present context, of the structural requirements of a purposeful productive process.

2.0.1 There are three main sections to the argument. The first concerns the properties of a controller: the minimal requirements; the need to be able to draw distinctions and the consequent need for a motive. This leads to the conclusion that a controller must be purposeful. Goals and purposes are defined, without recourse to notions of feedback, and this is shown to be justified. In the second part, the argument sets out to define W. W is part of E and therefore must be distinguished from the rest of E. The problem of achieving this raises questions of order, in the sense of pattern or form. There is a review of various approaches to the definition and measurement of order, with discussion of their similarities and differences. The problem arises of the meaning of randomness, and there is discussion of order in relation to the conservation of information. The final part of the argument deals with the problem of explaining order, once it has been detected. It is shown that explanations of order must be made in terms of other (corresponding) order. Certain views are discussed concerning the nature of explanation. Further discussion raises questions of order's antecedents, and shows that the explanation of order is bound to appeal to notions of purpose, and that indeed the existence of order will be taken to presuppose the existence of purpose. It appears that the order that is discovered will depend on the character of the observer who makes the discovery. As the observer takes the order discovered to be the outcome of purpose, he will postulate an agent to whom the purpose belongs. It is argued that a purposeful entity is a necessary and sufficient condition for the explanation of order.

Explaining order

SECTION 1 GOALS AND PURPOSES

2.1.0 Let A be a controller in so far as it determines W, that is, by 1.1.1, in so far as it produces W. In other words, allow the possibility of degrees of control. To be a controller, A must be able to make imperative statements of the kind, 'Do X', 'Let Y', and so on, or produce outputs equivalent in their effect. So A must include procedures capable of producing changes in E corresponding to such statements. It follows that when A acts as a controller it maps some of its internal states into E, namely those states corresponding to those of its control commands that are satisfactorily obeyed.

A maps into E

2.1.1 Thus A must be able to call names, such as 'X', 'Y' and so on. Adapting Spencer-Brown's (1969) argument, given in his definition of distinction; the name A calls indicates a value; the value is of a content; the content is indicated by drawing a distinction; there can be no distinction without a motive; thus, to call a name A needs a motive. But to be a controller A needs to be able to call a name. It follows that to be a controller A must have a motive. Conversely, if A has no motive, then A cannot draw a distinction; without a distinction there can be no content seen to differ in value from any other content, and therefore there can be no value; so the name A calls can indicate neither content nor value, which is to say, cannot be a name; and it follows that A cannot include procedures capable of producing changes in E corresponding to them. Thus <u>A may be a controller if and only if it has a motive</u>.

Controller needs a motive

2.1.2 By definition A is a controller and therefore, by the conclusion of the previous paragraph, must have a motive. Let A's output state at any given moment be the command it gives, and let this state correspond to its

Motives and goals

motive. As AW is defined as a process and must therefore be regarded as 'on-going', our concern is with its behaviour over some succession of moments, a time interval, T say. We therefore consider a succession of outputs of A over T and say that, if there is some internal state z of E that remains invariant under input from A over T, A's motive is invariant over T. We call z A's goal during T, defined as acting to keep (some part of) E in a state z during T. For the present purpose, A may legitimately be taken as part of E.+ It follows that the invariant state z - the goal-state - may thus itself be an internal state of A, z_A say. If this is the case, A's output will be seen to alter the state of its surroundings appropriately for keeping some part of its internal state invariant. In every case, it is only by discovering some invariant state that any goal-state may be identified.

2.1.3 More generally, if there is a subset V_A^* of the set of all the output states V_A of A, such that E's internal states remain confined to some subset Z_E* of their total set Z_E of internal states of E as long as A's

Purposes

output states are confined to the subset V_A*, we shall say that A is purposeful over that set of outputs and define its purpose in terms of Z_E*. More succinctly we interpret $Z_A \in Z_A^* \Leftrightarrow Z_E \in Z_E^*$ to mean that A has the purpose Z_E* provided at least some of the elements of E in Z_E* receive inputs from A.++ It follows that, unless A's internal states are observable,

---

+ Strictly, to avoid problems of self-observation, it is necessary to partition A into a part that entertains and executes the purpose and a part whose internal state is kept invariant. An appropriate partitioning is likely to prove complicated since, as in the case of the models, and models of models of Minsky's formulation (0.2.2.0), the interdependence of the elements of the partition is not clear-cut. Nevertheless, it seems justifiable to ignore these difficulties here, as there is an obvious sense in which an entity (organism) may be seen to be seeking to maintain an invariant internal state. Homeostasis denotes just this activity.

++ Put differently, we should say A had caused Z_E*. This formulation illustratesHume's objections to notions of causation.

the existence of purpose will have to be inferred. Informally, we call a purpose any goal of a system that remains unchanged under alterations of the system's internal state. As a rough guide, we may take a goal to be the aim of a tactic in a strategy whose object is the purpose. It follows from this definition that in general A's purpose will include more than one goal and will be defined over a set of internal states $Z_A^* \epsilon Z_A$ say, and a set of states of the system upon or within which it exercises its purpose, $Z_E^* \epsilon Z_E$ say.

2.1.4 Justification of proposed definition

This definition deliberately makes no appeal to notions of regulation or feedback, because this enables it to avoid constraints on A - such as, for instance, the need for it to have an input, which the existence of feedback would entail - before they have been shown to be necessary. This does not reduce the definition's plausibility or usefulness however, nor lead it to violate any of the demands of the commoner kinds of definition, such as that presented as a list of requirements by Klir & Valach (1965) for example.+ In fact it is equivalent to Ashby's (1964) likening of goal-seeking behaviour to the maintenance of a stable equilibrium, only seen from the point of view of the observer of the entity that maintains the equilibrium rather than of the entity itself, and with the qualification that we must exclude equilibrium-seeking-behaviour of a purely 'mechanical' kind, such as for example that exhibited by a simple pendulum. The distinghishing characteristic of true goal-seeking behaviour, as Sommerhof (1969) makes explicit, is that it places no re-

---

+ Klir & Valach list the following characteristics of a goal-seeking system:
'1. System S, which features goal-seeking behaviour.
2. A goal to which the system is directed by its behaviour.
3. A control element which directs the system towards its goal.
4. A representation of the goal in the control element.
5. Disturbing elements which hinder System S from obtaining its goal
6. A connection between the goal and the control element of System S.' (p.272)

strictions on the initial states of its variables, while in the pendulum these initial conditions are dependently linked: the acceleration depends on the displacement and arbitrary combinations of these variables is impossible. As a definition the one proposed here has the virtue of being based upon what Ackoff & Emery (1972) call 'the common properties of production (which) enable us to define the concepts of <u>function</u>, <u>goal-seeking</u>, and <u>purpose</u> with all the rigor of the concepts used in the physical sciences, and yet retain the core of meaning these terms have gained over the ages' (p.15). It is consonant with Sommerhof's concept of 'directive correlation'. (Sommerhof, 1969, p.174)

2.1.5 The proposed definition presupposes an ordered, multi-goal-seeking system. In this it conforms to the definition of Ackoff & Emery in which a purposeful system 'selects goals as well as the means by which to pursue them' (p.31). Though it does not follow these authors into making this the basis for attributing 'will' as the characteristic peculiar to such a system. For the assumption on which this would depend, namely that selecting goals demands mechanisms basically dissimilar to those required for selecting the individual responses called for by goal-seeking, does not appear sufficiently justified.+ Our view also dictates avoiding the distinction similar to that of Ackoff & Emery, that Sommerhof (1969) draws between 'goal-directed' behaviour, which he takes to be 'objective', and 'purposive' behaviour which in his terms involves conscious striving (p.154) and is therefore 'subjective'. For this latter kind of behaviour we reserve the name 'intentional' defined as 'consciously purposive'.

Purpose is not 'will'

2.1.6 Because it says nothing about V_A* or Z_A*, our definition may convey little

+ cf. The operation of systems such as Winograd's (1972) natural language system (6.3) for example support this contention.

of the quality of purpose as it is commonly understood to reveal itself in 'purposeful behaviour', for instance the quality that Sommerhof (1969)

Purposeful behaviour

calls a 'primary characteristic of all obviously living systems' (p.150). But the definition has no difficulty accommodating such characteristics. Sommerhof suggests that purpose has a hierarchical structure 'in which the goals of the various part-activities (or activities of its parts) are inter-related and integrated' (p.150). He distinguishes 'transient' and 'proximate' goals for each part-activity, arranged in hierarchies of subservience to 'less transient and less proximate' goals of the activity as a whole, which in turn is subservient to behaviour patterns, and so on. Intermediary goals may be concerned with attaining or maintaining conditions suitable for other activities, and the various activities may all be connected in mutual interdependence.[+] Such a view we may interpret as an assertion about the structure of A's output, V_A*. Thus, frequently it will be A's purpose to <u>bring</u> some part of E's internal state within certain limits, rather than simply keep it invariant. In the terms of our definition, this state of affairs, of aiming to bring a system within certain specified confines at some time in the future, corresponds to a purpose of maintaining invariant the <u>movement towards</u> the desired state. There is no difference in principle between holding either a static or a dynamic state of E's invariant. The difference is practical. Normally, purposes connected with changing states entail setting up sequences of sub-goals to be executed in turn, <u>en route</u> so to speak to the main goals.[++]

2.1.7 If A's purpose concerns its own internal state, it may appear to an ob-

[+] Chapter 7 formalizes a notion of purpose similar to Sommerhof's.

[++] The two kinds of purpose, those associated with keeping some fixed state invariant and those with keeping some dynamic state invariant, correspond as aims to the two kinds of regulation that Conant (1969) distinguishes as 'point regulation' and 'path regulation' (8.2.0).

observer to be altering its goals with respect to E. This is likely to occur whenever its internal state varies outside the set of constraints that the purpose defines. This kind of purposeful behaviour characterizes the homeostat. Without the invariance that homeostatic performance provides, the development of purposeful structures of any kind would be impossible, since there would be no way of 'preserving the effects of chance' (Monod, p.32). A's survival is super-ordinate to all other purposes A may have, it is the 'lynch-pin' of A's motivation (George, 1970, p.142) since it is a pre-condition for its achieving any other purpose whatever. A system giving preference to purposes other than the ones that combat threats to its survival will be extinguished with a probability approaching unity with time's progression. Evidently therefore, naturally occurring purposeful systems must incorporate mechanisms for over-riding other goal-directed behaviour in favour of survival goals whenever appropriate. The ultra-stability of Ashby's (1965) homeostat describes what is necessary for just such a performance. An ultra-stable system receiving an input that threatens to force any of its variables outside the limits that mark the boundaries of its viability, will automatically change the parameters governing its transfer function, and continue changing them until it either fails finally or finds a new transfer function that keeps its variables within the required limits.

Ultra-stability a primary purpose

SECTION 2 DISCERNING PURPOSE

2.2.0 Let A be motivated to realize some fixed goal g say. According to our definition, some internal state z of part of E will remain unchanged over that succession of outputs V_A* of A that correspond to the realization of g. As W is the trace that O observes A to make in E (by 1.1.2), we may

What makes W appear to reflect purpose

suppose that W represents a product corresponding to A's purpose - or purposes - up to the time when O makes its observation. As O can observe only E in which it can distinguish the product W of AW, one may ask under what conditions O will suppose that W is the product of a purposeful process.+

2.2.1 Supposing that O can observe W implies that it can distinguish W from the rest of E, we ask to what extent its ability to do this depends upon its own characteristics and to what extent on W's. First we may notice that

Distinguishing W

if E consists of distinguishable elements, O will need the ability to distinguish constraints between them if it is to perceive objects in E, in particular W. W is a 'thing' or 'object' (or 'event') with spatial or temporal extension. 'The essence of ... a "thing", a unity, rather than a collection of independent parts corresponds to the presence of ... constraint' (Ashby, 1964, p.131). Various ways are available for measuring such constraints, such as for instance the mean auto-correlation function computed across some space or interval as, $F(\zeta) = \int_0^\infty f(z)f(z+\zeta)dz$ or $F(\tau) = \int_0^\infty f(t)f(t+\tau)dt$, where ζ and τ represent the space and interval respectively in which the message develops. But such devices provide only the crudest kind of measure, as we shall see when we come to examine perceptrons and the like, and it seems unjustifiable to claim as Moles does that 'the sensation of form is the perception of auto-correlation' (p.88). What is important is that W presents itself as a <u>form</u> or pattern <u>relative to E</u>.

2.2.2 It follows that the notion of a 'thing' is necessarily vague, since it

+ Minsky (1972) notes that when something moves we attribute the movement either to a physical force or a purpose, seldom both.

depends upon what constraints O identifies, which in turn depends upon which constraints are important to it in terms of its goals.+ No 'thing' can be strictly independent of its surroundings, since this would imply that it was a closed system which would make it unobservable. 'Things' are merely relatively independent. We may picture objects as consisting of clusters of constraints, or regions of relatively high 'connectedness' between unitary elements, provided we assume that all elements in an observable world are connected, at least via one another - that is we picture some kind of net-work of inter-connections. This is why what makes an entity a thing must remain partly arbitrary, which is to say, subjective. Though generally, in practice, as we shall see shortly (3.3) the arbitrariness does not present too great difficulties.

W distinguished depends on O's purpose

SECTION 3 ORDER

2.3.0 There is an obvious relationship between the notions of constraint and order in a system. Von Foerster (1960) interprets Shannon's definition of 'redundancy' in such a way as to provide a measure of order; Ashby (1972) relates constraint, order and variety; Lofgren (1967, 1969, 1972) associates order with the existence of descriptions, which consist of programmes shorter than the sequences they generate. (These descriptions, as we shall see, he equates in turn with explanations; learning, explicability and theories, with various kinds of 'order extraction'.) Popper (1959), Von Foerster and Lofgren (1967) agree that it is desirable for any measure of order to reflect the relative connotations of the term as

Concept of order

Von Foerster

+ 'Noise can be defined only on the basis of intent' (Moles, p.100), and therefore complementarily the signal that stands out against the noise as background. Moles notes too (p.77) that numbers of 'forms' co-exist which have high or significant auto-correlations in their own domain though null correlations with one another.

it is commonly used. Among these is the notion that order may vary from absolute order on the one hand to absolute disorder on the other. Von Foerster interprets Shannon's definition of redundancy to make it furnish such a measure. In the formula

$$R = 1 - \frac{H}{H_m}$$

in which $\frac{H}{H_m}$ is 'the ratio of the entropy H of an information source to the maximum value it could have while still restricted to the same symbols' (p.37), R may be taken as the desired measure. It will vary from zero, for H equal to H_m (maximum entropy), to 1 when H is zero, that is when the position of one element of the system is sufficient to determine the positions of all the others. Löfgren (1967) formalizes Von Mises' <u>axiom of randomness</u>, cited by Popper, in terms of a universal Turing machine. The axiom of randomness is sometimes called 'the principle of the excluded gambling system'. It postulates that among the classes of sequences of events capable of indefinite extension - tosses of an indestructible coin, for example, - those that are random would be such as to make it impossible for a gambler to find any <u>system</u> of betting on their individual outcomes such as would improve his chances of winning. Löfgren (1967) uses a formalized concept of description as a basis for rigorous formulations of 'intuitive notions' of order and randomness. According to Solomonoff (1964) we say that 'S is a description of T with respect to machine M', if $M_1(S) = T_1$, where S is the input string to the machine and T its output string. Lofgren's definitions of order and randomness derive from this definition. They are respectively

Von Mises

Solomonoff

Löfgren

> ' ... there is a U such that for any two equally long patterns ξ and ζ, ξ possesses more order than ζ if and only if $s(\xi,U) < s(\zeta,U)$.'
>
> ' ... there is a U such that for any two equally long patterns ξ and ζ, ξ possesses more randomness than ζ if and only if $s(\xi,U) > s(\zeta,U)$.' (p.165)

where U is a univeral Turing machine that may generate the sequences ξ and ζ from different starting tapes, such starting tapes being known as the U-forms of these sequences, and where $S(\xi,U)$ in the shortest U-form of ξ, and $s(\zeta,U)$ of ζ.+

2.3.1 But none of these definitions indicates any effective general test for regularity; or, a fortiori, how to find s such that we may be sure that there is no s* say, such that $s^* < s$, not even how to know s when we have found it. 'Thus our tests of randomness are never tests which exclude the presence of all regularity' (Popper, p.359). So all we can do is apply successive tests for specific regularities though never knowing when or whether all regularities have been revealed. We have a criterion of success in our search, namely a contracted description - U-form, - but our goal, the shortest description, is specifically what is not known. Thus failure to reveal regularities after a number of tests does not mean that some regularity will not emerge when the next test is applied. Indeed, in an arbitrarily long sequence, we should perhaps expect to find at least some regularity, since 'No arbitrary long randomized sequence can be generated effectively' (Lofgren, 1969). On the face of it this conclusion of Lofgren's appears to conflict with Popper's assertion that 'we can develop a theory which allows us actually to construct ideal types of disorder' (p.359). But, by a 'type' of disorder Popper means one defined as the failure to discover by testing some 'specific regularities'. In agreement with Lofgren, Popper concedes that 'there are no tests for presence or absence of regularity in general' (p.359). So the observer remains bound by the limitations of the specific tests for regularity that he applies, the success of which, registered as the discovery of regularities, he denotes by phrases such as the 'discovery of order', 'the recognition of pattern' or 'form', and the like.

Tests for regularity

+ s is thus totally random since by definition there is no s* say, such that $s^* < s$.

2.3.2 The considerations of the preceding paragraph imply that order and disorder (randomness) are both relative terms, and may be regarded as predicates assigned according to the results of the application of particular tests. In so far as such tests must necessarily be conceived by some observer, we may take the concepts to apply relative to an observer.

Order is relative to O

> 'By saying that a factor is random, I do not refer to what the factor is in itself, but to the relation it has with the main system. Thus the successive digits of π are as determinate as any numbers can be, yet a block of a thousand of them might serve quite well as random numbers for agricultural experiments, not because they are random but because they are uncorrelated with the peculiarities of a particular set of plots.' (Ashby, 1964, p.259)

Notice that this view does not contradict Popper's injunction against the temptation to attribute randomness to lack of knowledge as to the order prevailing, if any order does prevail. If anything it clarifies his objection. Advancing this clarification further, Ashby (1964) points out that the observer of the output of a machine dependent upon an input of 'random numbers' will see the machine as determinate or indeterminate according, respectively, to whether he is aware or not of the numbers input. Consistent with what was said in 2.2.2, that what order is distinguished will depend upon the observer that distinguishes it, this implies that Lofgren's definition of order (2.3.0) must also be interpreted relative to an observer. If Ashby's notion of randomness is to be taken as equivalent to Lofgren's it follows that the magnitude of the correlation, in Ashby's case must be proportional to the ratio of the length of the U-form to that of the sequence generated in Lofgren's. We achieve this result by the following interpretation: in Ashby's case we interpret the agricultural experiment as the 'observer' and the successive digits of π as the sequence observed. Lofgren's use standardizes the observer as a universal Turing machine. The tape required to 'convert' some given Turing machine into a universal one describes the particular Turing machine in question. Lofgren's case thus becomes equivalent to Ashby's, if we substitute in Lofgren's definition of order a particular Turing machine for the universal one.

2.3.3 The relation between observer and order observed is the reason why, in practice, as we remarked in 2.2.1, little difficulty generally results from the evident arbitrariness with which 'things' are delineated and distinguished from one another; it is because the things distinguished in the first place are distinguished as a result of correlations between them and the observer that distinguishes them. Which is a manner of saying that individuals evolve in ways appropriate to their environments. Fogel, Owens & Walsh (1966) have 'evolved' machines in specific simplified environments (6.6). They conclude from their experiments that 'each living creature may be viewed as a tentative model of some significant aspects of its environment' (p.111). More generally one might put it that every entity of the kind able to maintain equilibria within the limits necessary to ensure its survival under conditions of a range of disturbances, - entities showing in some degree the characteristic that we have defined as intelligence - is a description of those aspects of its environment that have given rise to the disturbances that make up its history; and that the quality of the description might be defined as having been good enough for the entity's survival so far. It follows from this as a corollary that individuals evolving in similar environments are likely broadly to share some at least of their form-discerning propensities, thereby restricting the arbitrariness of the discernment of order referred to in 2.2.1, at any rate among the inhabitants of the same world. In general it is only changes in resolution level that lead to difficulties in delineating things (7.2.3), and this we should expect as such changes are equivalent to altering an entity's environment.

Individuals and environments

SECTION 4 CAUSAL EXPLANATIONS

2.4.0 Having distinguished W from the rest of E, suppose O now attempts to account in some way for W's existence, to explain how W came into being say, to provide what is called a 'causal explanation' of W. A priori there are no restrictions on what O might postulate. For instance, O might assert that W had come into existence of its own accord, out of nothing. An objection to such an explanation might be that it was 'subjective' and not communicable to or understandable by anyone at all, probably not even the individual making it. To be communicable in the sense of being understandable, an explanation would have to be able to win wider approval. O would have to be able to sustain it against criticism (Craik). Ashby (1962) points out that communication implies constraint; an event at A is communicated to B if its occurrence at A limits subsequent possibilities at B. Extending this idea to the notion of explanation, we might say that an explanation were communicable - more properly, understandable - in proportion as stating it limited the range of possibilities flowing from the explanation. For instance, the unqualified statement 'There is W' conveys almost nothing about W. Like a description, an understandable explanation E of W restricts the kind of thing W might be. The question is, How? Obviously explanation and description must be related. They have in common the adducing of consequent from antecedent conditions. Thus, according to Popper (1959)

Accounting for W's existence

> 'To give a causal explanation of an event means to deduce a statement which describes it, using as premises of the deduction one or more universal laws, together with certain singular statements, the initial conditions.' (p.59)

In Lofgren's (1972) terms a relatively effective explanation 'is effective in relation to... an interrogator in the same sense as a program is effective in relation to a computer. It permits the interrogator to check (like a computer) every step in the chain of arguments' (pp.343-4). Formally this view, with its overtones of Descartes' Method, expresses an hypothesis which 'cannot itself be proved because the meaning of "effectively understandable" is only intuitively understood' (p.345). Lofgren gives formal expression to the hypothesis as

> 'EXPLANATION HYPOTHESIS I. If an explanation $\underline{E}$ is effectively understandable (relatively effective), then it is understandable in terms of the rules (explanation arguments) and axioms (postulates) which constitute an r-formal theory $\mathfrak{T}$, such that $\underline{E}$ is a p-explanation in $\mathfrak{T}$' (Lofgren, 1972, p.345).+

Newell & others (1958) extend the detail of this view, distinguishing 'specification by state description' from 'specification by process description' (p.30) and illustrate the distinction with a number of examples. We may either, they assert, write down an explanation in logic or stipulate the operations on the axioms (the proof) that will produce it. This distinction that these authors draw is equivalent to Spencer-Brown's distinction between 'injunction' and 'description': a musical score is a set of injunctions to a performer which, if he follows them, will cause him to produce the sounds that <u>are</u> the musical composition.

2.4.1 But O faces an immediate difficulty, because of all the explanations he may offer, few are likely to be theorems in formal theories. Fortunately though, this difficulty is not necessarily insurmountable. For suppose, for example, O's explanation were a statement in a natural language, it might still be made effective in Lofgren's sense. Many probably most, relatively effective explanations must have begun as natural language statements, from which embryonic existence they gradually acquired their more rigorous form. The question is how this formative process was accomp-

Making explanations effective

+ 'A sequence $\underline{E}$ is a p-explanation in $\mathfrak{T}$ (relative to $\mathfrak{T}$) if $\underline{E}$ is a proof sequence in $\mathfrak{T}$ and $\mathfrak{T}$ is r-formal.' (Lofgren, 1972, p.344)

'<u>A formal theory</u> $\mathfrak{T}$ is a set $\underline{A}$ of wffs, called the axioms of $\mathfrak{T}$, together with a set of predicates, the <u>rules of inference</u> of $\mathfrak{T}$. When $R(\underline{Y}, \underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n)$ is a rule of inference of $\mathfrak{T}$, the wff Y is said to be a consequence of the wffs $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ in $\mathfrak{T}$ by R.' (p.399)

'A formal theory $\mathfrak{T}$ is r-formal if $\mathfrak{T}$ has a recursive set of axioms and a finite set of recursive rules of inference.' (p.396) lished. Is there an effective procedure for it? Von Mises offers the concept of connectibility.

> 'We call a group of words (sentences and sequences of sentences) "connectible" if they are compatible with a system of statements which, it is assumed, regulate the use of language - connectible ie., with respect to this system.' (p.73)

The 'place of linguistic rules' is taken by 'the totality of sentences, in which a given sentence is 'embedded' (p.76). The similarity between this concept and that of a 'formal theory' is obvious. Where two sentences are not connectible with one another, this must be because they belong to separate systems which need to be in some way 'joined' together to form a single system, if they are to be made connectible. But Von Mises does not say what makes two groups of words 'connectible' with one another in the first place. Within Lofgren's framework one would 'join' two expressions by deriving them from the same rules and axioms.

2.4.2 Explanations of Lofgren's kind are determinate. Within Lofren's framework an explanation would win wider approval or be sustained against criticism (Craik's requirements, mentioned in 2.4.0) if new sets of rules and axioms could be found that generated both the criticism and the explanation that it criticized. Thus if an explanation were not effectively understandable it might nevertheless be partially, and be made more, understandable, by making it sharper while at the same time broadening its basis. Which is to say by (1) increasing the explanation's restrictiveness through restating it in such a way as to impose narrower limits on the range of possibilities flowing from it (that is by increasing its determinacy), (2) incorporating in it, so to speak, criticisms made of it, (including the elimination of inconsistencies and contradictions, either by dropping elements of the explanation, or, where this proved impossible, by means of finding new axioms and rules from which to generate it). But notice that no effective procedure exists for carrying out this process, because, by

Determinacy in explanations

2.2.1, no effective procedure exists for finding regularities, and it is precisely upon the discovery of regularities that (1) and (2) depend. For even if we were to postulate that some statement of O's was a wff in some formal theory unknown to us, joining together numbers of separate wffs from separate theories would require finding some regularity of pattern such as would permit the separate wffs to be derived from common rules and axioms.

2.4.3 Thus, if O asserted that W had come into existence of its own accord (2.4.0), we might argue for example that, to have done so, it would have needed to be self-organizing in the strict sense of arriving at its organization - being the thing it is - without interaction with its environment. We might furnish a proof that this was impossible by recalling Von Foerster's (1960) argument to demonstrate that the existence of such a strictly self-organizing system would violate the Second Law of Thermodynamics.+ Or we might point to a similar conclusion that Ashby (1962) reaches from consideration of the properties of machines, without appealing to laws deriving from other theories.++ (N.2.3)

Unsatisfactory explanations

\+ Monod describes the growth of Escherichia coli in a suitably prepared medium, consisting of a few milligrams of sugar and certain mineral salts in a millilitre of water. Thirty-six hours is sufficient to permit the growth of several thousand million bacteria. 40% of the sugar will have been converted into cellular constituents, the rest, oxidized into carbon dioxide and water. If the experiment is carried out in a calorimeter, the entropy of the whole system (bacteria plus medium) will be found to have increased slightly more than is predicted by the Second Law of Thermodynamics. 'Thus while the extremely complex system represented by the bacterial cell has not only been conserved but has multiplied several thousand million times, the thermodynamic debt corresponding to the operation has been duly paid'.(p.29)

++ Ashby distinguishes two kinds of self-organization. The first, which he characterizes simply as 'self-connecting', may come into existence of its own accord, but will show no properties that relate to (correlate with) any other system (and therefore, in particular, any observer), and will thus always appear to be random (by 3.2); that is not organized, and so a fortiori not self-organized in the strict, non-ineracting sense. In the second kind of self-organization, the observer of the self-organizing system distinguishes 'good' organization from 'bad'. 'Good' organization of a system is that which is in some way appropriate to (correlated with) some second system. As this appropriateness could only come about by changes affecting the functional relationships between the states of the first system (the one that appears to be self-organizing), they could not at the same time depend on those states. Thus they would have to depend on the states of the second system. Organization would therefore occur as the result of this second system and would thus not be self-organization in the sense described.

2.4.4 If O's choice of what he asserted as his explanation - namely that W had come into existence of its own accord out of nothing - were dictated by a desire to be understood as widely as possible, he would have to choose that set of rules and axioms from which he could generate explanations of the largest possible number of things and events. For it is this set that would have the greatest chance of encompassing the explanations of other Os within its own terms, of making its own explanations connectible with other explanations. Thus, faced with an objection to its explanation of W that was rooted in thermodynamic theory, O would be obliged by 2.4.2 either, (1) to provide a new explanation of W, understandable within the terms - axioms and rules - of a theory that might accommodate the theory of thermodynamics also, or (2) to hold to its assertion in disregard of criticism of it, or (3) to abandon its assertion of W's spontaneous creation and concede that W must have come about as a result of some other, previously existing system. Alternative(1), if O can achieve it, is the equivalent of O's finding a meta-language to accommodate its own explanation and criticisms of it. It may be however that O's own language is insufficient even to understand the criticisms of it, which are themselves meta-linguistic for O. If O opts for (2) in the face of objections, it may be for this reason. But if it is not, that is if O can understand the meta-language of the objections to its explanation and still disreards them, we should say explanation was not O's aim. For what O said would offer none of the qualities of explanation that we have supposed to be desirable, and this would contradict the assumption with which we started, namely that O wished to be understood. It follows therefore that O will either have to find a single explanation to encompass simultaneously thermodynamic theory and the notion of spontaneous creation, or postulate some entity antecedent to W as W's causal explanation.

Communicability of O's explanation

SECTION 5 CAUSAL EXPLANATIONS AND EFFECTIVE EXPLANATIONS

2.5.0 Let us take it then that O supposes W has been brought about be the action of A. A suitable A would have to be the controller of some process AW that mapped A into E in the form of W, and by the conclusions of section 2, this process AW would have to be purposeful, with its purpose inhering in A. In other words, O's assumption would be equivalent to supposing that W were the observable trace of the operation of certain stable processes, and therefore, by the definition of purposeful (2.1.3), that it must represent the realization of the purpose or purposes inhering in them.

Antecedents

2.5.1 A definition of Lofgren's provides a formal basis for describing O's concept.

Productiveness

> 'An object A is productive in a surrounging S if A causes S to produce another entity B, symbolized $A \rightarrow d_S \rightarrow B$ and read as follows: the configuration (output state) of A forces S to produce B. Here d can be considered a description of B relative to S.' (Lofgren, 1972, p.358)+

We recall from 2.3.0 the formal concept of description as a programme that computes; that is as equivalent to the tape expression S (the description) from which a machine M will compute (describe) a sequence T. Interpreting this definition in terms of the previous paragraph it turns out, as we should wish, that it corresponds to the paradigm for a productive procedure adopted earlier (1.1.0). B may be taken as equivalent to the W O distinguishes, d to the purpose of which W appears to O to be the realization, and A to the entity A in which O supposes the purpose inheres.

2.5.2 So if O seeks a causal explanation of W, that is an explanation by means of an antecedent, then O will interpret W as definingthe antecedent's pur-

---

+ Notice that A's relation to S corresponds to the 'interaction' between what reproduces itself and the surrounding in which this occurs, by which Ashby (1962a) characterizes the process of self-reproduction.

pose. As the purpose will in turn determine how O defines the antecedent's characteristics, it follows that O will define A according to the W it distinguishes. In the broadest sense therefore, the way O defines A will depend on O's capacity for finding descriptions for E. This is because W is part of E and therefore O's descriptions for E will determine what Ws it discerns. In general, at best, the W O distinguishes will only partially accord with Ws distinguished by other Os or the W distinguished by the A that produces it (cf.1.2.2). In particular O's resolving power (7.2.3) and variety - temporal and spatial - will be determinants of the forms it distinguishes, that is its descriptions of W, and these descriptions may be unrecognizable as W to other Os. For instance if W were some random sequence of numbers, O's description of it might consist of the mean of that part of the sequence it had observed. Such a description would provide a sufficient rule for O to produce another sequence W* say, with the same mean as W, though possibly having no element the same as W. For O, W and W* might be indistinguishable, though another observer O' say, with a different description of W, based perhaps on W's specific elements, might discern no likeness between them.

Causal explanations

2.5.3 Thus going back to 2.4.0, if O distinguishes a form and wishes to give it an effectively understandable causal explanation, his explanation will have to postulate a purpose that the form defines. Moreover it will have to postulate an agent as the producer of the form and attribute the purpose to the agent. The identity of the agent will thus have to be consonant with the form O distinguishes and will therefore depend on O's capacity for distinguishing forms, that is of finding descriptions of E. In other words, the form O distinguishes will depend on O's characteristics.

Existence of W presupposes purposeful agents

SECTION 6 PLAUSIBILITY OF EXPLANATION (N.2.4.)

2.6.0 In general O will be unable plausibly to explain or account for W's existence. Which is to say its account will not be effectively understandable. Generally O's explanations are likely to be arbitrary, lengthy and complex, identifying somewhat tenatively, if at all, not some unitary system or 'entity' as W's antecedent, but numbers of more or less independent entities, none of which may be even individually plausible but merely the best O can do. The plausibility or otherwise of O's explanations will depend in the main upon their 'determinacy'. We shall call O's explanatory entities (the antecedents of W that O postulates because it cannot find a single one) determinate, if and only if, from the same initial conditions, they always lead to the production of the same W. Antecedents that behaved in an indeterminate way, sometimes leading to one and sometimes to another consequence, would be considered unsatisfactory.+

Lack of unitary antecedent for W

2.6.1 Thus for example if O were human and W were a stone, asked to account for the existence of the stone, O might be unable to attempt any account. Persistent interrogation might elicit from him a catalogue of antecedent conditions and events, including possible references to wind, water, plant growth, bacterial activity, temperature changes, movements of the earth's crust, gallactic upheavals and so on, in a sideways- and backward-stretching net. The net might extend as far as O were able to conceive, in some more or less arbitrary, subjective way, of more or less suitable antecedent entities; namely those characterized by the possession of such equilibria (goals or, in their more complex form, purposes) as appeared

Multiplicity of antecedents implies disorder

+ In this sense it may be misleading to speak of 'entities', A, or their purposes at all, since the concept of an entity depends upon the notion of an underlying unity. For consistency however, until we are better able to distinguish what constitutes such a unity, we shall retain these terms.

to him might have led to the emergence of those objects that he wished the entities to 'explain'. But even supposing each element of O's explanation were determinate in the sense described, simply the number of elements would quickly make his explanation a difficult or impossible predictor to handle, unless he furnished with it information on the relations between its elements, the constraints binding them; that is, unless he in effect reduced the number of independent elements in his explanation, by joining them together to make larger and fewer independent elements - by Lofgren's definition in 3.0, found more order.

2.6.2 But for O, the constraints that make the stone a 'thing' (2.2.1) are not such as enable him to point to some unified order that they reflect. (By 2.4.4 and 2.5.0) O attributes the arbitrariness of his explanation to the thing he explains, so that, from his failure to account for it plausibly, the stone acquires an 'accidental' quality for him. It appears to him no more than the fortuitous accretion of numbers of purposes diffused throughout a productive system that consists of numbers of separate, in the sense of more or less independent and unconnected elements, which has come about largely for reasons simply of their chance coalescence. We suggest in 2.7.1 that the reason for O's impression is connected with the amount of 'work' this kind of explanation leaves to him, compared with more 'satisfactory' explanations. Such an impression would show an intuitive appreciation of Occam's razor, 'an interpretation of which is that the more "simple" or "economical" of several hypotheses is the more likely' (Solomonoff, p.3), so that conversly the less economical more complicated explanation would come to seem more fortuitous. Solomonoff goes on to explicate the concepts of 'simplicity' and 'economy' by means of Turing machines, the 'simpler' of two hypotheses being the one with the shorter 'description', as defined in 2.3.0. Thus, as we concluded in the previous paragraph,

Disordered explanation reflects on what it explains the simpler description is also the more ordered.

2.6.3 Because O fails to provide constraints binding the antecedent entities it postulates for W, these entities cannot provide the determinacy required of an explanation (3.6.0). There is little about O's explanation that fits it to this particular stone; it would do, it seems, equally well for almost any stone. Moreover this vagueness is likely to grow roughly inversely with the explanation's 'usefulness' (shortness, relative simplicity), at least in relation to the particular stone in question. Conversely the more specifically O attempts to make his explanation refer to this particular stone, the lengthier, more complex and 'useless' it is likely to become.

Length of explanation and usefulness

2.6.4 In important respects, as we have said, the difficulties with O's explanation depend upon the antecedent entities he chooses for it. Though it is not necessarily the case that choice of more rigorously definable antecedents would alone be sufficient to arrive at a shorter, simpler or more specific explanation. An account of the stone made in terms of rigorously expressed chemical and physical laws for example would not necessarily do so, because, in fitting such explanations to this particular stone, they would become unmanageably long Even if the entities O postulated were all well-defined and behaved in a well-defined way, in their respective environments, objections to O's explanation would be bound to persist, because they result from the stone's apparently - to O - 'accidental' nature; that is from O's inability to detect regularities in it such as would enable him to furnish a unified explanation. To meet the objections, O would need to find a shorter and at once less vague way to account for the stone. The degree of his success would correspond exactly with a reduction of

Explanation and antecedents' relationships the stone's evidently accidental quality. That is the discovery of some single system or entity as the stone's antecedent, by furnishing a set of constraints - namely those that make the single entity single - that might be mapped onto the stone, would reduce the stone's apparently accidental quality.

SECTION 7 DESCRIPTION AND EXPLANATION (N.2.5)

2.7.0 While it is true that O's inability to account properly for W's existence makes W appear to have an accidental quality, it is true too that any regularity that O distinguishes in W identifies as an entity any antecedent that O may postulate to account for it. In this respect O's accounting for the stone's existence closely resembles his description of it as (by 2.4.0 and 2.5.0) we should expect. Were O able to discern regularities in the stone's appearance such as would enable him to identify this particular stone, he would be able to attribute the regularities to the outcome of a purpose and the purpose to some antecedent entity. That is, he would be able to suppose that the regularities of his description existed as the result of some pre-existing regularity, such as the equilibrial state of a pre-existing, purposeful entity. So the more concise O's description of the stone, the more concise (and therefore less accidental-seeming) his explanation of it. This would be the case even if, initially at any rate, O had characterized the entity he postulated simply as an 'entity A such that it produces W'. That is simply as some pre-existing coherence, reflecting a coherence he had discovered in W.

Regularities in A reflected in W

2.7.1 If, apart from W, O identified a dynamic entity A, whose purpose (set of equilibrial states observed as output) correlated with W, then O would use A to 'explain' W's existence. So for example the products of reproduction or 'self-reproduction', the off-spring of biological species say do not have the accidental quality of the stone; though compared with a stone, accounting for a member of a biological species, other than by pointing to its parents, would probably require an even longer and more complex explanation. What makes the difference between the two explanations is not that O has a more precise idea of the connexion between the biological offspring and its parents than between wind, water, &c., and the stone. It is that, in the case of the biological offspring, identifiable entities (the parents) exist, into which the offspring map, apparently with little residue, while no equivalent mapping is available for the stone. Such entities, - the parents - being true (coherent) entities, may themselves be thought of as determinate systems whose internal workings O is at liberty to ignore, since they occur <u>independently</u> of him. While the 'entity' wind, water, &c., into which he maps the stone, being an arbitrary collection of separate elements, requires that O should as it were operate it himself by defining the elements' relationships.

Correlation as explanation

2.7.2 It appears that we are here concerned with the <u>accumulation of purposes</u> such as enables them to be transmitted or realized as a body. What the purposes <u>are</u> or <u>how</u> they are transmitted is of secondary importance. An example of an entity able to 'accumulate' purpose in this sense is provided by Fogel & others. These authors have 'evolved' machines to perform particular operations in certain well-defined environments. One such machine for instance is required to predict the next input state of an environment whose input states consist of a cyclically repeated string of randomly distributed zeros and ones. Simplified, their method is to

Correlation by evolution

begin with some arbitrary machine (finite automaton), (1) let it run for a time, (2) measure its performance, using a measure based on the number of its outputs, zero or one, that are the same as the next input, (3) 'mutate' it, by making some random alteration in its internal or output states, and repeat steps (1) and (2). If the new machine performs better than the old one, producing more outputs the same as the next input than the old one did, it is retained. Otherwise, (4) the old machine is restored and a new mutation tried. Steps (1) to (4) are repeated until improvement in performance ceases, by which time the machine may perfectly predict its environment . If this occurs the machine's and its environment's outputs will be identical. At least they are likely to be very similar. The machine is then itself a perfect or very close description of its environment, one that in general is more concise than a simple enumeration of the environment's output states would be.+

2.7.3 Taken alone, the outputs of the machine are precisely as random as the string of random numbers it has been evolved to predict. To O they appear 'accidental', the products of chance, like the outputs of the environment.

Discovery of A-W correlation provides O's explanation

But if O notices the correlation between the two sets of outputs, it is likely to postulate a casual link between them.++ Irrespective of any postulated causal link with its environment however, the machine is likely to be supposed purposeful, simply because the alternative explanation, that the observed correlation between its outputs and the environment's came about by chance, in unacceptable. And this will be because O's in-

---

+ 'My hypothesis is ... that thought models or parallels reality - that its essential feature is not 'mind', the 'self' 'sense-data' nor propositions but symbolism, and that this symbolism is largly of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation'. (Craik, p.57)

++ In practice any causal link actually postulated is likely to depend upon temporal arrangements. For example, in the present case, A would probably see the machine evolved to predict its environment's outputs as if it controlled them.

tuitive definition of purpose matches the more rigorous one proposed in 2.1.2. Notice that the machine's evolution has come about as a response to the environment as a whole, particularly its cyclic character, not to its individual outputs. But this does not affect O's deeper assumption, namely that the correlation observed did not occur by chance but on purpose. It does not matter to O that he may be incapable of describing the machine beyond pointing to or naming it as the producer of the output observed. The machine itself might simply be a black box marked with a label or code number enabling O to identify it. It is always possible for O to find an explanation that appears satisfactory. Broadly speaking O either identifies an object or state of affairs by finding a description of it, and thenpostulates a matching purpose of which the object is supposed to be the realization. Or O identifies some purposeful entity which enables him to distinguish objects or states of affairs to which the entity's purpose is supposed to have given rise.+ In either case, the description of the object and that of the purpose are morphisms of one another.

Recapitulation of O's procedure

2.7.4. Recapitulating O's procedure when it sets about accounting for W's existence: first it distinguishes W as an entity in E. What makes W an entity is the constraints existing between its parts. O's distinguishing W depends upon its discovering these constraints, which is equivalent to O's finding a description of W sufficient to enable it to name W or delineate it in some way. To account for W's existence, O postulates some pre-existing, purposeful entity with which it supposes W stands in some kind of correspondence. The entity is equivalent to and may be an explana-

+ We have noted by implication in 2.2.2 that the object or purposeful entity O distinguishes will itself depend upon O's purpose.

tion. Finding an entity suitable for this object, consists in finding one that incorporates a procedure whose operation on E will be sufficient to bring W into existence. Provided O is satisfied that the entity is determinate and that it does in fact bring W into existence, O may take the entity as, in this sense, sufficient to account for W's existence.

SECTION 8 IMPROVING EXPLANATIONS

2.8.0 As we have asserted, an account by O depending upon such an entity may be regarded as equivalent to an 'effectively understandable' explanation (2.4.0), in the sense that it may be characterized as a sequence generated by an effective procedure, even though the effective procedure belongs to the entity and not to O itself. Such an explanation is (1) repeatable (determinate), (2) equivalent to the sequence that a universal Turing machine would compute given the proper starting tape, namely one consisting of the axioms and rules of the appropriate r-formal theory. Moreover an effectively understandable explanation of W would be that part of any sequence, from its start up to and including the p-explanation of W, provided that it was included in a finite segment of the total sequence. (If it were not, the explanation could not be effectively understandable, since each step requiring a non-zero time to be understood, the explanation would never come to an end.) Thus there is an equivalence between the explanatory entity O postulates and an effectively understandable explanation, in the sense that, given the proper initial conditions (either the entity or the appropriate starting tape for the Turing machine), the end conditions (the required explanation) will come about (automatically) by the execution of a determinate procedure. Both kinds of explanation, Lofgren's and O's, represent effective procedures.

O's account is relatively effective

2.8.1 But we have seen that a relatively effective explanation is not necessarily effective for any particular given observer O. The variety required for an effective explanation may exceed O's capacity as a channel (see 7.3.3), and even if it does not, we have observed (2.6.0) that O's explanation is unlikely in general to be effectively understandable. Now suppose $\mathcal{E}$ is an explanation that is not effectively understandable, then, by Lofgren's Explanation Hypothesis I, it is not a p-explanation in an r-formal theory. Nevertheless (recalling 2.4.3), it may still contain 'parts' that are or could be made effectively understandable, each perhaps consisting of a sub-sequence generated by an effective procedure from the set of rules and axioms peculiar to that sub-sequence. (In the limiting case the sub-sequence might consist merely of individual elements.) Such a 'partially' understandable explanation would become '<u>more</u> understandable', in the sense in which we are using the term, if a new set of rules and axioms could be found from which it were possible to generate more than one of the sub-sequences. That is, if the number of sub-sequences could be reduced by joining sub-sequences together to form longer sub-sequences.+

Extending explanations

2.8.2 This indeed is in accordance with the way O is bound to proceed in order to increase the relative effectiveness of an explanation. For, recalling the example of the stone, we see that O's procedure consists in naming antecedent entities, wind, water and so on, which we take to be equivalent to selecting, by their code numbers, appropriate starting tapes for a universal Turing machine for generating particular sequences, or of choosing finite automata, like those evolved by Fogel & others (2.6.4), such as have outputs of the kind required. And, supposing for the moment that we

+ The idea that this formulation expresses occurs more or less equivalently in other formulations such as Beer's (1972) in which an explanation's increasing understandability corresponds to rising through a hierarchy of meta-languages. Compare also Von Mises' view on connectibility(2.4.1); or Schrodinger's (1947) two 'mechanisms' for producing order, 'The statistical mechanism which produces order from disorder and the other one producing 'order from order'.'

are justified in assuming this equivalence, we may regard O's explanation as a collection of names arranged in a sequence with some degree of order relative to O. In this case, the understandability of O's explanation would increase as O discovered more order in the sequence, that is as O found entities capable of generating ever longer segments of the sequence. If our assumption is not justified, that is, if the antecedent entities O names turn out to be inconsistent or indeterminate through being insufficiently specified, then (by 2.4.1) these shortcomings may be remedied, as we have shown.

CHAPTER 3

AUTHORSHIP

SECTION 0 AUTHORSHIP AND ORDER-EXTRACTION

3.0.0 The previous chapter, together with the one preceding it complete the preliminaries to the definition of authorship. The definition adopted makes authorship transmit order but reserves the terms creative-authorship, for authorship that generates order, though with the qualification that generation of order is not an autonomous process, but occurs within a surrounding, in a manner similar to 'self-organizing'. The distinction between the two kinds of authorship provides a basis for answering questions like Minsky's (1972), '<u>When can one credit a machine with solving a problem and when must one credit its designer or programmer</u>?' (p.7) It has the advantage too that it permits emphasis of the gradual, evolutionary character of the 'creative process'.

Levels of authorship

3.0.1 The concept of order extraction and order-extraction work provides the basis for the definition of author. A purpose in A to produce W requires order in A, and the purpose makes A W's author; in so far as A has acquired the order that it passes on to W by its own order-extraction work, it is said to have created W and is called W's c-author. Definitions follow of partial authorship and there is discussion of the subjective nature of the definitions of the evolutionary character of author as defined.

Purpose

SECTION 1 APPROACH TO AUTHORSHIP

3.1.0 As the distinguishable 'thing' W represents a mapping of A into E – or of that aspect of A that we may regard as the purpose that led to W's production — we shall call A W's <u>author</u>. Although A will be embodied in some sort of physical hardware, it is not the hardware itself, but its organization that is of interest to us. This is the reason why we defined A as a set of procedures (1.1.0). Where A is a single coherent entity, our use of the term author will be seen to correspond to its normal dictionary meaning (1.0.3). A's coherence, as we showed in the previous chapter, will depend upon the coherence of the purpose that O discerns in W. If, as in the case of the stone in the preceding chapter, a human observer discerns little coherence in the purpose realized in W, he will attribute authorship to a series of entities A_1, A_2, ...

Author and controller

3.1.1 A number of simplifying assumptions have been made in arriving at this definition of author. First, it has been tacitly supposed that authorship may be regarded as some kind of step-wise process whereby A is activated and W emerges on an all-or-none basis, emitted as a single output. It will become clear from more detailed examination of the authorship process (chapter 7) that this assumption is not necessary to the definition and has been made simply for the sake of initial simplicity. Later the restriction will be relaxed. Secondly, and consonantly with this assumption, we have treated A and W as if they were unitary entities. Moreover, although A's purpose has been distinguished from A itself, the matter has been left vague. Clearly we shall ultimately want to be able to discuss complex and complicated As and Ws. Later these simplifying assumptions will be justified and the constraints they imply relaxed, without damaging the concepts concerning which they have been made. Thirdly, we have so far avoided all mention of how A exercises control during the production procedure that gives rise to W. Regulation and control is the subject of chapter 8.

Assumptions underlying controlled approach

SECTION 2 ORDER EXTRACTION

3.2.0 Implied by the proposed definition of authorship is the notion that whatever results in the production of W may be regarded as the outcome of an accumulation in A, before W was produced, of the regularity that W reflects. That is as having been part of A's organization or structure.

'Learning programming hypothesis'

Now since A is an entity of some kind, possibly a loosely bound set of objects or events, it has some form of organization, which it must either have acquired or possessed when it came into being.+ In the latter case, A itself may be thought of as a W produced by some previous A, to which similar remarks about its organization apply. Thus the purpose realized in W must have been either acquired by A or channelled through it, although the possibilities need not be mutually exclusive, and generally will occur together.. Lofgren distinguishes these two cases as forms of <u>order extraction</u>.

> '<u>Learning-programming hypothesis</u>. An object <u>A</u> can learn from a surrounding <u>S</u> if <u>A</u> can extract order (regularities) from <u>S</u>. The more order that is extracted (the shorter the description of <u>S</u> produced by <u>A</u>), the more genuine is the quality of learning. The amount of work done by the learning mechanism (the order-extraction mechanism) in A represents the amount of learning done by A and is hence a subjective measure. If A obtains the properties of <u>S</u> without any proper order-extraction work, <u>A</u> is said not to have learned <u>S</u> but to have been programmed by <u>S</u>. (Löfgren, 1972, p.356)

Notice that, according to this hypothesis, order-extraction work may be regarded as a necessary, though not a sufficient condition, for learning, as order extraction can occur without learning. A procedure capable of solving problems or recognizing patterns for example is not necessarily modified by the results it achieves, except at the instant when it reaches its solution or makes its recognition.

+ In 2.4.3 we mentioned arguments to show the impossibility of self-organization. It follows therefore that if an entity manifests any new organization it must have 'acquired' it at least in the sense of having generated it as a response to changes in its surrounding that have affected it.

3.2.1 Within the terms of this hypothesis we may suppose that A's initial structure, when A 'came into being' as an entity, must itself have been the result of order-extraction work carried out by some entity, antecedent to A, A' say. Similar observations apply to A' and to its antecedents, resulting in a backwards evolutionary chain, modified only by successive acquisitions of order, where these occur, of the entities that make it up. Corresponding to these different origins of A's structure, namely order-extraction work carried out by A (learning), or structure acquired by what, corresponding to the idea of an evolutionary process, we shall call 'heredity', we shall distinguish two separate forms of authorship. In so far as the purpose to produce W existed in A from the outset (by heredity), or if A acquired the purpose without proper order-extraction work (by being programmed), then in either case we shall refer to A as the apparent author of W. In so far as the purpose to produce W came into existence as the result of proper order-extraction work performed by A, we shall refer to A as W's creative author, written c-author. In general any particular W will be taken to be equivalent to the extent to which the purpose realized in W has originated in A, by A's order-extraction work.

Evolutionary nature of authorship

3.2.2 Returning to the question of purpose, we may restate the definition of authorship as follows: given a purposeful process AW, an observer O, able to observe events in E but not in A, will attribute W to the creative-authorship of A so far and only so far as he judges that (1) W corresponds to the execution of a purpose entertained by A, and (2) the purpose emanates from (has been generated as a result of acquisition by) A. We arrive at the definition of authorship by relaxing the second of these conditions.

Definition of author and creative author

3.2.3.0 A number of further definitions follow.

3.2.3.1 A productive process AW will be called a <u>creative process</u> (c-process) if and only if the controller A is a c-author, in which case we shall denote the process by c-AW.

c-process

3.2.3.2 If A denotes a number, n, of distinct controllers A_i of W (ie. if A is made up of a number of separate entities A_1, A_2, ... , A_n), then these will share the authorship of W and will be called the <u>partial authors</u> (or <u>particl c-authors</u> as appropriate) of W.

partial authors

3.2.3.3 Correspondingly we call W a c-W (or partial c-W) if and only if it is the product of a c-process (or partial c-process).

partial c-process

<u>SECTION 3 A, W AND THE ORDER-EXTRACTION PROCESS</u>

3.3.0 These definitions all depend upon the subjective measure of order-extraction work which has been made the basis of the definition of authorship. To show that this subjectivity is no mere artefact of the definition but is inherent in the character of what is being defined, we recall that A, A's purpose and W are all defined by and therefore dependent upon the characteristics of O. (see for example 1.2.2 and 2.5.2).

Authorship a subjective concept

> 'Because any real 'machine' has an infinity of variables, from which different observers (with different aims) may reasonably make an infinity of different selections, there must first be given an observer (or experimenter); a SYSTEM is then defined as <u>any set of variables</u> that he selects from those available on the real 'machine'. It is thus a list, nominated by the observer, and is quite different in nature from the real 'machine'.' (Ashby, 1965, p.16)

For 'machine' in this quotation we may read W. Our concern is not with 'the real' W but with what we may regard as an abstraction of the observer's.

3.3.1 Although the definitions proposed are subjective, they are embedded in effective definitions of order and will be shown to provide a foundation for extending those principles such as hopefully will furnish a basis upon which to build an art machine. We shall go about this by seeking to show that W plays a part in the process of A's order-extraction and is not merely the outcome of order-extraction work preceding its production. We shall begin by examining briefly various principles of order-extraction work. This is an extremely wide field covering practically the whole domain of artificial intelligence and including those studies usually denoted by the names 'problem-solving', 'pattern recognition', 'learning', 'evolution', as well as 'creativity' and 'imagination'. We shall touch only those parts of the field relevant to indicating some of the principal forms of order extraction. We shall follow this review with a more detailed examination of purposes and show how order extraction works to produce 'teleonomic' (2.0.0) entities and how these transmit purposes. Thereafter we shall deal with the realization of purposes: control and regulation, and the constraints exercized on order extraction by the growing structures of the teleonomic entities and their environments. These last two topics, constraint and control, will lead naturally to questions as to the kind of W that we should expect different kinds of teleonomic entity to produce.

W may form part of order-extraction process

CHAPTER 4

PRINCIPLES OF ORDER EXTRACTION

SECTION 0 ORDER AND ENTROPY

4.0.0 This chapter aims at drawing together the concept of heuristics with the wider concept of order to which it belongs. What emerges is a conservative view of order. Overshadowing the acquisition of order stands the Law of Requisity Variety, as a reminder of the work order extraction demands. Heuristics are slow and difficult to acquire and may lead to error. Matching the power that each new heuristic confers is a countervailing loss of flexibility. Without heuristics order extraction is slow; with them it may be rapid but ultimately disastrous.

A conservative approach

4.0.1 The chapter begins with a review of a number of definitions of the two concepts, algorithms and heuristics. Heuristics fill a role where algorithms are unavailable. Heuristics are aids to order extraction, but unlike algorithms, heuristics may not merely fail but actually lead to error which would not have occurred if the heuristic had not been used. Especially in novel situations, there may be a trade-off between error and heuristic power, which may be thought of as analagous to statistical type I and type II errors. The power that heuristics confer on an order-extraction process is often wrongly attributed to the process's authorship rather than to the authorship of the system to which the heuristics are due. Heuristics' power and error risk both relate to the 'distance' from the goal at which the heuristics are applied, power increasing inversely and error risk directly with increasing 'distance'. The chapter ends with a suggestion for measuring both heuristic power and error risk.

Speed and error

SECTION 1 HEURISTICS AND ALGORITHMS

4.1.0 The object of setting out to discover the principles for building an art machine is to arrive at an effective theory of art. We have shown that there is a strong relationship between art and order; that order comes into being through a process of order-extraction; and, by 2.3.1, that there is no effective procedure for extracting order. In these unhelpful circumstances we are therefore led to examine the principles of order extraction, especially the use of heuristics, with the object of discovering necessary features that an art machine will have to incorporate.

Principles of order extraction

4.1.1 Heuristic methods are concerned with order extraction. They are often contrasted with algorithms. Minsky (1965) makes 'algorithm' synonymous with effective procedure (p.105). Klir & Valach define the term as 'every precise instruction which uniquely determines the procedure leading from the initial information to the sought-for resulting information' (p.145). Beer uses algorithm to mean ' a comprehensive set of instructions for reaching a known goal' (p.305). Defining heuristic is more difficult. A number of authors do without the term. Ashby (1964) avoids both algorithm and heuristic, though his discussion of trial and error - what he prefers to call 'hunt and stick' - is essentially a treatment of heuristics. The '<u>objective properties of getting success by trial and error are shown when a Markovian machine moves to a state of equilibrium</u>' (p.230). Such movement is '<u>homologous</u> ... to movement by a determinate trajectory, for both are the movement of a machine to a state of equilibrium' (p.231). Other general texts that do not mention heuristics are those of Klir & Valach and Minsky (1972).

Heuristics and algorithms

4.1.2 Polya (1948) starts from the <u>Oxford English Dictionary</u> definition of heuristic as an adjective meaning 'serving to find out'. He associates the notion with an absence of 'complete certainty'. Heuristics are 'provisional', the 'plausible guess' that precedes attainment of the 'final goal'. 'We need heuristic reasoning when we construct a strict proof as we need scaffolding when we erect a building' (p.102). This contrasting of provisional and strict procedures appears to imply a difference of function. The provisional procedure somehow gets one to the goal while the strict procedure justifies one's having got there. These functions are distinguished by Reichenbach (1968) as the 'context of discovery' and the 'context of justification' (p.231).+ According to Reichenbach, an example of what may be thought of as occurring might be as follows. Suppose one's heuristic were an analogy (1.3.1). Suppose T were a theorem in a formal theory, an electrical network theory say, for concreteness; and suppose one were faced with a problem in hydrodynamics. Then, if one noted an analogy between the network to which T applied and one's particular problem in hydrodynamics, it might be possible to use this heuristic (the analogy) as the basis for a 'plausible guess' at the solution to the problem in hydrodynamics. One might set up the hypothesis that the analogy one had noted between the two systems might be extrapolated to include the result T (in its appropriate hydrodynamic form, call it T_{hydro}), thus suggesting a solution for the hydrodynamics problem. T however would not provide a 'strict proof' of T_{hydro}. It would stand as a 'conclusion' that might then act as a statement of what was required to be proved, in a form that might itself indicate lines along which the proof could possibly be expected to run. That is the analogy might act as a kind of goal-setting or problem formula-

The working of heuristics

+ Reichenbach comments, 'the act of discovery escapes logical analysis; there are no logical rules in which a "discovery machine" could be constructed that would take over the creative function of the genius' (p.231). Five years later Newell & Simon (1956) reported on a Logic Theory Machine that proved theorems in Russell & Whitehead's Principia Mathematica. Chapter 6 will provide detailed examples of the use of heuristics in related machine enterprizes.

tion device. This argument and Polya's are clearly close to one another.[+] (cf. also Newell & others (1958) fourth criterion for creativity, N.3.2)

4.1.3 Newell & others (1958) denote by heuristic 'any principle or device that contributes to the reduction in the average search to solution' (p.22). The shift of emphasis away from questions of the validity of heuristics, towards those touching their operation, reflects the authors' concern with the realization of effective procedures in the form of computer programs for actually solving problems. An heuristic for such a program is quite specifically a means of reducing the number of branches to be searched by some tree-searching procedure for instance. The hope would be to eliminate only those branches <u>not</u> leading to solutions, while leaving for search those others along which solutions might be found. Newell & others (1958) point out that this hope is not necessarily always possible to fulfil. For heuristics may eliminate fruitful as well as barren branches.

Heuristics and machine operation

SECTION 2 GOING OUT ON A LIMB

4.2.0 It appears from this that 'error' may arise from two sources due to the use of heuristics, one which leads to the neglect of aspects of a problem where solutions are to be found, and another, which allows the retention for investigation of aspects where no solutions exist. This suggests a measure for the efficacy of heuristics, analagous to the familar statistical procedure for distinguishing <u>type I errors</u> (barren branches left) from <u>type II errors</u> (pruning fruitful branches). That is we picture two

Heuristics as sources of error

[+] Selfridge, in the discussion following a paper presented by McCarthy (1958), extends this point of view: 'he [Professor Bar-Hillel] made a remark that conclusions should not be drawn from false premises. In my experience those are the only conclusions that have ever been drawn' (p.86). This sums up the basis of the idea that Popper (1963) advances, namely that scientific knowledge grows by means of a series of conjectures and refutations. Popper contends that a consequence of Hume's conclusions concerning causation is to preclude any other view. This argument receives closer examination in Part II.

populations of branches, barren and fruitful. A priori there is no way of knowing to which population a given branch belongs. Heuristics may be regarded as tests for classifying branches. The better the heuristic the fewer 'wrong' classifications it will lead to. In the absence of an error-free heuristic, knowledge of the relative magnitudes of the type I and type II errors will clearly help in choosing the 'best' among a number of heuristics that may be available for a given task. Indeed such a choice might itself be heuristically systematized in order to optimize the balance of the two kinds of error in such a way as to maximize problem solutions. Evidently then, a feature of heuristics, as distinct from algorithms, is that they give no assurance of reaching a particular goal. Indeed they may actually block progress towards a goal that would have been reached if another heuristic had been used or search had proceeded without the aid of heuristics, 'randomly'.+ Minsky (1958) makes the point explicitly. We

> ' ... use the term heuristic in describing rules or principles which have not been shown to be universally correct but which often seem to be of help, even if they may also often fail. The term "heuristic program" is thus used in reference to the distinction between programs which are guaranteed to work (and are called "algorithms") and programs which are associated with what the programmer feels are good reasons to expect some success' (p.36).

SECTION 3 DISTINGUISHING ALGORITHMS AND HEURISTICS

4.3.0 Beer appears to make a somewhat stronger claim for heuristics. Like Minsky, he contrasts heuristics and algorithms. A heuristic he defines

+ Problems devised by early Gestalt psychologists, such as those for instance that were supposed to demonstrate 'insight' may be interpreted in these terms. Thus typical insight problems are of a kind that can only be solved when the subject recognizes that he is using a wrong heuristic. The problems devized are such as would be likely to call forth some habitually used heuristic that would not lead to a solution. The subject would show 'insight' when he abandoned his wrong heuristic in favour of a better one. We shall show that this important feature of creativity corresponds to one of the criteria of creative thinking suggested by Newell & others (1958) (see N.3.2). From this viewpoint, random procedures are likely to prove most creative, since they will be likely to achieve the most novel and unconventional solutions to problems, along with a high degree of wastage, (just as random choices in the football pools are likely to win the biggest prizes).

as a 'set of instructions searching out an unknown goal by exploration, which continuously or repeatedly evaluates progress according to some known criterion' (p.306). Or again, 'An heuristic specifies a method of behaving which will tend towards a goal which cannot be precisely specified because we know what it is but not where it is' (p.69). The 'basic trick' is to provide an 'algorithm determing an heurstic' (p.71). So we end up with an algorithm. It is not immediately clear what the relationship is between the 'known goal' that Beer refers to in his definition of algorithm (4.1.0) and the 'known criterion' associated with the heuristic. Nor is it obvious how the algorithm's 'known goal' differs from the heuristic's 'unknown goal'. The difficulty may become clearer by means of an example: a multiplication algorithm computes a product. Such an algorithm might be embodied in some kind of mechanical device, such as a Turing machine. Given the appropriate starting tape, including on it the multiplier and the multiplicand, the machine would perform the desired computation. It would execute its procedures in a determinate way and to that extent its starting tape would determine its end tape (cf. N.2.4). But beyond this, it could not be said to 'know' its goal. The machine's programmer (the observer) might know (be able to state in his meta-language) that the machine's goal was to compute the product of two numbers printed on its starting tape; that is, might know what the machine 'aimed' to do, namely carry out a certain computation. But in general the programmer would not know what the product was going to be, any more than the machine would, unless he had previously performed the computation himself, using an algorithm isomorphic with the one given to the machine. The observation that, in the case of an heuristic goal, 'we know what it is but not where it is', is not helpful in the present example. For the Turing machine will execute the multiplication algorithm without knowing its goals in any usual sense. Whereas the programmer who knows the goal (that is, what the goal is, namely to perform a multiplication with given multiplier and

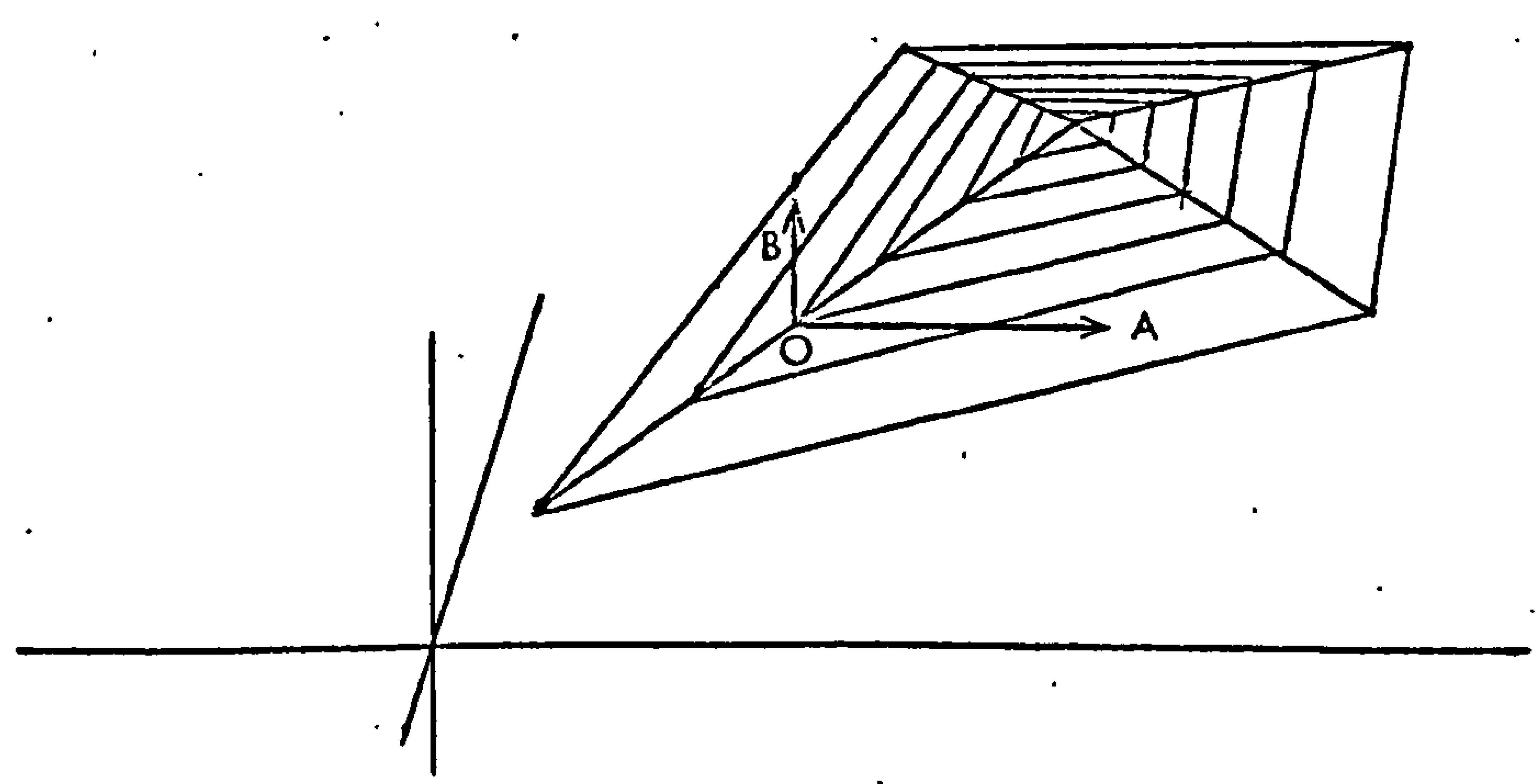
Stronger claim for heuristics

multiplicand), will not necessarily know whether or not the machine has attained it, if he does no more than inspect the answer it reaches.

4.3.1 Hill-climbing heuristics

What then are the characteristics of heuristics that distinguish them from algorithms in Beer's terms? Let us look at an example that Beer provides: to climb a conical peak in a mist you could use the heuristic 'keep going up'. As long as a peak existed and was conical, such an instruction would get one to it.+ Given a compass and a tape measure, one could plot one's path as one went, so that it would subsequently be possible to specify a path leading from one's particular starting point to the top, as an algorithm of the form: 'Go a distance d_1 in a direction s_1; change direction to s_2 and travel a distance d_2', &c. What then are the differences between the algorithm and the heuristic? The algorithm is sure while the heuristic is not: given a map of the area not showing the peak, the algorithm would

+ Though even this instruction is not as straight forward as it looks. For it immediately gives rise to the question how to carry it out. An example of O. Selfridge, cited by Minsky & Papert(1969) provides an instance of a 'good hill [conical] with a bad algorithm' (p.179) for executing the heuristic that Beer suggests.



A and B are lower than O

(Sketch reproduced from Minsky & Papert, 1969, p.179)

The sharper the ridge, the more likely is it that a search of 'local' surrounding points by someone standing on it will fail to discover any point higher than the one he is on. In general the problem is even more complicated, since hills are not normally conical. 'Success depends ... on the extent that the summit is not as globally defined as it might appear. In cases where the hill has a complex form, with many local relative peaks, ridges, etc., hill-climbing procedures are not always advantageous. Indeed, in extreme cases, a random or systematic search might be better than a procedure that relentlessly climbs every little hillock' (ibid. p.178).

enable one to plot the position of the peak, provided one could find the starting point to which the particular algorithm applied; not so the heuristic. To do the same by means of an heuristic would require a contour map of the mountain which would of course show the peak. The algorithm is a set of instructions for getting from some particular point on the mountainside to the summit; the heuristic is independent (in principle, if not in practice) of the starting point. Without either algorithm or heuristic though, one could still probably get to the top of the mountain by going on a series of 'random walks'. (The number might be extremely large, but might arbitrarily be made finite by treating possible starting points, changes in direction, and so on, as discrete and finite in number, by the superimposition of some kind of grid say.) But even on a random walk one would need to have a way of recognizing the top when one reached it; an altimeter say, and a rule for interpreting its readings. Such a rule however would be bound to overlap with the heuristic 'keep going up'. For the heuristic would have to include an instruction such as, 'Go to the "local", highest point; if it is the point you are at, stop. You are at the top.' And the rule for interpreting the altimeter would have to be of a form such as, 'Is the altimeter reading higher than it was before you made your last move? If not, go back to the point of your last move and stop. You are at the top.' Thus, given an algorithm, one <u>need not know the goal at all</u>, (cf.430) since blindly following the algorithm to the instruction 'stop' will suffice to get one to it. While on the other hand, lacking an algorithm, one must have at least a definition of the goal sufficient to allow recognition of it. When Beer associates algorithms with known goals, the goal is only known in the meta-language of the algorithm's <u>author</u>. The known criterion (4.3.0) that he associates with the heuristic refers to the object-language of the heuristic's operator.

4.3.2 If one plotted in three dimensions a large number of random walks, some reaching the summit and some not, one might note that a common feature of paths that reached the peak was that they all kept going up. Clearly it is by noting this common feature that one arrives at Beer's heuristic. But the business of noting it is precisely what we have been calling order extraction. So the heuristic does not avoid order-extraction work. It represents its results.

Heuristics from algorithms

4.3.3 It appears therefore that including heuristics in an algorithm may assist in order-extraction work (such as is performed by the General Problem Solver of Newell, Shaw & Simon (1963) for instance) by providing the algorithm with the benefits of the previous order-extraction work that has resulted in the discovery of the heuristics that it incorporates. And this furnishes the basis for a kind of heirarchical procedure by which, roughly speaking, the algorithm is enabled to proceed by ever larger steps. We may interpret this view in the terms used in the previous chapter, discussing order, to provide a stricter statement. (The present interpretation corresponds to the conclusion of 2.8) Accoring to this formulation we should say that an heuristic, in so far as it represented the results of order extraction, might be regarded as a description that was relatively shorter than what it described. In passing from the starting point to the goal therefore, the algorithm that incorporated the heuristic would need to pass through relatively fewer steps than would have been required in the heuristic's absence. This suggests lines along which one might attempt an effective definition of heuristics which would make them analogous to the partial explanations and partial descriptions of 2.7.

Heuristics increase order-extraction power

4.3.4 Solomonoff (1964) provides a formulation of inductive inference that suggests a basis for such a theoretical measure of heuristic 'power'. Solomonoff defines a 'probability' in the following way. Suppose we have a universal Turing machine M with a starting tape S that makes M generate T, written M(S) = T, then we say that 'S is a description of T with respect of M' (p.9). The probability that Solomonoff defines takes into account all the possible outputs of a universal Turing machine and all possible descriptions of each output, with the object of constructing a function for computing the 'probability' P (a, T, M_1) that might be called ' "the probability with respect to M_1, that [the string] a will follow [the string] T " ' (p.8). If we regard an heuristic as a description, which we have argued is plausible, then $P(a, T, M_1)$ would be reduced for any machine using the heuristic, since the heuristic would limit the numbers of descriptions of strings that the machine could output. The relative extent of this reduction might provide a basis for the measurement of the heuristic's 'power'.

Measurement of heuristic power

SECTION 4 MINIMAL HEURISTICS AND ALGEDONIC CONTROL

4.4.0 Notice that, in any heuristic process, the heuristics used may derive from order-extraction work that has occurred outside the process, like Beer's hill-climbing rule for instance. Or the process may provide its own heuristics by means of its own order-extraction work. For example, order extraction occurring within a problem-solving procedure might produce partial solutions in the form of heuristics. These might then be incorporated within the procedure (possibly in a way laid down in advance) to modify and strengthen it. The distinction between these two possible origins of heuristics is not always strictly observed. Sometimes it appears

Authors and users of heuristics

tacitly to be assumed that a procedure may as easily be detailed for developing its own heuristics as furnished with them from the outset; even though, in the latter case, its heuristics may reflect the order-extraction work carried out by very much more powerful systems, such as the human programmers of computers, for example. The failure to maintain a distinction between the two possible origins of heuristics occasionally leads to confusion as to the source of the power that heuristic procedures demonstrate, what is due to the system that has produced the heuristic being credited to the program that employs it. Such confusion fails to recognize that the problem of discovering heuristics is precisely the problem of extracting order. Many procedures that employ heuristics aim at order extraction (as for example all problem-solving procedures[+]). Finding heuristics to find heuristics evidently is one way of describing the study of order extraction!

4.4.1 Once provided with appropriate descriptions, structures and other suitable heuristic devices, their systematic use, by effective procedures into which they are incorporated, may greatly increase such procedures' power. But order extraction itself, as we shall see in the following chapter, is subject to severe restrictions. Thus Beer's 'algedonic' loops[++] make powerful control elements, though one should be careful to notice that their power may often conceal the extent of the order-extraction work that has been needed to produce them. For example, it is misleading to suggest that the dog trainer, who trains his dog by means of rewards and punishments 'does not understand "how the dog works" ' (ibid. p.78). Simply understanding what constitute rewards and punishments for

Hidden order-extraction work

[+] These may be taken to include all situations as defined by Newell & others (1958) in which 'Given a set, P, of elements, the problem is to find a member of a subset, S, of P having specified properties' (p.11); a wide class indeed.

[++] Procedures that 'reward' right moves and 'punish' wrong ones.

the dog implies 'understanding' of an important kind, not necessarily easy to come by. It is easy to overlook how much is (correctly) assumed about what gives dogs pleasure and pain, by simple extension from self-knowledge. How much needed to be learned would be more obviously apparent if it were required to discover rewards and punishments for a complex system of quite unknown properties. Even among humans, whose properties are not quite unknown, the discovery of algedonic principles is not necessarily always simple. 'Masochism', for example, has no immediately obvious explanation. A 'random' search procedure for algedonic loops in a system of comparable variety might easily demand computational power exceeding the limits calculated by Bremermann.+

4.4.2 It follows from this argument that the power of algedonic systems may be less than it at first appears. When outputs emitted by one system alter the organization of other systems, without the first system's knowing what might act on them as rewards and punishments, by 1.1.1 the name 'controller' does not apply to the first system, because the first system is unmotivated (being unable to distinguish the states of the system whose organization it is affecting). Only where one system demands and can distinguish in another changes of a particular kind or confined to a particular class, does the name controller properly apply. And this requires that the first system should be able to control and hence communicate with the systems in which it wishes (entertains the purpose) to bring the changes about. Analagously with our observations about the goals of algorithms and heuristics (4.3.1), we notice that any outside agent that knows the control effects of some given system on another, may use the first system to control the second, even though the agent has no idea how or why the control

Meta-controllers

+ Bremermann (1962) used quantal considerations to calculate an upper limit to the rate at which computation may be performed. The rate works out at 10^{42} bits per gram mass of computer per second.

procedure works. But he will at least have to know what it achieves, a state of affairs S say. Knowing this will be the same as knowing an heuristic for achieving S. The outside agent in this situation has the role of a meta-controller.

SECTION 5 THE CONTROL LANGUAGE

4.5.0 'Pleasure' and 'pain' may be taken to be the two words of a necessary and sufficient (binary) language in which to communicate minimal control commands. To learn (extract order), any search procedure must possess a way of 'knowing' when it has succeeded: it requires an answer, 'Yes' or 'No', to the question, 'Have I reached the goal?' Without such knowledge learning is impossible since the procedure will never halt. To bring about, in the organization of two systems, changes such as to produce correspondences (or correlations - 2.3.2) between them - that is, to permit order extraction to occur - the two systems must be able to communicate. The minimal language that permits this is a mutually interpretable binary language. Ashby's 'good' and 'bad' forms of organization (2.4.3) are those forms respectively that the algedonic controller rewards and punishes. Both pairs of terms, pleasure and pain, good and bad, refer to the organization of one system altering in a way that tends to favour its survival in conditions of disturbance brought about by another system in relation to which the terms are defined.

Binary control

4.5.1 Controls based on algedonic principles are error-free. This is because algedonic control refers only to the goal of the procedure with which it is concerned; (that is, the two words of the language refer only to whether or not the goal has been attained) and not to any earlier phases of the

Algedonic heuristics are error-free

movement towards it. For this reason, heuristics of this kind, if they may properly be called heuristics, cannot block possible goal paths, as they might if they referred to choice points along the route to the goal (4.1.3).

4.5.2 We see from this that the magnitude of the error that an heuristic may produce must be related to the phase of the goal-search procedure to which it refers. In a tree search procedure for example a tree might have k levels of branching and m branches at each branch point. Such a tree would have m^k branch ends. If only one of these ends represented the goal, then the probability of reaching it by chance (random search) would be $\frac{1}{m}k$, assuming equi-probability of branch choice. An heuristic that eliminated $\frac{m}{2}$ branches at the first branching would double this probability. But it would <u>increase</u> the size of the maximum possible type II error (4.2.0) by the same proportion. An heuristic that eliminated $\frac{m}{2}$ branches at the second branching would eliminate only $\frac{m^{k-1}}{2}$ paths overall. Such an heuristic - assuming it worked - would not so greatly increase the probability of reaching the goal as the first heuristic, but neither would it so greatly increase the risk of type II errors, because it would be eliminating far fewer paths. There is a trade-off between the efficiency of the heuristics - in speeding up goal attainment - and risk of error. In the limiting - algedonic - case no paths are eliminated, since in this case there are no further choices of paths to be made.

Error-risk-power trade-off

4.5.3 Notice that in this algedonic case, the procedure accords with none of the cited definitions of heuristics. It differs in two respects from the notion described: (1) it is error-free; (2) it does not accelerate goal

Error-risk increases with control vocabulary

attainment. The operation of the algedonic procedure may possibly be better understood from another point of view. One might regard the kind of 'heuristic' of this limiting case as a sufficient though possibly not a necessary condition for any kind of instruction. In this case we might let the two words of the binary language be 'correct' and 'incorrect'. The schoolmaster who set his pupils problems and then restricted his teaching to marking ticks and crosses against their efforts to solve them might find that the pupils learned little. Yet undeniably the dog-trainer, who is unable to give his dogs any more positive instructions than the hypothetical schoolmaster, does get them to learn. He may increase his successes by dividing the dogs' tasks into numbers of phases; that is by sub-dividing the tasks and carefully watching the dogs' behaviour, a technique quite familiar from Skinnerian reinforcement schedules. In other words, he resorts to a repertoire of understanding that he has achieved and that now represents his own learning (acquired through his own order-extraction work). Moreover, though the algedonic language may be binary, it is not necessarily so. The dog-trainer is likely to employ rewards and punishments of different stengths; the schoolmaster to award answers marks instead of simply ticks and crosses. But notice that in general any increase in the vocabulary available for the algedonic language will have to be bought at the price of (1) an equal increase in the number of final states its user is able to distinguish in the system he wishes to control with it, and (2) his gaining sufficient understanding of that system to enable him to describe those end states to it, in a way that it will be able to interpret correctly. Thus the effect of the controller's procedure for shortening the search is once again seen to (1) depend upon how much order their is in the heuristic's power, and (2) increase the possibility of error (by risking, in this case, wrong identification of some intermediate task state, or misinterpretation by the controlled system of the controlling system's language).+

+ A corollary to the conclusions of this paragraph suggests a basis for the resolution of differences between Skinnerian and opposed psychologists. It is evident that though the Skinnerian framework is a sufficient one for learning, it does not follow that it is all there is. If psychologists' endeavours had been restricted to homeostatic search it is doubtful whether Skinnerian psychology would exist at all.

SECTION 6 HEURISTICS AND HOMEOSTATIC PRINCIPLES

4.6.0

Heuristics aid not replace order extraction

> 'We should say what is a better and what a worse result, but the computer has to determine a better strategy, a better control system than we ourselves know. And of course it can do it. Because its algorithm, what it is programmed to do, specifies an heuristic. Alter the solution you are now using a little bit, says the algorithm, and compare the outcome with the erstwhile outcome. If this is more profitable, or cheaper, or whatever else we say, adopt it. Go on like this until any variation you make leads to a worse result than you already have. Then hang on to this solution, until the situation changes: whereupon you may do better once again by producing a new variation' (Beer, p.71).

We have already drawn attention to the observations of Minsky & Papert (4.3.1) concerning the critical importance of local and global features in hill-climbing procedures, such as the one Beer describes above, and have commented (4.4.0) on the dangers of simply passing the buck of order extraction to a computer, in the name of suppressing variety. In the end, the order extraction required for effective modelling, description, explanation, theory development, &c., is inescapable, and the simple 'power' of computers is in itself a poor substitute for structure. In general we might regard purely homeostatic principles as laying down the minimal conditions necessary for two systems to become correlated, in Ashby's sense (2.3.2). But it is necessary to recognize that these principles are no more than minimal. The following chapter will show more clearly the way in which heuristics add to powers of order extraction.

4.6.1

Heuristics easier to use than to find

In practice the use of heuristics in 'heuristic programming' and the like appears to have been confined principally to the use of heuristics provided 'from without', by some (human) agent, outside the procedure that uses the heuristics. In general it seems that, while it is possible to modify and improve such heuristics by the operation of the program or procedure that uses them, their innovation is more difficult.

> 'One of the key heuristics that underlies physical intuition in dynamics is the notion that forces produce changes in velocity (rather than producing velocities). Evidence from which this idea might be derived is

available to anyone with eyes. Yet at least hundreds of man-years of search by highly intelligent men were required to discover the idea, and even after it was enunciated by Galileo, another century of work was required before even the most intelligent scientists had cleared it of all obscurity and confusion' (Newell, Shaw & Simon, 1958,p.73).

4.6.2 To show how homeostatic principles may alone be sufficient to extract order gives little idea of how heuristics operate to enhance the power of order-extraction procedures. In the homeostatic case, the actual business of extracting order is left to 'self-organization', with little attention paid to the formation of the heuristics that the homeostatic procedure may use. Homeostatic principles alone do not furnish a full account of how order extraction occurs. It is essential to show fully both how heuristics help and how they come into being in the first place. For, as we have shown, finding heuristics and using them are part of the same broad order-extraction process. Minsky & Papert (1969) sum up some of the shortcomings of enquiries that fail to penetrate deeply enough into notions of 'self-organization' and the like.

Complexity of heuristics

'A perusal of any typical collection of papers on "self-organizing" systems will provide a generous sample of discussions of "learning" and "adaptive" machines that lack even the degree of rigor and formal definition to be found in the literature on perceptrons. The proponents of these schemes seldom provide any analysis of the range of behaviour which can be learned nor do they show much awareness of the price usually paid to make some kinds of learning easy.' (p.16)

CHAPTER 5

LIMITS OF ORDER EXTRACTION

SECTION 0 FURTHER QUESTIONS OF STRUCTURE

5.0.0 This chapter takes up in more detail some of the general conclusions reached so far, with the object of arriving at a deeper understanding of where and in what form we shall encounter restrictions on the kind of art machine that might be constructed. We have suggested, in the previous chapter, that the task is not hopeless, but that we shall have to look for success of a more limited kind than we might have hoped for at the outset. Here we examine this idea with a view to discovering where we should expect the limits to fall; what kinds of questions we should ask about the machine's capabilities; what in general will be the relative demands for time, power and structure and what are the prospects for meeting them. Above all we persist with questions as to the sources of structure. The answers to these questions concern the scope and power to be creative that we may reasonably expect of any art machine we might be able to construct.

Limits of creativity

5.0.1 The chapter begins by classifying various forms of order extraction, according to the time scales within which they proceed, and the degree of definition of their goals. The classification is presented as a table of kinds of order extraction, from pattern recognition to biological evolution. Technical differences are discussed between parallel and serial processes before an examination of perceptrons as paradigms for parallel processes. Some of the conclusions of the investigations of Minsky & Papert (1969) are contrasted with earlier more optimistic forecasts for

Features of teleonomic systems

perceptrons. Conclusions are generalized to the wider field of order extraction, with reference to the Law of Requisite Variety. It is conjectured that order-extraction rate grows exponentially with order extracted and that the kinds of order extracted will largely determine the kinds of order that will be extracted. Finally it is suggested that the transmission of order leads to the development of teleonomic systems.

SECTION 1 THE CLASSIFICATION OF ORDER-EXTRACTION

5.1.0 Pattern recognition, learning, adaptation, problem solving, thinking, creativity, development and evolution are among numbers of headings under which order extraction of various kinds has been studied. Computer science and cybernetics both contribute an extensive literature in experiment and theory. Our particular concern is with the principles of order extraction, what kinds of order extraction may be distinguished, the minimum conditions that permit it, what facilitates it, the rates at which it occurs, what forms it may take. Broadly speaking cybernetics has studied order extraction largely under three heads; as a perceptual problem, for the most part under the name 'pattern recognition'; as problem solving, which has included thinking, creativity and the like; and as evolution. Learning and adaptation have been treated, both as aspects of all the former studies, as well as in their own right. Though learning entails order-extraction work, order extraction does not necessarily produce learning. A procedure for example that solves a problem or recognizes a pattern, though it may be said to have done order-extraction work (in the sense of 3.2.0), need not be in any way modified by the result it achieves. In the absence of some degree of retention, even temporary of the results of the procedure, we should not call its performance learning.

Categories of order extraction

5.1.1 No sharp divisions delineate the groupings mentioned. They differ more as to the aspects of order extraction they emphasize than in points of principle. Broadly, in practice, much of the difference between the various approaches is associated with the time. Thus in general, preception, or at any rate recognition, occurs over short time intervals, and accordingly it is chiefly in the area of pattern recognition that relatively rapid parallel, as distinct from generally slower sequential, procedures have been investigated. Problem solving (chess, theorem proving and the like) typically takes place over longer time intervals, which seem more 'naturally' to call for sequential searches aided by heuristics. Pattern recognition and problem solving may both occur without learning. Evolution on the other hand requires at least some kind of homeostatic adaptation (which, at its simplest, may not warrant the name learning - see 5.2.2). Evolution is normally associated with long durations and may enjoy little heuristic guidance initially at any rate, beyond what the most primitive of binary controls, survival and non-survival, provides.

Time scales

5.1.2 Kinds of goals and their definition

With these differences have become associated corresponding differences in the kinds or order with which the various approaches are concerned. These relate to goal definition and the kinds of heuristics used. Pattern recognition procedures tend to specify goals most fully. Uhr (1966) refers to pattern recognition as the 'problem of assigning a name to, or classifying the many different exemplars of a particular class' (p.2). In practice this has usually meant naming correctly various shapes or handwritten alpha-numeric characters say, projected on some sort of 'retina'. The goal's specificity - to name the character or shape - arises partly as an artefact of the experimental situation. It is necessary for the experimenter, in this case the meta-observer, to be able to say whether or not the procedure he is using - which is the observer - is making what he,

pattern recognition

the meta-observer, decides is a correct classification of some given pattern. This means that the experimenter specifies in advance what the procedure (observer) is to recognize. (If the procedure identified a pattern that the experimenter could not identify, this would necessarily fall outside the domain of the experiment and receive no credit; that is the experimenter would put it down to mal-function of the procedure.) In problem-solving procedures the goals are generally less specific than in pattern recognition. While the pattern recognition experimenter provides the various 'exemplars of a particular class', the experimenter in problem solving merely provides a class description according to which exemplars may be identified - criteria for logical proof provided to a theorem-proving programme for instance - allowing the procedure to produce the exemplars themselves - in the form of particular proofs of a theorem, say. Evolutionary procedures are unique in requiring no goal at all as a necessary condition of their operation. Survival or non-survival will alone determine what order such procedures extract. Simple 'self-connecting' (2.4.3) is a sufficient operating principle for the procedure. Survival need not even be a goal. It will emerge as one, because survivors are likely to evolve procedures for retaining and protecting survival mechanisms just as they evolve the mechanisms themselves. While goals, in this sense, are not necessary, evolutionary procedures do not preclude specific goals, as for example in the procedures of Fogel & others (2.6.4). We present the differences that we have outlined in this section as a table, which is offered simply as a general indicator of differences whose numerous, readily-apparent inconsistencies show the inadequacies of the arbitrary categories into which the study of order extraction has fallen.

problem solving

Evolution

	Kind of Order	Goal Criterion	Model	Heuristics
1.	Patterns to be classified: patterns given by experimenter	Correct or incorrect classification stipulated by experimenter	Pattern recognition	'Natural' shape recognizers: Lettvin, Maturana, McCulloch & Pitts (1965) Sutherland & Mackintosh (1971) Guzman, SEE programme Rattner, SEEMORE programme (chapter 6)
2.	Problem solution: solution pre-known to experimenter	Criterion for solution stipulated by experimenter	Problem solving	Tree-pruning rules (non-adaptive) Tree-pruning rules with adaptation based on feedback and hindsight analysis (adaptive).
3.	Evolution to pre-scribed criterion, in specified environment	Acceptable 'fit' to prescribe form	Development, evolution	Knowledge of results in n-ary language: Fogel & others (2.6.4)
4.	Classification of 'new' patterns	Consistency of classification as judged by experimentor	Pattern recognition	
5.	Solution of unsolved problems	'Fit' to prescribed criteria, e.g. criteria for proof	Problem solving	'New' proofs of theorems: Newell & Simon (1956)
6.	Evolution to prescribed criteria in 'open' environment	As for 3	Evolution	'Structured' changes: Pask (1959,1969) Theory of Evolution; Löfgren (1972) Breeding experiments, &c.
7.	'Free' evolution	Survival	Biological, social theory, &c., evolution	Biological structures, social structures, &c.
8.	'Free' evolution in 'open' environment	No goal	Connectivity	None

TABLE: 5.1.2

SECTION 2 PARALLEL AND SERIAL PROCESSES

5.2.0 Order extraction's formal treatment leaves many of its practical aspects concealed, though an important difference that it examines explicitly is that between serial and parallel processes. Parallel processes gather in-

Time-cost trade-off

formation on a wide class of events, taking the various separate 'pieces' of information independently. Order extraction that occurs - a decision taken, based on the 'evidence' of the gathered information say - may depend upon only a tiny fraction of the total information that has been assembled. That is only a relatively extremely small part of the total information gathered may be used to determine the order extracted (the decision taken). The information not used for the decision is 'wasted'. The cost of gathering this wasted information is the price paid for being able to acquire order (reach some required decision) in a single step by the simultaneous assessment of the information gathered, rather than by the series of steps needed by a sequential procedure.

> 'The total amount of computation may become vastly greater than that which would have to be carried out in a well-planned sequential process ... whose decisions about what next to compute are conditional on the outcome of earlier computation. Thus the choice between parallel and serial methods in any particular situation must be based on balancing the increased value of reducing the (total elapsed) time against the cost of the additional computation involved.' (Minsky & Papert, 1969, p.17)

5.2.1 Perceptron is the name given to a class of abstract machines that use parallel procedures. They were first conceived as devices for pattern recognition. 'Perceptrons make decisions - determine whether or not an event fits

Perceptrons

a certain "pattern" - by adding up evidence obtained from many small experiments.' (ibid. p.4) In an early paper on perceptrons, Rosenblatt (1959) makes it clear that the hope for these machines was 'that a properly designed perceptron will be capable of spontaneously forming meaningful classifications of the stimuli in its universe without being taught by an experimenter' (p.421) (Italics added). That is perceptrons were conceived as promising to capture the properties of order, proceeding from a <u>tabula rasa</u> (cf. 4.6.1 for example). Thus, although in practice the order that the perceptron is required to extract has been pre-determined by the experimenter, real order-extraction work is nevertheless required of the perceptron, which itself must discover a pattern that it has not previously identified.

5.2.2 Minsky & Papert offer the following paradigm and definition for a perceptron.

Paradigm for a perceptron

'Let $\Phi = \varphi_1, \varphi_2, \ldots, \varphi_n$ be a family of predicates. We will say that
<u>Ψ is linear with respect to Φ</u>
if there exists a number θ and a set of numbers $\{\alpha_{\varphi_1}, \alpha_{\varphi_2}, \ldots, \alpha_{\varphi_n}\}$
such the $\Psi(X) = 1$ if and only if $\alpha_{\varphi_1}\varphi_1(X) + \ldots + \alpha_{\varphi_n}\varphi_n(X) > \theta$
The number θ is called the threshold and the αs are called the coefficients or <u>weights</u>. ... We usually write more compactly $\Psi(X) = 1$ if and only if $\sum_{\varphi\epsilon\Phi}\alpha_\varphi\varphi(X) > \theta$' (p.10)

'<u>DEFINITION</u>: A PERCEPTRON is a device capable of computing all predicates which are linear in some given set$^\Phi$ of partial predicates.' (p.12)

It was hoped, at the outset of investigation of perceptrons, that appropriate alteration of the weights, achieved by feedback from the truth or falsity of computed predicates might provide the basis for a sufficient form of adaptation, to enable such devices to come to distinguish 'patterns'. Yet the promise seen in the first perceptron separability theorems remains unfulfilled. Minsky & Papert write that 'There is not yet any general theory of this topic' (p.181). In particular, 'the problem of relating speeds of learning of perceptrons and other devices has been almost entirely neglected' (p.181). They argue that perceptron learning should entail a better procedure than simply the ability to find a separating predicate. A simple homeostat might achieve this with a finite number of errors. But 'it would be hard to justify the term "learning" for a machine that so relentlessly ignores its experience' (p.181).

SECTION 3 THE COSTS OF ORDER EXTRACTION

5.3.0 The broad conclusions reached by Minsky & Papert apply not only to perceptrons or pattern recognition, but to all systems capable of this kind of formal treatment. Ashby (1962) looked forward ten years earlier to the kind of treatment Minsky & Papert undertake. He pointed out that '<u>there is no getting selection for nothing</u>'. 'I suggest that when the full implications of Shannon's Tenth Theorem [Ashby's <u>Law of Requisite Variety</u>, (1964)] are grasped we shall be first sobered, and then helped, for we shall then be able to focus our activities on the problems that are properly realistic, and actually solvable' (p.273).

Shannon's Tenth Theorem

5.3.1 The observations of Minsky & Papert on 'The Seductive Aspects of Perceptrons' aim in this regard 'to separate reality from wishful thinking':

> ' ... it is easy to imagine a kind of automatic programming which people have been tempted to call <u>learning</u>: by attaching feedback devices to the parameter controls they propose to "program" the machine by providing it with a sequence of input patterns and an "error signal" which will cause the coefficients to change in the right direction when the machine makes an inappropriate decision. The <u>perceptron convergence theorems</u> ... define conditions under which this procedure is guaranteed to find, eventually, a correct set of values.' (p.14)

Perceptron repertoire

But we may not simply leave the matter there. These authors present the following formulation: if Φ is a set of partial predicates and $L(\Phi)$ is the set of all predicates linear in Φ, then $L(\Phi)$ is the repertoire of the perceptron, the set of predicates it will compute when its coefficients α_ϕ and its threshold θ range over all possible values. In practice $L(\Phi)$ is limited by the size of Φ possible for some actual device, and this puts a limit on the repertoire of any (physically real) perceptron. 'The ease and uniformity of programming have been bought at a cost! We contend that the traditional investigators of perceptrons did not realistically measure this cost.' (p.15)

5.3.2 The authors list three 'crucial points' neglected by investigators:

Kinds of information and rates of learning

(1) As a paradigm, $L(\Phi)$ tends to lose the 'geometric individuality of the patterns', because it leads to the treatment of classes of geometrical objects as classes of n-dimensional vectors $(\alpha_1, \ldots, \alpha_n)$. There are, they show, '<u>particular</u> meaningful and intuitively simple predicates that belong to <u>no</u> partically realizable set' (p.15). '$\Psi_{CONNECTED}$' is an example of a predicate that cannot be computed by a wide class of perceptrons. ($\Psi_{CONNECTED}$ is the name the authors give to a function that computes 'connectedness' - the property of being able to get from one part of a 'thing' to another by a path that lies wholly within the thing.) Yet some analysis of connectedness is clearly important, if not actually a necessary condition, for defining 'things' in physical space. (cf. Ashby, 2.2.1) As a more concrete example ' ... we can construct, by special methods, a perceptron that could learn either to recognize squares or to recognize circles. But the same machine would probably not be able to learn the class of "circles or squares"! It certainly could not describe (hence learn to recognize) a relational compound like "a circle inside a square".' (Minsky & Papert, 1972, paragraph 4.3).

(2) The authors find that the question of the information content of the parameters $\phi_1, \ldots, \phi_n$ has been neglected. There are examples, 'which we believe are typical rather than exceptional' (p.15), of 'meaninglessly big' ratios of the smallest to the biggest of these coefficients. 'In some cases the information capacity needed to store $\phi_1, \ldots, \phi_n$ is even greater than that needed to store the whole class of figures defined by the pattern!' (p.15)

(3) <u>Time of convergence</u> in a 'learning' process is a related area of neglect. As perceptrons are in practice finite state devices, it is always true that a correct setting of the $\phi_1, \ldots, \phi_n$ can be found, if one exists, by an exhaustive or random procedure, simply by trying different combinations of settings until the right one turns up. The significant question is how quickly some given perceptron might arrive at the desired state, compared with one that used such a procedure. There are 'sets of some geometric

interest for which the convergence time can be shown to increase even faster than exponentially with the size of R' (p.16), the retina.

5.3.3 'Prior structure' — Broadly, we may take the conclusion that Minsky & Papert reach to be that 'significant learning at a significant rate presupposes some significant prior structure' (p.16). This conclusion might be taken as an extension of Ashby's dictum, 'There is no getting selection for nothing' (5.3.0). Put in the terms we have been using, we might give it the generalized interpretation: order extraction will be accelerated by making use of order already extracted. Which amounts to a conjecture that the amount of order extraction is likely to grow according to some kind of exponential function of the amount of order extracted. This corresponds to the conclusions of the previous chapter about the use of heuristics - which may be regarded as equivalent to the prior structures to which Minsky & Papert refer. As these authors' conclusions are neither intuitive, nor of only a practical application (though practicality is always part of their concern), but rest upon detailed and rigorous analysis, they provide important support for the broad conclusions we have reached so far. Most significant in their analysis is the detailed indications the theory suggests of the limits and benefits to be expected from structures of various kinds.

5.3.4 Likeness between art machine and its maker — These remarks on heuristics and prior structures imply certain conditions that will apply to the requirements of an art machine. We shall need to see to it that the machine is able to supply itself with structure (cf. N.4.0) and shall therefore have to be able to say how it is to be acquired and in what form. If we accept the conclusion of Minsky & Papert cited in the previous paragraph, as well as our own conclusions in the foregoing chapters, then to ask that structure be found by prior order-extraction work of the machine is precisely what we cannot do. We shall therefore ourselves have to 'invent' structures for the machine or select from those already available to us. We may suppose, on the basis of the view of Minsky & Papert, that we shall be better at inventing structures than any unstructured machine, because of our ability to draw upon the structre we already incorporate in ourselves. We may suppose further that we are likely to remain better inventors for as long as our structure exceeds that evolved by machines.+ This means that it would be likely to prove extremely difficult if not impossible to evolve within the life-time of an experimenter, wholly novel structures, owing simply to an insufficiency of time and computing power. It follows that all the structures we provide will necessarily reflect us to some extent. This is no more than a re-interpretation of Lofgren's view (2.5.1) that the product of a productive process will bear a resemblance to at least part of the productive element in the process; or more broadly, as we agrued in chapter 2, that every purposeful process is partially self-reproductive. Even if we did succeed in building a machine to generate 'new' kinds of structures, we should probably be unable to distinguish them, precisely because they would not be cast in our image and would therefore appear to us merely 'random' (cf. 2.3.2). By the Law of Requisite Variety, capacity as a selector cannot exceed capacity as a channel of communication. In our terms this amounts to saying that what has to be given (by programming) to a system as stucture, because neither the time nor the power are available for it to evolve for itself from scratch an equivalent structure, - measured in quantity of order-extraction work - may not be acquired by the system through its own efforts, except by an equivalent amount of order-extraction work. But, as it was the lack of time and power to to do ~~the order-extraction work that made it necessary~~ to give it the structure in the first place, programming is unavoidable.

+ If we take order as quivalent to structure, and adopt Lofgren's definition of order, we might take, as an indication of the high degree of structuredness of humans (and other mammals), the ratio of the genetic variety that determines the individual to the variety represented by the individual's cells.

5.3.5 Equivalence of acquiring and communicating order

More precisely, suppose we have two order-extracting systems A_1 and A_2, and that both systems have been provided with a structure S that represents a given quantity Q of order-extraction work. A_1 is then set to work extracting order from a surrounding E. After performing a quantity Q' of order-extraction work in this surrounding, A_1 has succeeded in adding to its structure S the further structure S'. Suppose A_1 now attempts to communicate this newly-extracted order to A_2. On the assumption that A_1 has no more information regarding A_2 than it had of its surrounding E at the start of the experiment, then by the Law of Requisite Variety, communicating S' to A_2 will require precisely the same order-extraction effort Q' as extracting S' in the first place. If A_2 had lacked A_1's structure S, then A_1 would first have had to provide it with this structure, since S' is based upon S, in the sense that A has used S in acquiring S', and because of the growth of order-extraction rate with order extracted, S or its equivalent is indispensible. If the reason for providing A_1 with the structure in the first place was to economize on A_1's time and effort, then the same reason will apply to A_1's 'passing on' S to A_2. But if A_1 cannot pass on S, then neither can it pass on S' (cf.1.2.4)

5.3.6 Advantages and disadvantages of teleonomy

There is no particular reason why one should find this discouraging, unless because it frustrates the wishful thinking that seeks an escape from the consequences of the Law of Requisity Variety. Evidently we should maintain a relatively conservative policy regarding order. For, even if it were true that the structures we had developed were far less powerful than they might have been had we followed some other course, the Law of Requisite Variety bars the way against evolving wholly novel ones without difficulty. The problem is to balance the advantages of tenaciously sticking to an unsatisfactory heuristic against the disadvantages of the flexibility that can perhaps only be won by doing without the heuristics altogether. Chapter 2 argued that teleonomy resulted from structure. Here it appears that the overall teleonomic purpose is bound to be the structure's reproduction of itself.

CHAPTER 6

PRIOR STRUCTURE

SECTION 0 PHASES IN THE USE OF STRUCTURE

6.0.0 The results that this chapter discusses are at once encouraging and disappointing, though their achievement - or lack of it - furnishes no apparent basis for any clear-cut argument on progress: whether for instance Turing's (1950) prediction of successful machine mimicry of human intelligence by the turn of the century was over-optimistic or not. Turing's analysis failed to distinguish sheer 'size' of computers from structure, easily assuming that human programmers would be able to guide the evolution of the structures required. But if the rate of structure acquisition grows exponentially (cf.5.3.3), as this chapter suggests, achievements like Winograd's program (6.3) are perhaps more hopeful. For while Winograd's robot is not half way to human intelligence it might be half way on some logarithmic scale. What this chapter aims to show is the inseparability of the two sets of requirements, those for an art machine and those for machine intelligence. Obvious as this may now appear, the link between artistic and intelligent capacities is neither always clear nor acknowledged. It may be that analysis of pattern-recognition procedures tends to emphasize the intimacy of creativity and intelligence, but this is no more than incidental to its showing the importance of structure to both.

Creativity and intelligence

6.0.1 The chapter examines examples of 'prior structures', in particular descriptors developed for use in pattern recognition procedures. Brief descriptions are given of Guzman's SEE program (Minsky & Papert, 1972). Winston's program (ibid.) for concept formation, based on the SEE program provides an example of how structures may be built up into higher level structures.

Levels of structure

Finally Winograd's (1972) natural language program shows a way of imparting structure acquired in one domain for use in another. There is discussion of the trade-off between the power gained from using descriptors and flexibility for further development.

SECTION 1 OBJECT IDENTIFICATION

6.1.0 To understand more precisely the nature and extent of its advantages and restrictiveness we shall briefly examine some examples of the use of prior structure. In the realm of pattern recognition, prior structures may take the form of descriptors. A pattern-recognition procedure based on such structures recognizes a description of a geometrical figure rather than the figure itself. For example in the SEE program the aim is to divide a three-dimensional scene into 'objects'. The objects for the purpose are rectilinear blocks in two-dimensional projection. This program works at several distinct levels. The first level operates on the optical data, identifying certain optical features: regions, edges and vertices. As the program is to use the vertices as its descriptors, it first identifies them andthen classifies them into types, the most important of which are shown in figure 6.1.0.0.

The use of descriptors

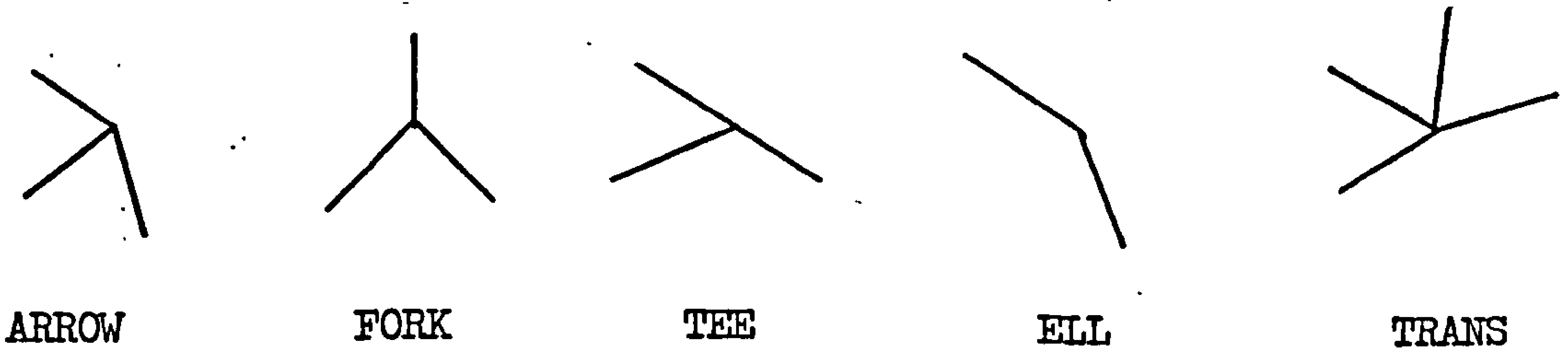


Figure 6.1.0.0

Now it uses these classified structures, interpreting them as evidence that indicates 'links' between regions, and thereby decides which regions should be grouped together as belonging to the same objects - those for which there is the strongest evidence of linkage. In this example the programmer decides both upon what descriptors shall be used and what they are supposed to indicate. They are not evolved or developed by the program. Thus the ARROW-type of vertex is to be taken as resulting from an exterior corner of an object, where two of its plane surfaces form an edge. It is therefore to be regarded as evidence of a link between the two surfaces. A FORK-type vertex is to be taken as evidence for the meeting of three planes of a single object and thus for links between them. These links may then receive abstract representation, as in figure 6.1.0.1 (iii).

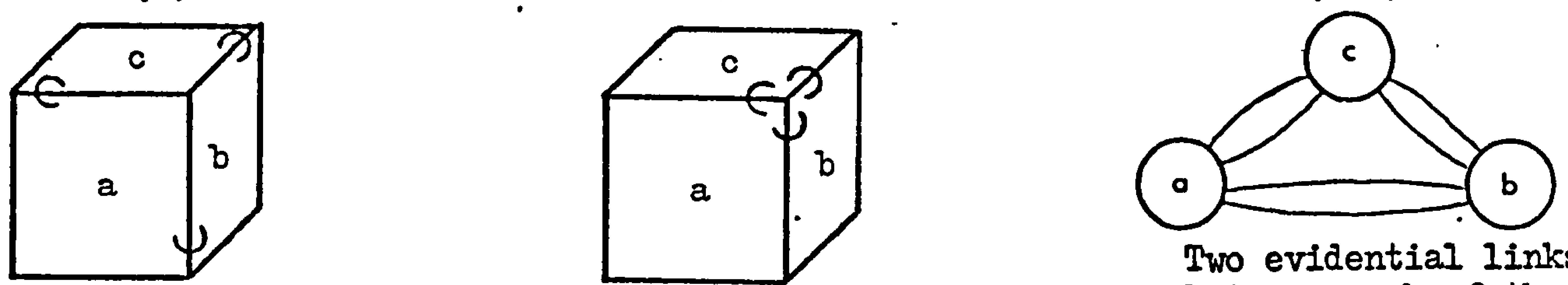


Figure 6.1.0.1

At a further level, the program assesses its own descriptions. For example groupings for which there appears to be strong evidence of connections (regions grouped by more than one link) have single link connections severed (figure 6.1.0.2(ii)).

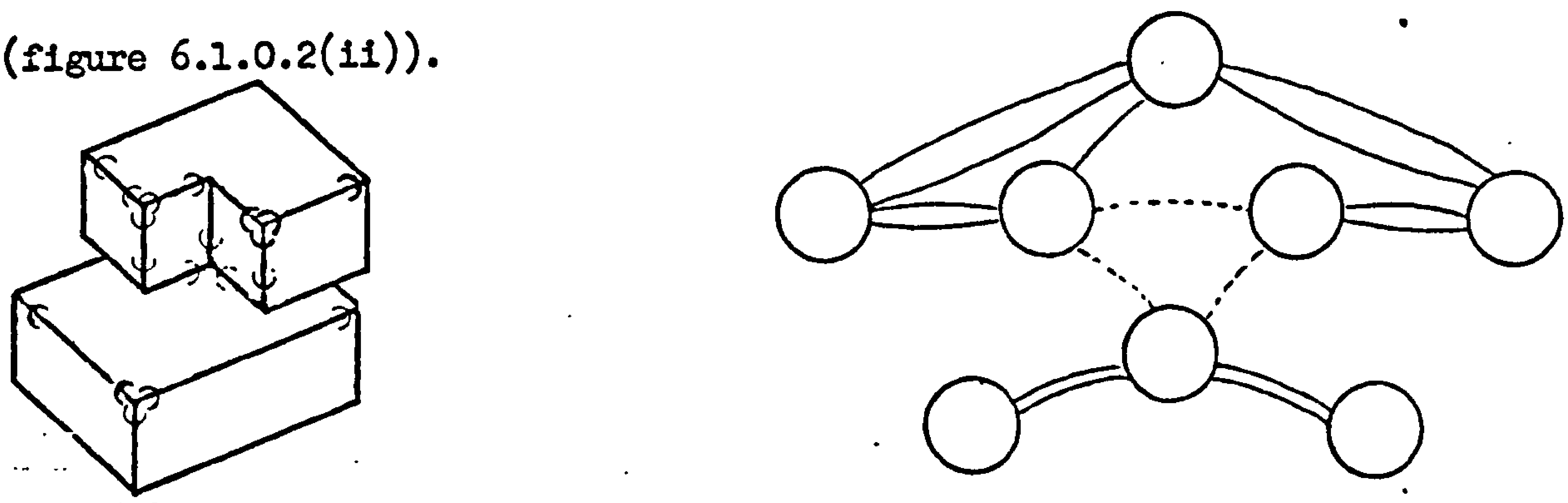


Figure 6.1.0.2

And there are other devices to assist correct identification. Figure 6.1.0.3 shows an example of a scene in which the program successfully separates all the objects

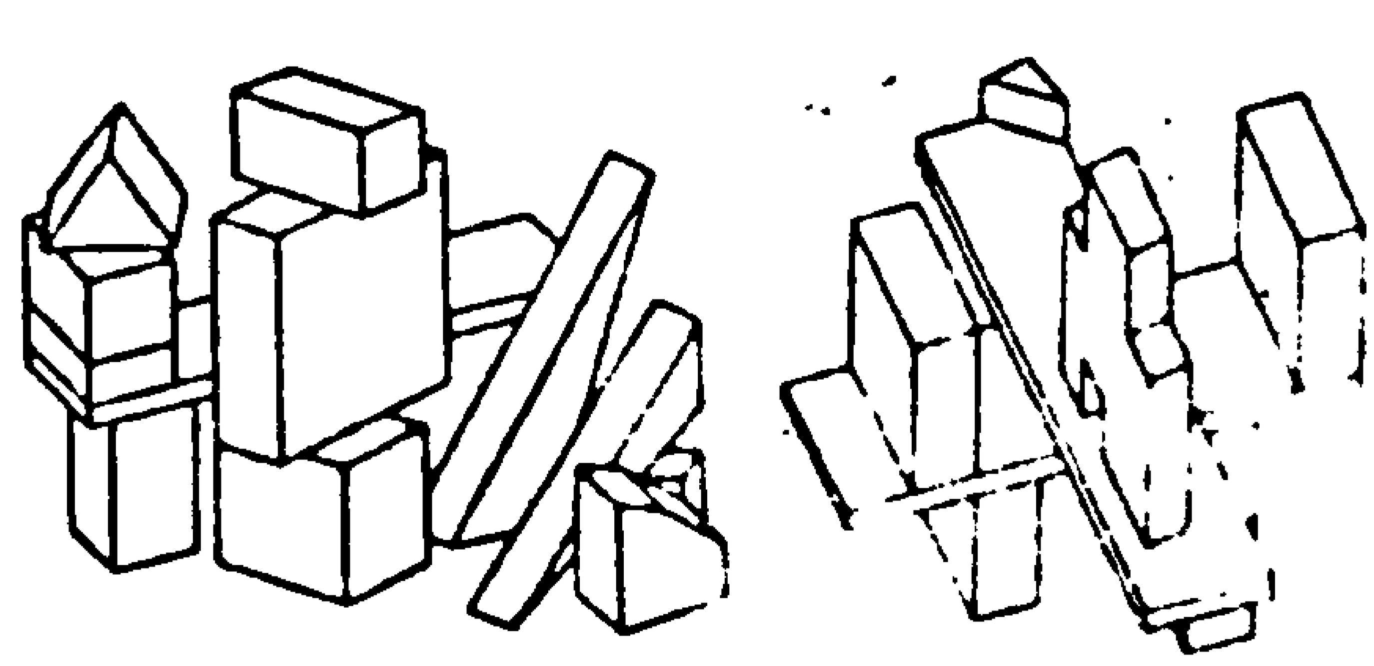


Figure taken from Minsky & Papert (1972)

Errors still occur due to various causes, such as the coincidence of lines as in cases like that depicted in figure 6.1.0.4(i) that may mask the situation of (ii).

(i) Figure 6.1.0.4 (ii)

And errors of identification persist. Some at least of the difficulties experienced by the program are similar to the difficulties that humans have, when faced with similar scenes.

6.1.1 What is chiefly of interest here is not the details of the operation of the programs, but the use that they make of the descriptors (structures) that the programmer provides. Bruner, Goodnow & Austin (1956) suggested the operation of similar kinds of procedures in human perception. They refer to descriptors as 'configurational attributes'. Their remarks draw attention both to the way these configurational attributes provide <u>evidence</u> for perceptual decisions as well as to how they might lead to error. 'A bird has wings and bill and feathers and characteristic legs. But the whole ensemble of feathers is not necessary for making a correct identification of the creature as a bird. If it has wings and feathers, the bill and legs are highly predictable.' (p.47) (cf. Moles's definition of symbols, 10.1.1) The program's descriptors are heuristics for the attainment of goals <u>defined by the programmer</u>. Using such heuristics, the programmer is unlikely to be surprised by the program's results, for it is not to be expected that these will reveal anything previously hidden from him. This is precisely because any procedure using these heuristics will closely mimic the programmer's own behaviour, since the

Programmer seeks program's mimicry

heuristics are likely to resemble those he uses himself. It has been the object of the programmer to produce a program that will make of the world, or at least that limited aspect of it to which it is confined, the same kind of sense as the programmer has already made of it. The programmer never intended the program as an instrument of discovery of new visual images.

SECTION 2 CONCEPT FORMATION

6.2.0 Winston extends the devices of the SEE program that recognize objects, to provide a procedure that can identify assemblies of objects and simple constructions. Winston's program builds up descriptions of complex assemblies by adding to descriptions it already has. The descriptions the program uses are of classes of objects, not individuals. New classes come into existence by means of enumerating the differences between the description of a new individual and some class into which it does not quite fit. The new class consists of the old class plus the (critical) differences. Thus description, and so also identification, of new classes has to proceed to some extent stepwise or the program will end up with a complex set of relations that it is unable to use.

Descriptions of assemblies

6.2.1 As a form of learning Winston's technique represents a considerable departure from traditional 'adaptive' theories. His procedure uses the SEE program as a sub-process and is further equipped to recognize spatial relations, such as contact and support, between the objects that it uses the SEE program to identify. By means of these provisions it is possible to furnish it with descriptions of assemblies. Thus an assembly of three bricks, one supported by the other two might be presented as an example of an arch (figure 6.2.1.0(i

Forming and generalizing a concept

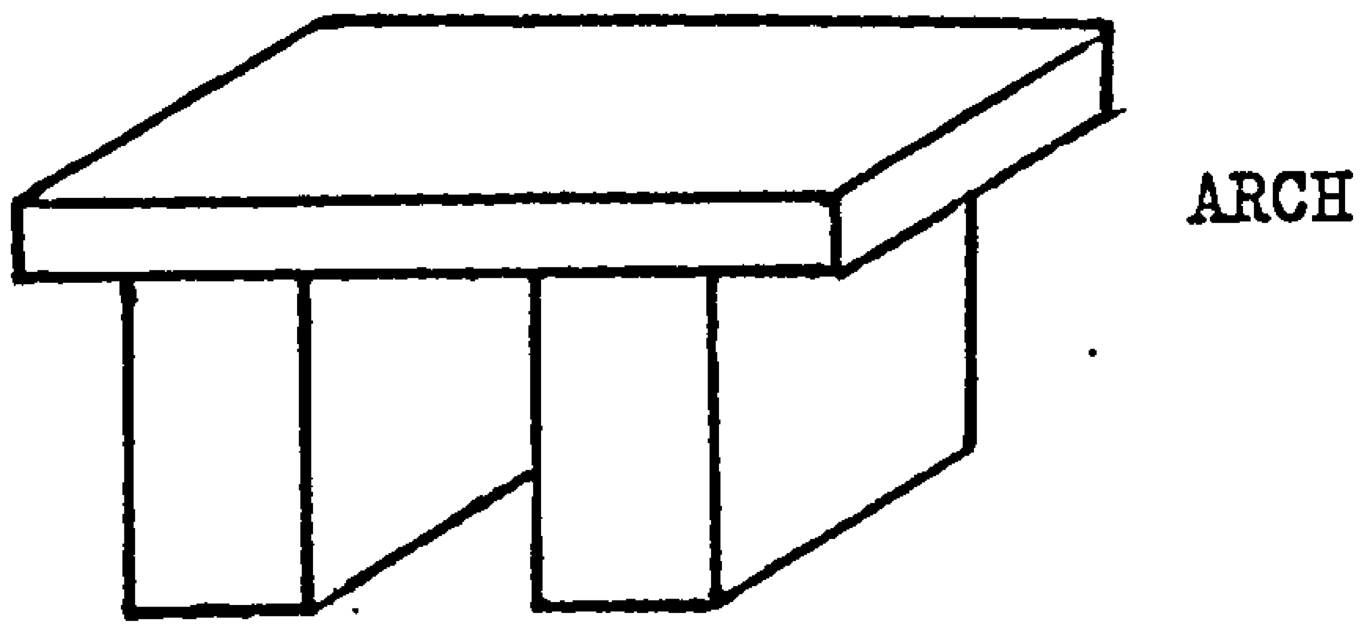
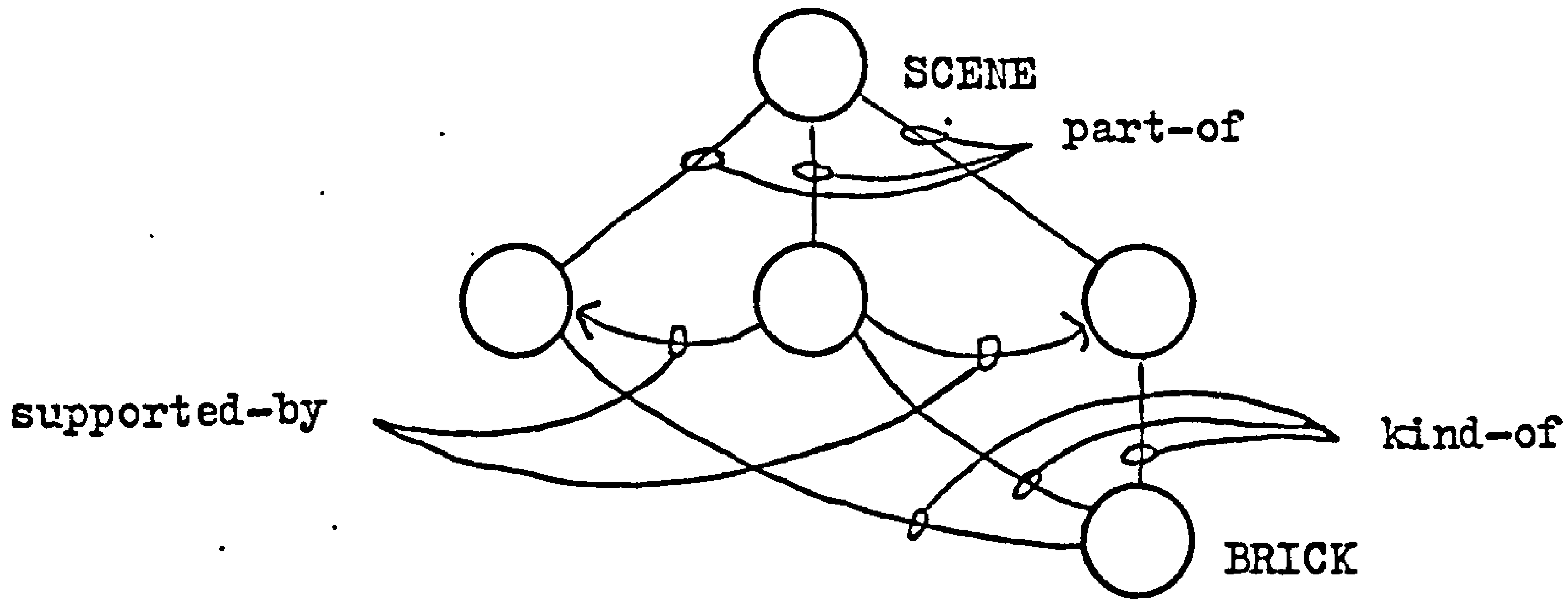


Figure 6.2.1.0 (i)

The program then stores a description of the example, using the terms it has available such as 'brick', 'supported-by' and so on, in a form which may be represented by a network.



Large circles represent particular physical objects, small circles other kinds of concepts, and labels on the arrows relations.

Figure 6.2.1.0 (ii)

Given another example (figure 6.2.1.1(i)) it is told that this is not an arch.

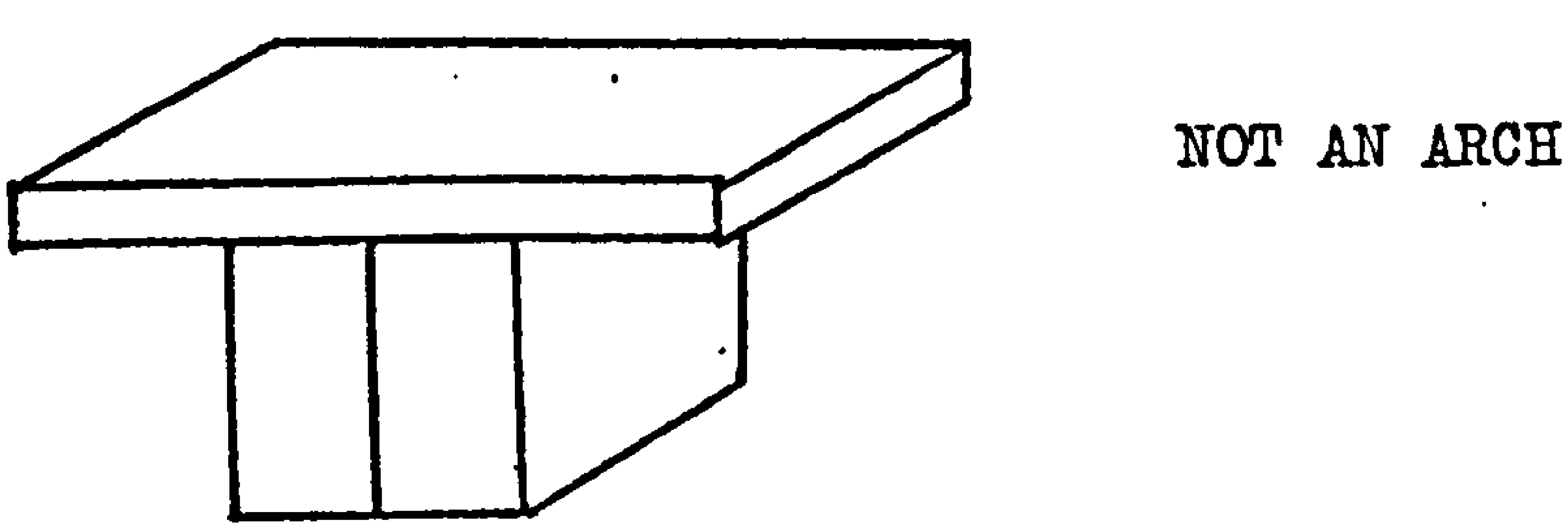


Figure 6.2.1.1 (i)

It forms a description of the new construction, compares it with the arch description, notes the differences (contact between the support bricks) and adds to its arch description that there should be no contact between its support bricks (figure 6.2.1.1 (ii)).

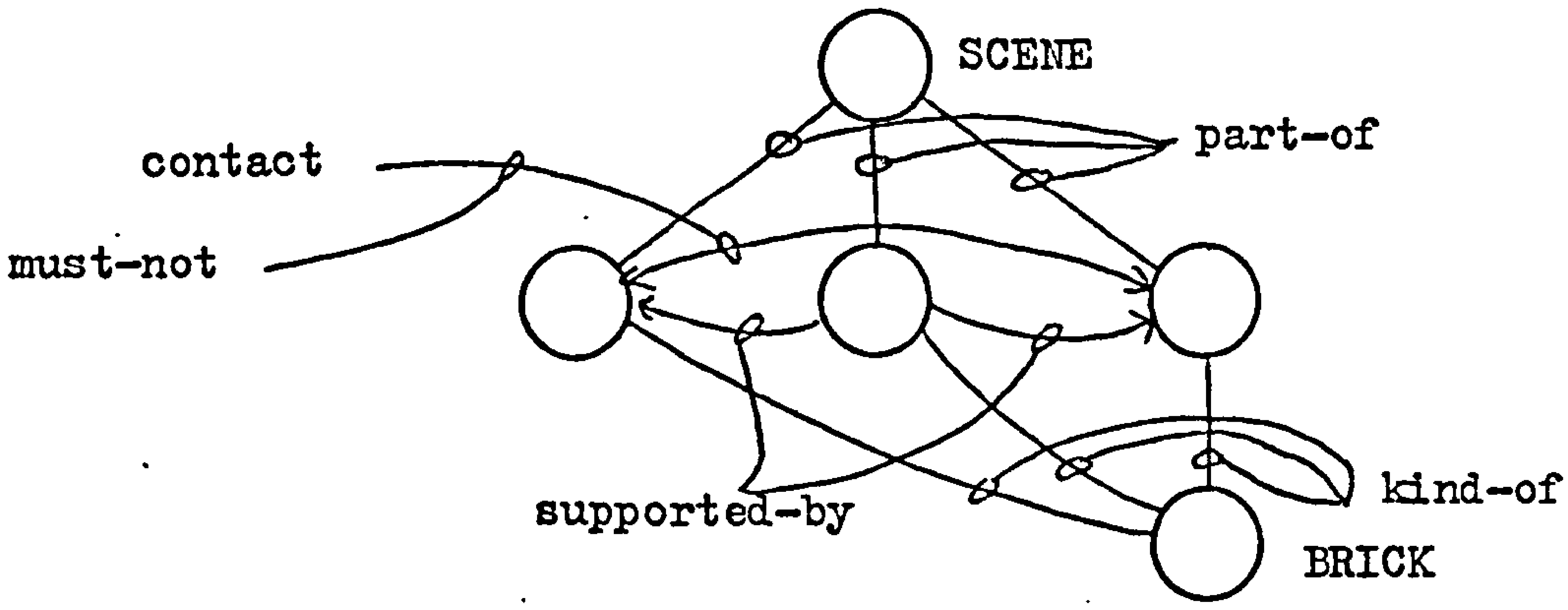
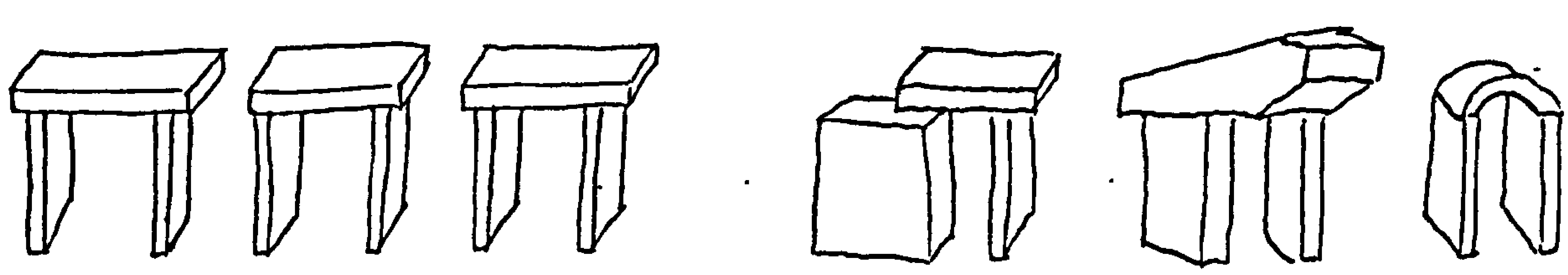


Figure 6.2.1.1 (ii)

Further examples of arches and non-arches may be given, and the arch and non-arch descriptions modified accordingly. For instance the support bricks may be supporting an object drawn from a wider class than the class of bricks, the class of prisms say. Or the supports themselves may be drawn from this wider class. Using these examples and its procedure for noting differences, the program 'generalizes' the description of its 'concept'. So that, starting with Guzman's descriptors and a few classifications of relative spatial relations, it can build up (learn) concepts that enable it to identify, first objects, then assemblies of objects and then, using these new concepts, relative spatial relations of these assemblies. Thus it will learn to identify a complicated scene such as figure 6.2.1.2 (i) or figure 6.2.1.2 (ii) by means of the network of figure 6.2.1.3, as a row of arches, a good example of what Gagné (1962) refers to as 'productive learning' (p.355)



(i)

(ii)

Figure 6.2.1.2

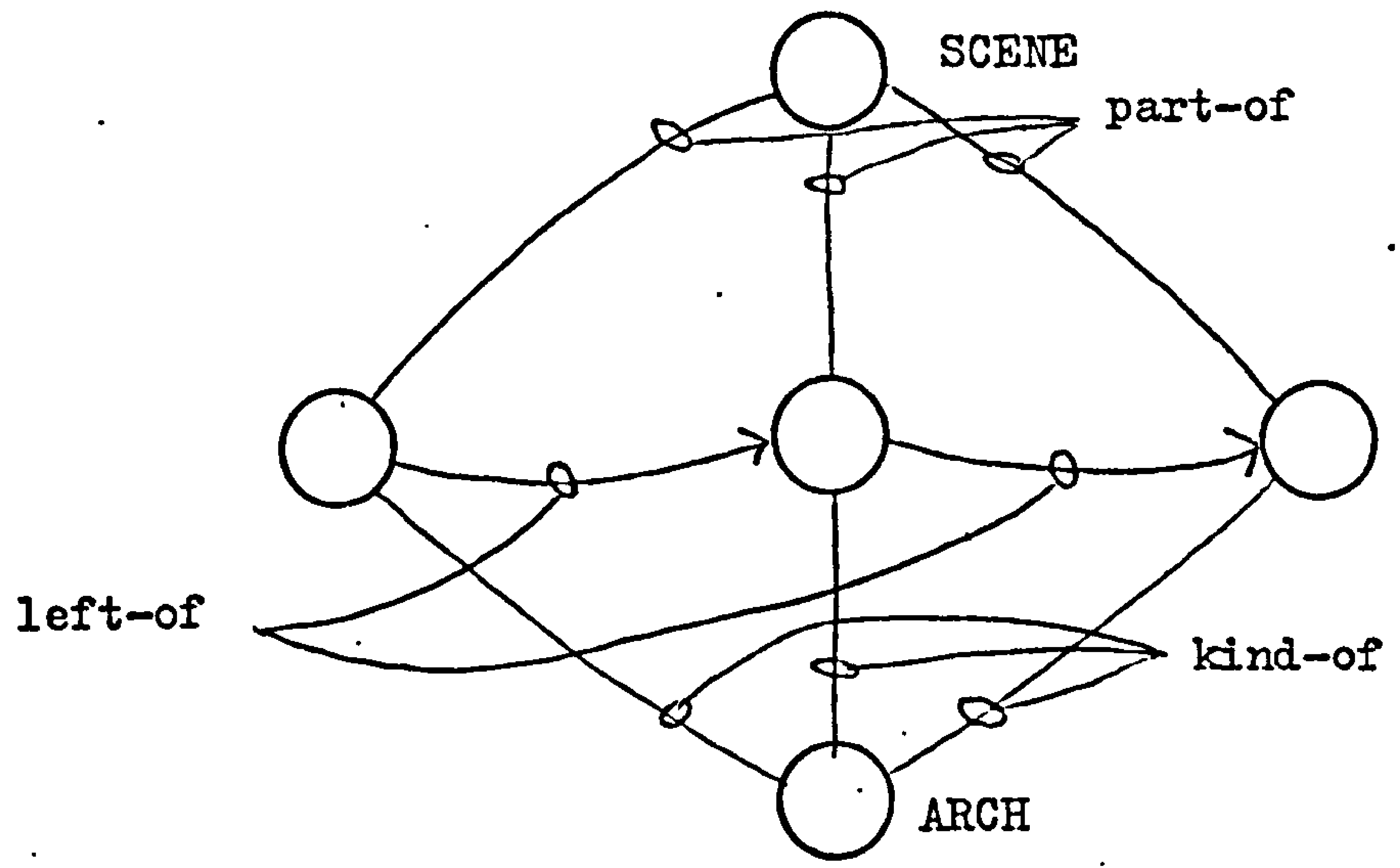


Figure 6.2.1.3

Determinants of concept formation: 'blinkering' effect

6.2.2 Like the SEE program, Winston's procedure relies for its capabilities upon descriptors provided for it by the programmer, both those that enable it to break scenes up into objects - the Guzman vertices - and those that relate the objects spatially. Taking order extraction as an hierarchical process, as Winston's program does, we should conclude that the program's power derives from its use exclusively of relatively 'high-level' concepts - its descriptors. By the conclusions of the previous chapter, the fact that they are high-level implies that these descriptors must have been slow and difficult for the programmer to have acquired in the first place, acquisition dating far back into evolutionary history. Notice that using these descriptors makes other 'kinds' of concepts literally inconceivable for the procedure. It has already been intimated by the observations of Newell & others (4.6.1), how difficult new descriptions may be to digest; and the 'blinkering' effect that partly accounts for the difficulty is intensified by making use of the descriptors relatively early in the identification task, that is before much other order-extraction work has occurred. Evidently the power of the descriptors is bought at the cost of narrowing the set of attainable goals. (cf. 4.5.1) Notice too that the concepts that the procedure acquires will depend upon the order in which the program encounters its examples, though not in an entirely inflexible way. The noting of differences between concepts must involve a procedure for comparing descriptions. If descriptions are held in the form of a series of networks, then the description of the difference between two such networks will itself be a network in which each node might refer to a pair of nodes, one from each of the compared descriptions. Such comparisons would require conventions for deciding which nodes of the compared descriptions were to be matched and which matchings were to receive priority. And the conventions and priorities might themselves be subject to adjustment by the application of heuristic rules. In general, although what concepts are formed will depend upon the order in which modifications occur, flexibility will derive from the mode of operation of the program. This is because the procedure operates by matching a description to an example, as an hypothesis, which is then whittled down into a plausible form by noting differences. Flexibility will partly depend on whether hypotheses are relatively generous or tend to be conservative, that is whether descriptions are broad or narrow.

SECTION 3 SUPPORTS FOR LANGUAGE STRUCTURE

6.3.0 Going on from here, Winogard adds to the objects and their spatial relationships of Winston's program <u>actions</u> that may be performed upon them. The 'blocks world' becomes a <u>micro-world</u> able to provide a language-understanding system with a subject domain for discourse. To make the micro-world suitable for this purpose, Winograd provides it, besides objects, properties, - a class including relations - actions with goals, processes and simplified forms of concepts like space, time and purpose.

The 'blocks world' as a domain of discourse

> 'We can describe the process of understanding language as a conversion from a string of sounds or letters to an internal representation of "meaning". In order to do this, a language-understanding system must have some formal way to express its knwoledge of a subject, and must be able to represent the"meaning" of a sentence in this formalism. The formalism must be struc-

tured so the system can use its knowledge in conjunction with a problem-solving system to make deductions, accept new information, answer questions, and interpret commands.' (Winograd, 1972, pp.23-24)

'In practical terms, we need a transducer that can work with a syntactic analyser, and produce data which is acceptable to a logical deductive system. Given a syntactic parser with a grammar of English, and a deductive system with a base knowledge about particular subjects, the role of semantics is to fill the gap between them.' (p.28)

6.3.1 In Winograd's system meanings are mostly represented by programs written in PLANNER (Hewitt, 1969), a language for proving theorems concerned with actions and goals (of a robot), which is able to guide its proofs by heuristics provided from the knowledge of the world in which the actions and goals occur. PLANNER sets up series of sub-goals to be carried out in a set order, to bring about an event. Thus, told to GRASP a block, the robot which performs the actions on the blocks (in this case, a simulated robot depicted together with its world on a CRT screen), may first have to GET-RID-OF a block it is already grasping. To do this it must FIND-SPACE in which to set the block down. Having got rid of it, it may have to CLEARTOP of the block it has been instructed to grasp, by removing a block covering it. This may involve finding space to which to move this covering block, and so on. The names in capitals represent PLANNER theorems. Each action which cannot be carried out causes the program to back up to a preceding theorem, until this process is exhausted, at which stage failure is reported. A series of actions constitutes an 'event' which is remembered and which furnishes information about the past as well as reasons (explanation) for carrying out actions. Events are timed by a clock which starts at zero and is increased by one every time a motion occurs. A second kind of memory keeps track of the positions of objects as they are moved.

Actions and meanings

6.3.2 The success and flexibility of Winograd's system results from the interposition of its semantic system between the syntactic constructions and the PLANNER programs that define the meanings of words and other constructions, so that it can deduce from these, new procedures for the deductive system to use in answering questions about obeying commands and acquiring new knowledge from and as a result of the dialogue. Since the dialogue is about the blocks world, syntactic problems are usually soluble by reference to this world. It is the ability to make such reference that enables the program to avoid building an unwieldy, purely syntactical system to overcome all the difficulties and ambiguities of the syntax. In the terms formulated in the foregoing chapters, the success of Winograd's system depends on the arrangement that succeeds in providing the syntax of the system with a richer source of order than is represented by syntactic rules alone. Heuristics like the edges and vertices of the SEE programme, embody the results of considerable order-extraction work, which is absent from the quickly-growing systems of rules in language-understanding systems that rely purely on syntax.

'Blocks world' structure carried over into language structure

6.3.3 This interpretation relies on the theoretical considerations so far developed and upon the supposition that the blocks world descriptors reflect large amounts of order-extraction work. Notice therefore that several independent lines of empirical evidence support this supposition. For instance human retinal function acts to increase the acuity with which 'edges' are detected, while the SEE program provides edges as data. But anatomical and physiological features responsible for edge acuity in humans evolved over extremely long periods, though the SEE program almost wholly conceals the extent of the order-extraction effort that the period's length suggests. Other naturally occurring descriptors are also comparable with Guzman's vertices, and also represent results of order-extraction work accomplished over long periods of evolution. The mechanisms of the frog's eye, identified by Lettvin, Maturana McCulloch & Pitts (1965) suggest specific simple analytical features for detecting kinds of order important to the frog's survival, namely shapes related to food and to danger. Considerable evidence for similar kinds of mechanisms

Descriptors are 'high-level' structures

in other species is collected in Sutherland & Mackintosh (1971). Gestalt psychology provides another line of empirical evidence of a kind suggesting that descriptors play an important part in 'recognition'. So does developmental psychology. A child's 'matchstick' drawing of a man is more intelligibly seen as a visual description than as an attempt at visual representation.

CHAPTER 7

EXTENDING RECEIVED ORDER

SECTION 0 THE MILIEU OF PURPOSE

7.0.0 The examination of purposeful systems thus far has paid little attention to the surroundings in which they operate. In reality these surroudings are all-important. The milieu of a structured dynamic environment, filled with the activity of mutually competing and co-operating systems in continually changing organization is not merely the backdrop to purposeful behaviour, but forms the fabric of purpose itself. We have observed (N.3.0) that the creative additions to organization that any structured system achieves are likely to be small in relation to the complexity of interconnections and constraints of the evolved, existing order into which such systems are 'born'. Ashby (1964) remarks that 'the organism can adapt just so far as the real world is constrained, and no further' (p.132). So even its creative effort - its adaptation - merely reflects its world - and its ability to survive in it. Commonly adaptation and survival, especially in varied circumstances, are taken to signify intelligence. But we have argued for the particular importance of that aspect of intelligence to do with extracting order from its surroundings, what we have called creativity. This chapter seeks to step back from the structures that have been our concern, to see them in that wider context.

Purpose and adaptation

7.0.1 Attention shifts away from structure, descriptors, heuristics, forms and the like, to the organization of goals and purposes. Goals and sub-goals are discussed, particularly when sub-goals operate through purposeful systems. This raises questions of specifying goal-seeking systems. Problems of goal 'transmission' raise questions of coding and communication between systems.

Co-existence of purposes

Concepts of compatibility and redundancy and apparent compatibility and redundancy of specifications are defined and examined in the light of notions of requisite variety. This leads to conclusions as to the limits to the possibilities of goal transmission, both absolute limits and those due to problems of interpretation. There is discussion of the relative limits between two systems in relation to the concept of fuzziness. The chapter then proceeds to apply its conclusions thus far. The object is to obtain a view of purposeful systems within the constraints of an active surrounding that restricts the possibilities of goal-generation and, thereby, of the capacity for c-authorship. Constraints are defined in terms of goal specifications. This enables the physical world and goal-seeking systems to exhibit their mutually dependent contributions towards forming a world for A, There is discussion of chains of goals, with the broad object of capturing the notion of development, both past and future; and there is discussion of specification by injunction and description in relation to limits on originality.

SECTION 1 LINKED PURPOSEFUL SYSTEMS

7.1.0 In 2.1.2 we defined purpose in terms of a purposeful system A acting within (and as part of) a surrounding E. A's purpose was to maintain invariant some internal state of E - or part of E, possibly itself - by suitably varying its own output. In many instances, as we have mentioned, A's purpose was to <u>bring about</u> some particular state in the future. We interpreted this, in accordance with our definition, as the maintaining invariant by A of the 'movement' towards the desired state. We referred to A's description of the state as a <u>concept</u> of A's. The generality of such abstractions, although providing clarity of definition, obscures the richness and complexity of the

Goal hierarchies

detailed working of purposeful behaviour in complicated systems. In particular, it ignores the structures of A and E. Thus, for A to bring about some internal state Z_E of E may entail its first bringing about z_E and so on. Such ordering of goals and sub-goals may be expressed in terms of procedures such as PLANNER or given more general representation as 'knowledge structures' such as those of Gagné (1962, 1964 for example), or other kinds of 'relational nets' such as Pask's 'entailment structures' (see for instance Pask, 1975). We arrive at a concept of 'an integrated sequence of activities' (Sommerhof, 1969), which '<u>stands for a relation between these activities which enables us to attribute an individual goal to each, and at the same time an ultimate goal to the whole sequence</u>' (p.188). This concept leads directly to the further concept of 'hierarchies of directive correlations, integrated by a single ultimate goal' (p.189). Fogel & others note that complete goal specification calls for a full statement of penalties and pay-offs for each point of time in the future, though in general, as we shall see (7.2.1.1), we shall be able to assume a limit to the number of future scenarios that a finite machine will be able to discriminate. The same reasoning permits the reduction of possible histories of a machine to a finite number. (See for instance Minsky, 1972). These discriminable scenarios are the distinguishable states of the system.

7.1.1 An extension of the use of sub-goals toward the attainment of an ultimate goal of particular importance to authorship as we have defined it, occurs when the sub-goal of a system itself involves other purposeful systems. In this case A does not achieve the sub-goal 'directly', but instead specifies to the intermediary purposeful system, call it A+, the goal which that system should aim for. The specifications that A provides take the form of instructions that A issues, to be carried out by the intermediary system A+. If A's sub-goal is g(A) say, then A's problem is to provide A+ with an input state

Goal transmission

such as will cause it to emit an output that will constitute the execution of g(A).

7.1.2 Specifying a goal to a purposeful system

We have defined A as a set of procedures and shall extend this definition to provide a definition of goals in terms of sets and combinations of sets. Roughly speaking we shall regard the goals of a system as so many target areas, each defined by a set of specifications, with intersections and unions of such sets representing corresponding intersections and unions of respective target areas. We shall suppose that specifications are compatible (see 7.1.3) and, to capture the sense of an hierarchy of specifications, monotonically narrowing the target area. We shall suppose further that redundant specifications (7.1.3) have been deleted. The sets we are left with will therefore form a monotonic sequence corresponding to a partially-ordered system, presented as a list. Pask's entailment structures are examples of such a system. Notice that presenting such systems as hierarchies is to some extent arbitrary. Entailment structures for instance begin as closed 'relational nets' in which certain nodes are selected as 'head' nodes and the nets pruned accordingly, to yield partially-ordered sets. If A is a set of procedures there must be among them subsets corresponding to each goal (including its sub-goals) that A may entertain. Suppose $\mathcal{G}$ is the set of all such subsets. We shall call a subset $G \subset \mathcal{G}$ a <u>specification</u> of A's goal. A sequence of such specifications, G_i $(i = 1, \ldots, k)$, say, defines a subset $G' = \bigcap_i G_i \subset \mathcal{G}$. G' might be an insufficient specification of g, as would be the case if for example it left ambiguous what was required of A+ at some point in the procedure it specified. We shall say, in such a case, that g is an <u>under-specified</u> goal of A. But if G' is sufficient and A+ is able to use it to set up its own goal state g(A+), to be the same as A's sub-goal g(A), that is if W(A+) the realization of g(A+) and W(A) the realization of g(A) are identical, we shall call G' a <u>full</u> specification of A's goal and say that G is a <u>fully-sepcified goal</u> of A.

7.1.3 Suppose G' is such a full specification of g and G^* is an additional specification, such that $G^* \neq G_i$ for any i, then, together with G', G^* defines a new set of specifications $G' \cap G^* = G''$ say.

Compatibility, incompatibility and redundancy

(1) If $G'' = \emptyset$, we shall set that G^* is <u>properly incompatible</u> with G'. This will occur if for example G^* represents an instruction that is mutually contradictory with G' or if G'' is ambiguous or incomplete as above (7.1.2).

(2) If $G'' \neq \emptyset$ and $G'' \neq G'$ represents a full specification of a new goal $g(A)$. We call this <u>proper compatibility</u> of G^* with G'.

(3) If $G'' = G'$ we shall say that G^* is <u>properly redundant</u> with G', since the goal $g(A)$ specified by G'' cannot be distinguished by A from that specified by G'.+

7.1.4.0 But as G' is a specification by A for the direction of A+, it must be interpretable by A+. Clearly this raises questions of <u>requisite variety</u>. Thus, given case (1) of the previous paragraph, A+ might interpret G^* either as compatible or incompatible with G'. In the former case we shall say that G^* is <u>apparently compatible</u> with G', with respect to A+. Similarly cases (2) and (3) may give rise to interpretations that we should call respectively <u>apparent incompatibility</u> or <u>apparent redundancy</u>.

Apparent compatibility, incompatibility and redundancy

7.1.4.1 Examples of these various cases are as follows. Suppose G' could be mapped into a white plane surface to make a black pattern on the surface, and suppose this pattern were then to be mapped onto a regular grid of squares of given fineness of 'grain' (corresponding to length of side of squares), according to the rule that a square more than half covered by the pattern should by

Illustrations of definitions

+ This definition is consistent with that of Shannon and Weaver (1949). For, if the redundancy R of an information source is given by $R = 1 - \frac{H}{H_m}$, where H is the entropy of the source and H_m is the maximum possible entropy it could have while restricted to the same variables, and if we take A as the information source, and the set of instructions G' emanating from A, as the variable, then by hypothesis, G^* will not alter the information carried by the source; that is $H(G' \cap G^*) = H(G')$, though, since $G^* \neq G_i$ $(i = 1, \ldots, n)$, H_m will be increased.

made entirely black, otherwise entirely white. Such a 'retina' could represent 2^n different patterns, where n was the number of squares in the grid. (In fact each grid pattern would represent an infinity of 'real' patterns, since there are an infinity of ways of more than - or less than - half covering a square with a pattern. Figure 7.1.4 illustrates the various cases that arise.

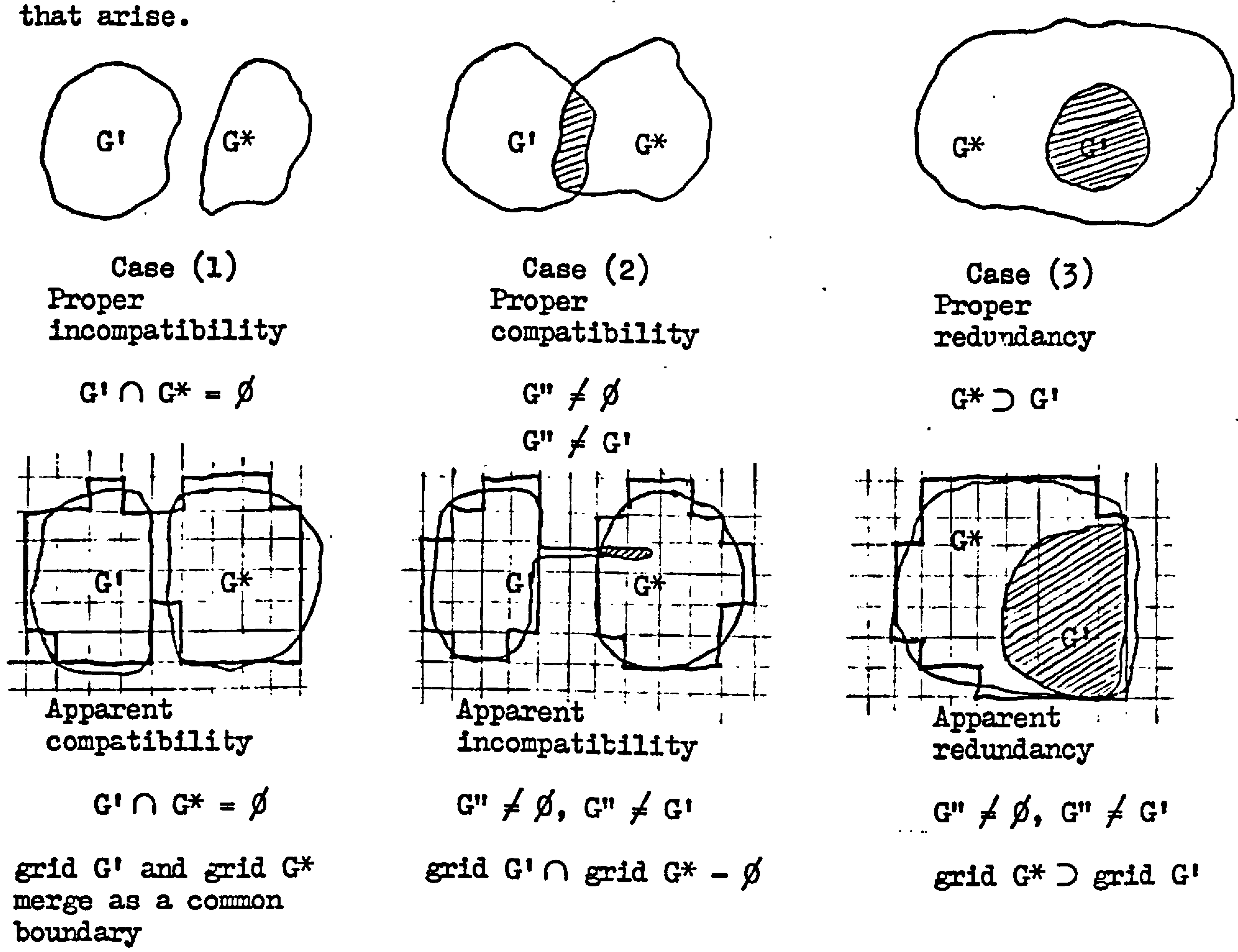


Figure 7.1.4

7.1.5 If (1) $G' \neq \emptyset$ and (2) G* is either properly (apparently) incompatible or properly (apparently) redundant with G' for every $G^* \subset \mathcal{G}$, then the goal g that G' specifies is (apparently) unique, since it cannot be specified in any other way. Thus the (apparent) redundancy of G' is zero. We call g a <u>specific goal</u>.+

+ cf. Löfgren's definition of order (2.3.0). The specifications of a specific goal are equivalent to the shortest starting tape x from which a universal Turing machine will generate the sequence ξ.

SECTION 2 INTERPRETATION

7.2.0 Clearly, as the figures of 7.1.4 suggest, difficulties arising out of what we have here called the interpretation of goal specifications or instructions are part of the whole field of pattern recognition and order extraction. The grid examples show for instance that much that was said about perceptrons will apply literally here too. More broadly, the ability of A+ to interpret goal specifications provided by A will involve many of the questions discussed in connection with heuristics and order extraction generally.

The context of goal specification

7.2.1.0 The difficulties of what we might call goal transmission reveal problems of some depth, to which we referred in chapter 2 when dealing with questions relating to observers. Basically we are concerned with the effectiveness of goal specifications. What we have so far said tacitly assumes the existence of a 'real' set of goal specifications underlying, as it were, the specifications entertained by any given entity, (as we might assume some 'real' length 'underlying' the approximations to it of actual measurements). We might approximate such a set of specifications using the grids of 7.1.5.1, by allowing the grain of the grid to become arbitrarily fine. But such a procedure would require giving an interpretation to the notion of a goal with an arbitrarily long specification; one, that is, that took an arbitrarily long time to specify and which therefore could not be a goal at all, as the term is ordinarily used.

'Real' goals and actual goals

7.2.1.1 Suppose for example an entity A were given the goal of partitioning the set R of real numbers by means of a Dedekind Cut at a point y, into two sets L and U. It would appear that such a specification to A should be effective, since it provides a procedure according to which, given any number $x \in R$ it would be possible for A to decide to which set, L or U, to assign it. But R is an infinite set and, in particular contains an infinity of members designate by more than some arbitrary number k of (say) decimal digits. In particular, R contains an infinity of members of this class whose first k digits are the

Finite machines

same and which differ only in digits following the k^{th}. It follows that if y consists of k or more digits, it will take A an arbitrarily long time to check any number having the first k digits the same as y (assuming it takes some finite time to check each digit). If A were a Turing machine, it would require arbitrarily long tapes, or if it were represented by a grid, the grain would have to be arbitrarily fine. Turing (1937) uses an argument similar to this to show that, 'If we admitted an infinity of states of mind, some of them will be "arbitrarily close" and will be confused' (p.250). A's task becomes impossible when y is a non-computable number. In other words, in general A will find the goal specification for partitioning R non-effective. 'Reality', in this sense, is thus unknowable by any finite machine and approximations to knowing it depend on the machine's variety.

7.2.2 What then does A understand by the instruction to partition R? In one sense one might argue to the effect that A has a 'belief' that it is possible to partition R. Such a belief need not have a merely intuitive basis. It might be based upon a p-explanation in an r-formal theory (2.3.2). However, such a p-explanation would itself depend ultimately on a set of axioms that remained unexplained (the 'final explanans .... without explanation' (Lofgren, 1972, p.341)).+ It appears that the 'fuzziness'++ of instructions given to A, as A is able to interpret them, depends upon A's variety. We might put it that A's goal in the case of partitioning R, even A's notion of R, has the status of an <u>hypothesis</u> not of an <u>experience</u>, (chapter 10). It represents knowledge by <u>description</u> rather than knowledge by <u>acquaintance</u>.

Variety and 'fuzziness'

+ It follows from this that no entity exists that could effectively understand (in the sense of 2.3.2) the 'perfect continence' on which Spencer-Brown bases his definition of the primary distinction.

++ Zadeh (1965) defines a 'fuzzy set' A in X, where X is a collection of points $\{x\}$, by means of a characteristic function $\mu_A(x)$, which assigns a number in the interval $[0,1]$ which represents the 'grade' of membership of x in A. The nearer the value of μ_A to unity, the higher the grade of membership of x in A, and conversely. We might relate the term 'fuzziness' as we have applied it in this paragraph to Zadeh's concept, by associating the value of Zadeh's characteristic funtion μ with the fineness of the grain of the grid. In view of the argument of 7.2.1.1 the grid's fineness has a finite limit, so that in effect μ must fall in the semi-open interval $[0,1)$, unless it is a measure applied by a meta-observer.

7.2.3 In general, if A specifies a goal for A+ in the form of an effective procedure, A+ will need variety at least as great as A's in order to be able to execute the procedure. Recalling the grids of 7.1.4, we might put it that the A+'s

Resolving power

resolving power should be at least as great as A's. But sufficient variety may not in itself provide a sufficient condition, since A+ will require also all those forms of orderedness (constraints on its variety: what we have discussed under the headings heuristics, descriptors and the like) that A exhibits in specifying the goal in question. A lower variety in A+ or an absence in A+ of some form of orderedness such as is present in the goal specifications from A, will lead A+ to judgements of compatibility, incompatibility or redundancy in these specifications, which are not evident to A. It follows from these arguments that the terms proper and apparent as we have used them to denote kinds of goal specifications, are meaningful in effect only with respect to some given A. The implication of this is that the term 'apparent' as we have defined it, may only be used by an observer to describe his judgement of some particular operator executing a given set of goal specifications. When an observer in such a case uses the term 'apparent', it implies that the description 'proper' applies to the observer's own judgement of the same specifications. Clearly it would not make sense for the observer to apply the word 'apparent' to himself, though having in mind an argument like the present one, he might do so as a meta-observer.

SECTION 3 CODING AND COMMUNICATION

7.3.0 These observations imply certain broad restrictions that would apply to making an art machine. We have shown reasons why it is necessary to furnish the machine with structure: heuristics, descriptors and such. They are that (1) significant learning, (order-extracting ability, on which depends the capacity

The dilemma of providing structure: recapitulation

for c-authorship;) at a significant rate pre-supposes some significant prior structure (5.3.3); and, as 'There is no getting selection for nothing' (5.3.0), the cost in computing time and power of evolving the required structure from scratch would make such an evolution impractical; (2) even if the machine were able to evolve a sufficient structure for itself, the likelihood is that the structure would be of a kind that its observers would fail to distinguish As we have said (5.3.4), to be able to make art that its maker will recognize, an art machine will need to resemble him. This was previously shown to be the case in the domain of order extraction. Here the notion is extended to apply explicitly in the realm of purpose. This implies that the machine's maker will provide it with structure that resembles his own - including all the structure he has acquired by his own order-extraction work. He does this to permit communication between himself and the machine, and because he could not do otherwise, having no other structure to offer. Such structure will in turn restrict the kinds of goals that the machine will be able to generate and thus, in terms of our definition (3.2.2), its powers of c-authorship. How much scope for c-authorship may the machine enjoy? Too little structure provided to the machine will render it incapable of significant results. The evolution of structure too unfamiliar, corresponding to too wide a scope as a c-author, will cause failure to recognize its products. Here is a dilemma for the machine's maker: unless he furnishes the machine with previously extracted order, its order-extracting ability will be limited; provided with all the order its maker could give it will make it indistinguishable from him. Clearly the machine will have to be constructed along the lines of heuristic chess-playing programmes for instance, beginning with some simple device and adding structure to it. Chapter 12 looks forward to possible devices we might begin with. Here we note that before a machine could show significant creativity, it would probably need to incorporate most of its maker's (human) structure. The rest of this chapter is devoted to questions as to the kinds of limits that given structure will impose. As an art machine will be a purposeful system, we shall examine questions of the relation of structure to goal-generating capacity, that is to c-authorship.

7.3.1 A's structure restricts its freedom to set goals. Its structure represents a diminution of its variety and therefore of its capacity to be a c-author. According to Ashby (1964) a constraint is a relation between two sets which limits the variety under one condition, making it less than under another. It follows from this definition that a constraint on one system may be brought about by another system or systems. Thus we take it that A's structure that restricts its freedom, includes its goal structure as part of it. If a condition C restricts $\mathcal{G}$, the set of all possible goals of A, to some proper subset $G \subset \mathcal{G}$, we shall call C a <u>constraint</u> on $\mathcal{G}$. Thus the constraint C has the force of a goal specification, since it has the effect of making A include G among any set of goal specifications $G' = \bigcap_i G_i$, $(i = 1, \ldots, n)$. It is a condition that restricts the goals that A may generate. We distinguish <u>internal</u> constraints due to A's structure from <u>external</u> constraints resulting from structures existing in A's surroundings. Among external constraints on A we distinguish those due to other purposeful systems from those arising out of non-purposeful elements.

Goals and constraints

7.3.2 Suppose a purposeful system A suffers constraints due to another purposeful system A*. Specifically, suppose A*W is a process in which A*'s goal is to operate on E to produce W(A*). Suppose further that this entails A*'s working through the agency of A. Instead of itself operating directly on E, A* utilizes the operations of an intermediate process AW that in turn will operate on E to produce W(A*). But A is itself a purposeful system and will therefore produce W(A) as the realization of its own goal g(A). Thus A* must <u>control</u> A to make it produce W(A*). Controlling A represents A*'s sub-goal which we write $g_s(A^*)$. The execution of $g_s(A^*)$ results in the production of $W_s(A^*)$, where $W_s(A^*)$ has the properties which A* supposes necessary to exercize the required control of A in AW. That is, $W_s(A^*)$ constitutes a set of instructions provided by A* to A from which A* supposes A will derive specifications G(A)

Two purposeful systems in series

of a goal $g(A)$, whose execution will result in $W(A) \equiv W(A^*)$. When A* aims to realize a goal in this way we shall call the aim A*'s <u>intention</u>. We may imagine a process AW such as we pictured in 1.1.0, consisting of a purposeful system A that operates through a transducer M. M is the <u>effector</u> system through which A acts on its surrounding E to produce W. By extension, in the present case we may suppose that A* acts through M* to produce W*. For A*, M* includes the whole process AW, along with whatever other elements there are in its own effector system. (See figure 7.3.2)

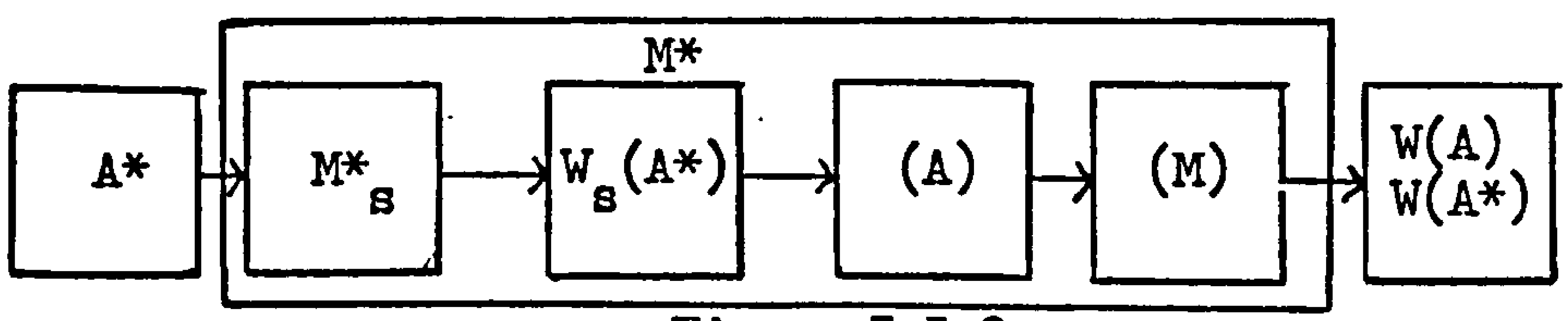


Figure 7.3.2

We may picture this process extended.

7.3.3 If, as in the last paragraph, $W(A^*)$ represents the realization of both $g(A^*)$ and $g(A)$, that is if $W(A^*) \equiv W(A)$, we shall say that A* <u>controls</u> A to produce $W(A^*)$. Moreover, insofar as A is purposeful, we shall suppose that it must

Voluntary acceptance of control

have 'consented' to A*'s control (or 'contracted' to be controlled by A* - cf. for instance Pask (1968), Pask & Scott (1973)). That is A must have entertained the goal of being controlled by A* (with respect to the production of W) as part of its goal specification $g(A)$. In such a case, where A is a purposeful system controlled, as defined here, by another purposeful system A*, we shall say that A is <u>voluntarily</u> controlled by A* if and only if the goal to accept A*'s goals exists as part of A's procedures.

7.3.4 A*'s goal specifications for A may take various forms. Supposing that A voluntarily accepts A*'s control, we distinguish different degrees of con-

Injuction and description

straint on A that A* may exercize, beginning with (1) the most restrictive case in which A* controls A step-by-step, as a starting tape controls a Turing machine. Following Spencer-Brown (p.78) we define such a state of affairs as representing control by <u>injunction</u>. If A proceeds according to A*'s injunctions, A* may direct A to any goal whatever (insofar as A and A* are able to communicate - see 7.4.0), regardless of whether or not A is capable of conceiving of (generating) or even entertaining (distinguishing) such a goal. (This definition of injunction corresponds to what we have up to now been calling description - a starting tape for a machine that makes it generate an end tape. The present departure from this usage aims simply to facilitate the analysis in this section.) Relaxing the conditions of (1) results in (2) the case in which A* fully specifies A's goal, though not how to reach it, and we shall reserve the term 'description' for goal-setting of this kind. Here A must proceed according to a goal description offered by A*, the description itself representing the realization of A*'s sub-goal. We may regard these conditions as providing a <u>problem-solving situation</u> for A and, following Newell (1962) for example, view the problem as a combinatorial one, involving the choice of a sequence of operations out of a set of possible sequences. In this combinatorial case, AW will terminate only when A has chosen a sequence (assuming one exists) that leads to the production of $W(A) \equiv W(A^*)$. Notice that, if A is successful, AW will terminate with the same result, <u>irrespective</u> of the sequence it has followed reaching it. In other words no evidence of A's control or part in the process of producing W(A) inheres in W(A) itself. Whatever A's role has been, it remains concealed; problem solving frequently leaves little or no observable trace to indicate how solutions have been reached, (which is the reason why experimental design in problem-solving psychology so often aims specifically at displaying the normally hidden aspects of problem-solving processes). And the solutions themselves, the end products of the processes - the answer to a complicated calculation; a move on a chess board - mostly provide little or no evidence to an observer of any complex process preceding them. An observer inspecting W will attribute c-authorship to A*. Only an observer ignorant of A*'s instructions to A will attribute A c-authorship.

SECTION 4 COMMUNICATION BETWEEN PURPOSEFUL PROCESSES

7.4.0 Clearly, in practice, there are limits to the kinds of goals that might be wholly specified by description. Specifically these limits depend upon what we have broadly referred to (5.3.0, 5.3.4, 7.3.3) as the 'communication' between A* and A. In the language of information theory, this is what is sometimes called the <u>transinformation</u> between A* and A, or the average mutual information or transmission between A* and A, which Conant shows is 'an essential component of the regulatory process' (p.336). Goals specified by description (in the present sense) depend on a generative process 'within' A, which does not occur where specifications are by injunction. A's generative process is thus subjective to A and may therefore lead to production of the wrong W - in A*'s terms - namely, $W(A) \neq W(A^*)$. In more detail, A* is supposed to have two goals, a goal $g(A^*)$ that when executed leads to the production of $W(A^*)$, and a sub-goal $g_s(A^*)$ leading to the production of $W_s(A^*)$. $W_s(A^*)$ takes the form of a set of instructions to A from which A may derive the specifications $G(A)$ of $g(A)$. Strictly, A*'s sub-goal g_s is to cause A to generate its own g such that its execution will lead to the production of $W(A^*)$. $W_s(A^*)$ is no more the realization of $g_s(A^*)$ than any W is the realization of g. It is simply the best A* can do given its own limitations due to M and additionally its possible ignorance of A's characteristics.

Imperfection of descriptive control

7.4.1 Evidently it is unwarranted in practice to draw too sharp a dividing line between injunction and description. For goal specifications provided by A* to A (in the form of commands or instructions) are <u>all</u> injunctive, though not necessarily complete. To execute A*'s instructions, A must 'interpret' and then act upon them. Only step-by-step instructions that need no interpretation call for no goal-generation by A. This case is trivial since it entails A's voluntarily surrendering all its capacity for c-authorship to A*. Where A*'s control is less than absolute, A's interpretation and subsequent genera-

Autonomy and authorship

tive process consists in interpolating missing computational steps. A's interpretive role increases as A*'s description becomes more fuzzy. In practice therefore the difference between injunctive and descriptive specifications depends upon the degree of autonomy that A* accords to A for realizing the goals specified; and hence A's opportunities or potential to be a c-author. By definition (2.1.2), A*'s goal is to hold invariant some state, or movement towards some state, of E (or part of E). The specifications that A* provides to A stipulate what the state is, either all at once, or by means of continuing direction as A proceeds. A must then maintain the state or determine the path to it, possibly by itself setting up sequences of sub-goals.+

7.4.2 The correspondences respectively between injunction and description and algorithms and heuristics will be clear. Thus for example a theorem-proving procedure might provide proof heuristics based on some proof criterion and a facility for checking its results against the criterion. The criterion might be some general definition of proof, in terms say of well-formed formulae. But the procedure might equally get on without the heuristics, simply by combining axioms say, according to some algorithm for forming wffs, and checking after each step to see if the required proof had been accomplished, that is whether the newly formed wff was the required one. Similarly we might provide a hill-climbing procedure with a general concept of a peak, as say the highest point in a region. Or we could give a particular definition of a particular peak, based upon route directions. Notice that the more general the definition of a concept, the more widely different the procedures that may adopt it as a criterion, while the more narrowly it is defined, the more restricted its applications will be. In the limiting case for generality, A* need do no more than retain a veto over what A produces, thus ensuring that anything

Algorithms and heuristics

+In the following chapter we shall see that A* and A may be characterized respectively as what Ashby (1964) distinguishes as controller and regulator.

finally produced conforms to the realization of g(A*). This is the homeostatic procedure we examined in chapter 4, under the heading algedonic control. All the remarks made in that chapter on the relative advantages and limitations of heuristics and algorithms apply, mutatis mutandis, equally here. Notice that purely injunctive specifications are tied to particular processors. If the processor is a universal Turing machine, the starting tape will need to include instructions for producing the particular machine required to execute the programme.

Injunctions and descriptions: choice in practice

7.4.3 In practice A will sometimesprefer (find it easier) to use descriptive, in others injunctive specifications. To use Spencer-Brown's example, it is easier to describe a cake or a piece of music by means of the injunctions respectively of a recipe or a musical score than in terms of descriptions of taste or sound. Conversely, it might be easier to leave the solution of a theorem to a suitably programmed computer than prove the theorem oneself. But 'the description is dependent upon and secondary to the set of injunctions having been obeyed first' (Spencer-Brown, p.78). Once the cake recipe is known its name will be enough to describe it. In practice numbers of factors will dictate the choice between injunctive and descriptive control procedures. Generally A*'s instructions to A will be a mixture of both. A limiting consideration for A* is its ignorance of A's structure, which in general will differ from its own.+ A* must compensate for its ignorance by the specificity of its instructions. Paradoxically, by the conclusions of the previous paragraph, the more specific the instructions the more closely bound they will be to a particular processor. But the less A* knows of A the greater its uncertainty of its properties as a processor. So there is a point beyond which it does not pay A* to make its instructions more specific. In practice the problem does not present itself simply as a straightforward one of coding, although it reduces to a coding problem if A* incorporates its knowledge of A

+. If A* and A are identical then questions of control between them are equivalent to questions of their own internal control.

into its coding procedures.

SECTION 5 CHAINS OF CONTROL

7.5.0 In general we picture the two-stage process A*W envisaged in 7.3.2, in which A* produces $W_s(A^*)$ to provide A with a basis on which to select a set of specifications $G(A) \subset \mathcal{G}(A)$. To these, according to this formulation, A provides

Co-author additionally, self-generated specifications, to obtain finally the specifications $G'(A)$ of $g(A)$. We define such a case as a special case of partial c-authorship and call A co-author with A* of W, and write W as $W(A^*,A)$. In principle at any rate, the relative contributions of A* and A to W correspond to the degrees of order in W attributable respectively to the two entities.

7.5.1 The two-stage process due to A* and A extends naturally to a multi-stage system, consisting of a sequence of processes $A_i(W)$, $(i = 1, \ldots , n)$, say. Each A_i generates $g(A_i)$ and possibly one or more sub-goals, corresponding to

Extended authorship sub-processes of the sequence. We write $g_{s_j}(A_i)$ to denote the sub-goal of A_i whose execution is intended to control A_j. $W_{s_j}(A_i)$ denotes the product of A_iW when this goal is executed. This W consists of a set of instructions to A_j, which may include both injunctions such that they will lead A_jW to produce $W(A_j)$, or descriptions from which A_j will be able to construct $g(A_j)$ whose execution will lead A_jW to produce $W(A_j)$, or both injunctions and descriptions. As a sub-goal $W_{s_j}(A_i)$ has as its ultimate object either, (1) $W(A_j) \equiv W(A_i)$, if A_jW is the penultimate process or, more generally, (2) $W(A_j)$ should form the basis for A_k say, to set specifications for $G(A_k)$ (where k may but need not equal $(j + 1)$), to make $g(A_k)$ and so on.

7.5.2 A_i intends (in the sense of 7.3.2) that the execution of $g_s(A_i)$ and $g(A_j)$ should lead to the production of the same W. But as we have noted, this may not always occur. If A_i's instructions to A_j are not effective, and A_j is not able to use them as a basis upon which to form a fully specified goal $g(A_j)$ (or sub-goal $g_{s_k}(A_j)$), or if A_j's variety is greater or less than A_i's, or in the presence of noise, A_j may generate a goal different from the goal A_i intended it to generate. That is it may happen that the execution of $g_{s_j}(A_i)$ and $g(A_j)$ lead to the production of $W_{s_j}(A_i) \neq W(A_j)$.+ In general, if $W_{s_j}(A_i) \supset W(A_j)$, then we should suppose that A_j will regard $g(A_j)$ as under specified by A_i. While, if $W_{s_j}(A_i) \subset W(A_j)$, we should suppose A_j's variety or power of resolution to be lower than A_i's. Other coding discrepancies between A_i and A_j, or the presence of noise might give rise to disjoint or overlapping W's.

Goal specification and variety

7.5.3 In general we need not suppose that the process AW will necessarily consist of a simple sequence of nested processes and sub-processes $A_iW(A_i)$ as we have pictured so far and that we might write as

Goal nets

$$AW = A_1(A_2(A_3 \ldots (A_nW(A_n)) \ldots W(A_3))W(A_2))W(A_1).$$

We need not suppose that one process occurs at a time, nor that any given sub-process $A_iW(A_i)$ will occur only once, but we may picture sequences of nested loops of recurring processes in complex arrangements. But other forms of constraint, apart from those already considered, will play a part in determining the structure of such arrangements. In cases where such processes have parallel components they may be reduced to sequential forms in the same way as information processing procedures.

+ W is an object or an event rather than a set of processes. Nevertheless we may describe it by means of a monotonically convergent sequence of sets, exactly as we described goals. (7.1.2)

SECTION 6 DESIGNERS

7.6.0 The controller A_i of AW's sub-processes have so far all been supposed to act voluntarily (7.3.3) under the control of superior processes and in the absence of competing goals within AW, or of their own. But this may not always be the case. The individual A_js may entertain their own goals and sub-goals. Moreover, this may be so whether or not the A_is act voluntarily under the control of superior controllers. More strictly one may imagine qualified voluntariness in the sense of voluntary control being accepted only if it does not require actions that conflict with or prevent the pursuit of an independent goal. Purposeful processes and sub-processes with conflicting or mutually interfering goals, in which control is not voluntary, give rise to problems that are of more than tangential interest to our present object, since we are concerned with possible conflicts between the purposes of our art machine and its controller, and the orders of linked systems on which it may draw to augment its own structure. The voluntarily-controlled sub-processes of such compounds will act in a way that is the converse of that in which conflicting sub-processes act that do not accept voluntary control by superiors. That is such sub-processes will pursue their independent goals only if these goals do not require actions that conflict with the pursuit of the goal of the superior. A sub-process in this case will have to 'accommodate' its own purpose to the superior purposes. Such accommodation will detract from the c-authorship of the sub-process in executing its own independent goal, insofar as it restricts the goal. The superior processes whose goals it has had to accommodate will not be co-authors, according to our definition (7.4.3).

Conflicting goals in linked processes

7.6.1 Ashby (1972) shows the relative importance of constraint on A due to the conflicting and other goals of other purposeful systems. He pictures a designer furnished with a set of instructions which may include instructions from goals differing from A's, but by which A is nevertheless bound. He represents the

Effort required of a designer

situation by a diagram: Figure 7.6.1

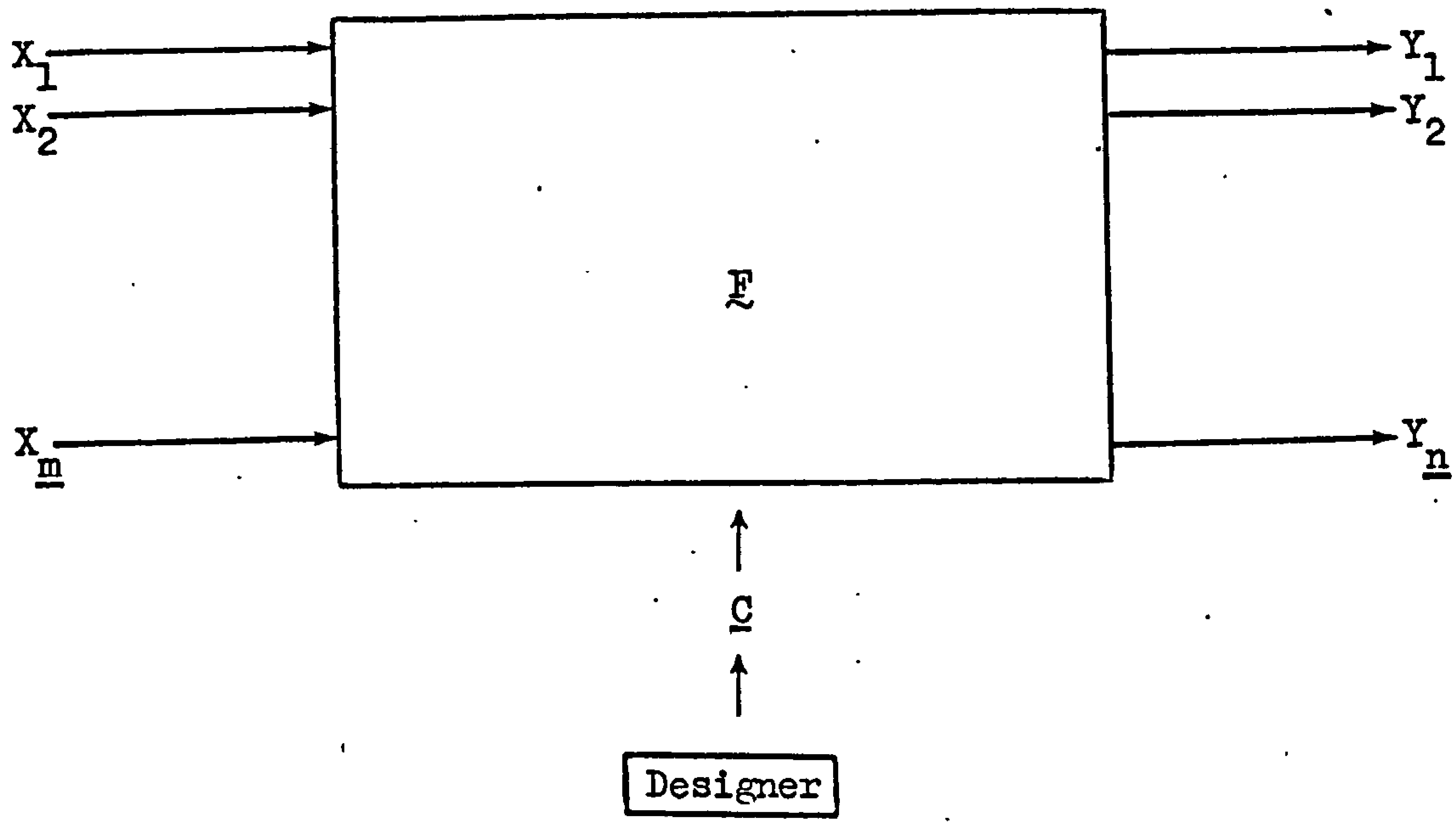


Figure 7.6.1

Ashby asks how much information processing by the designer is implied by his translation of the demands made upon him into completed designs that he produces such as meet these demands. He enquires how the <u>least safe capacity</u> (ibid. p.90) in channel C is related to the capacity in the channel from X to Y, that is what is the minimum capacity of channel C which will ensure that, <u>whatever</u> the actual situation between some given X and Y, C's capacity will be sufficient to enable the translation to take place. The designer receives a message via the channel C in the form of demands X to produce Y. From a set of functions he selects F such that $Y = F(X)$. F is the 'message sent'. Ashby notes that the number of functions from a domain of d elements to a range of r elements is r^d. If the numbers r and d are not small, d will be relatively far more important than r in <u>increasing</u> the number of functions and therefore the work required of the designer.

7.6.2 This acts as a severe constraint on the capacity of any system for c-authorship. For we may generalize this formulation to include in the 'demands on the designer' A, which constitute the constraint, instructions from other purposeful systems, either pursuing the same goals as A or different goals, or

Justification for the definition of c-author

'demands' not necessarily due to other purposeful systems at all, but merely to the structure within which the designer operates. All such structure, whether it is the physical structure of the world, the structure of A's effector system or the structure of a problem domain, detracts from A's capacity to be a c-author. Thus the product of a problem in multiplication is fully determined by the rules of arithmetic, and one would not call a system that performed the multiplication the product's c-author. Evidently the extent to which A is 'free' to generate its own goals is severly restricted whatever form A may be given.

CHAPTER 8

REGULATION AND CONTROL

SECTION 0 THE PURPOSEFUL INDIVIDUAL

8.0.0 Treatment of the subject of control and regulation has deliberately been left to this, the concluding chapter of Part I. The object of the delay has been to show as vividly as possible, not simply the necessity, but the sufficiency of a controlled productive process as a basis for our purpose of making an art machine. From the outset we have insisted on the necessity for AW to be a controlled process, though we have avoided any mention hitherto of how A's control might be exercized. This avoidance has led to certain difficulties and anomalies which we shall shortly be able to remove. The first difficulty was to define purpose without appealing to control. This was an obvious problem though we believe it was solved in a way that did not detract from the utility of the definition adopted: as we observed, control is an inferred aspect of purposefulness, at least for one observing it. In N.3.1 we remarked on problems of interpretation of W, exacerbated by differences between A and O, as in chapters 0 and 1 we drew attention to the related problem of how to know a WOA. Here we shall show that A as a controller is also O, and that the way the productive process proceeds is sufficient to ensure that A produces WOAs if that is his object. Moreover we shall infer that Os, other than A, must undertake a productive process not unlike AW in order to apprehend and appreciate a WOA in the role of audience or critic.

Controlled process necessary and sufficient basis for art machine

8.0.1 The argument of this chapter bases itself on the formulations of Ashby (1964) and Conant. We begin by outlining a paradigm for regulation, which we apply by degrees to the process AW, examining the implications for A's structure at each step. Different varieties of regulation are discussed, in particular

Control of production

cause-controlled and error-controlled forms. The initial paradigm is extended to apply to a purposeful individual, and it is shown, by means of diagrams of immediate effects, that a purposeful individual is homomorphic with a controlled procedure. It is found that a cause-controlled regulator needs a model of the world in which it operates and this implies a need for it to have an order-extraction facility so that it can acquire the model. A modified paradigm for an art machine is suggested and its operation as a productive process observed.

SECTION 1 REGULATION, CONTROL AND PURPOSE

8.1.0 We come at last to control. From the outset (1.0.3) we have defined A as a controller. As a purposeful system, A must be able to control its actions so that they serve its purposes. In 2.1.2 we defined A's purpose in terms of invariance: an observer O that identified some characteristic in A's output which remained invariant over a time interval T, supposed that the characteristic represented the realization by A of a goal entertained by A over the interval. If the characteristic remained invariant under changes in A's input state, A was said to be purposeful. Thus, broadly speaking, A must be able to (1) set the goals it will pursue; (2) 'protect'+ whatever procedures correspond to those set goals from disturbances; (3) execute the goals. The procedures required by (2) and (3) are of a similar kind, but differ from the procedures required by (1). We shall show that these two kinds of procedures, those of (2) and (3) and those of (1), correspond respectively to what Ashby (1964) distinguishes as <u>regulation</u> and <u>control</u>. He pictures them by means of a diagram of immediate effects as in figure 8.1.0, in which D represents a source of disturbance, C a controller (goal setter in our terminology), R the regulator, T 'the hard external world, or those internal

Ashby's paradigm for goal-seeking

+The notion of the source of disturbance as an adversary of regulation is due to Conant (p.334).

matters that the would-be regulator has to take for granted' (p.209), and E the set of 'outcomes'. R's job as regulator is to provide inputs for T such as will, when taken together with T's inputs from D, force T to emit E, such that E shows the required characteristics of invariance, - or such that certain values of E remain within some acceptable range.

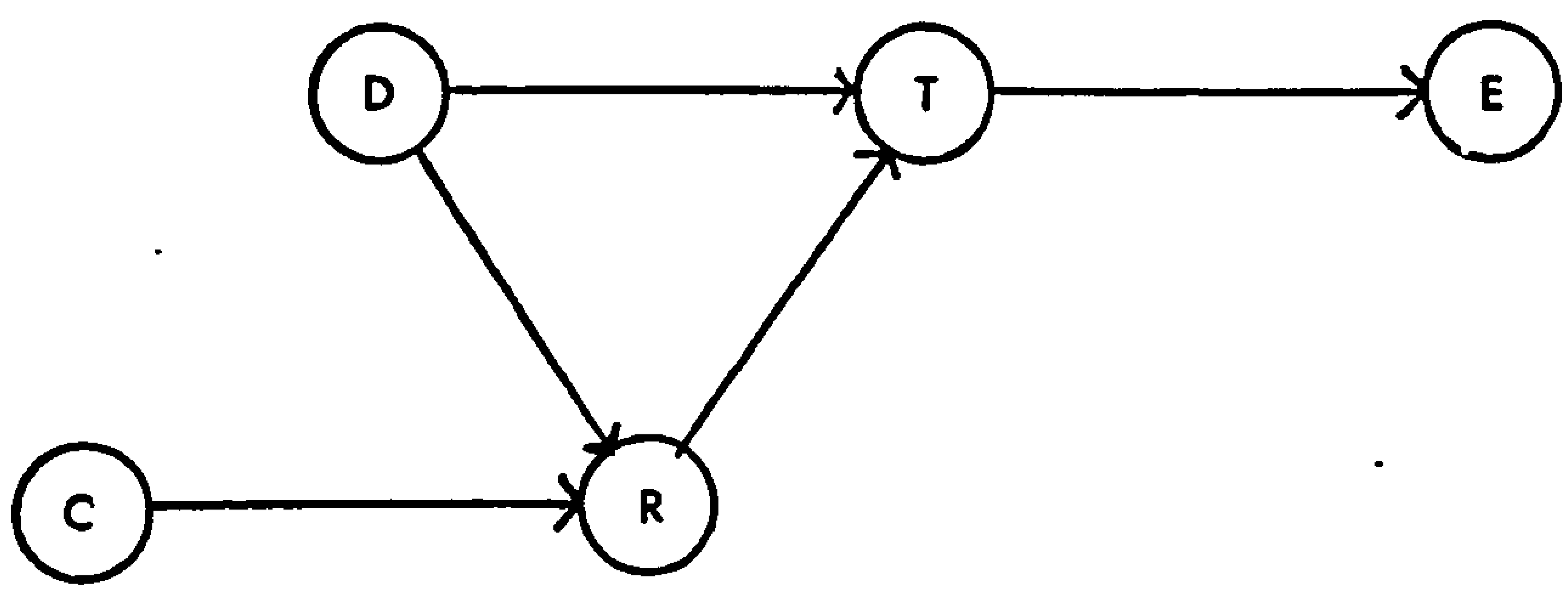


Figure 8.1.0

8.1.1 We might re-arrange Ashby's diagram of immediate effects to represent A, within the terms of our definition, as in figure 8.1.1.0, in which D represents disturbances affecting T.

Representation of a purposeful individual

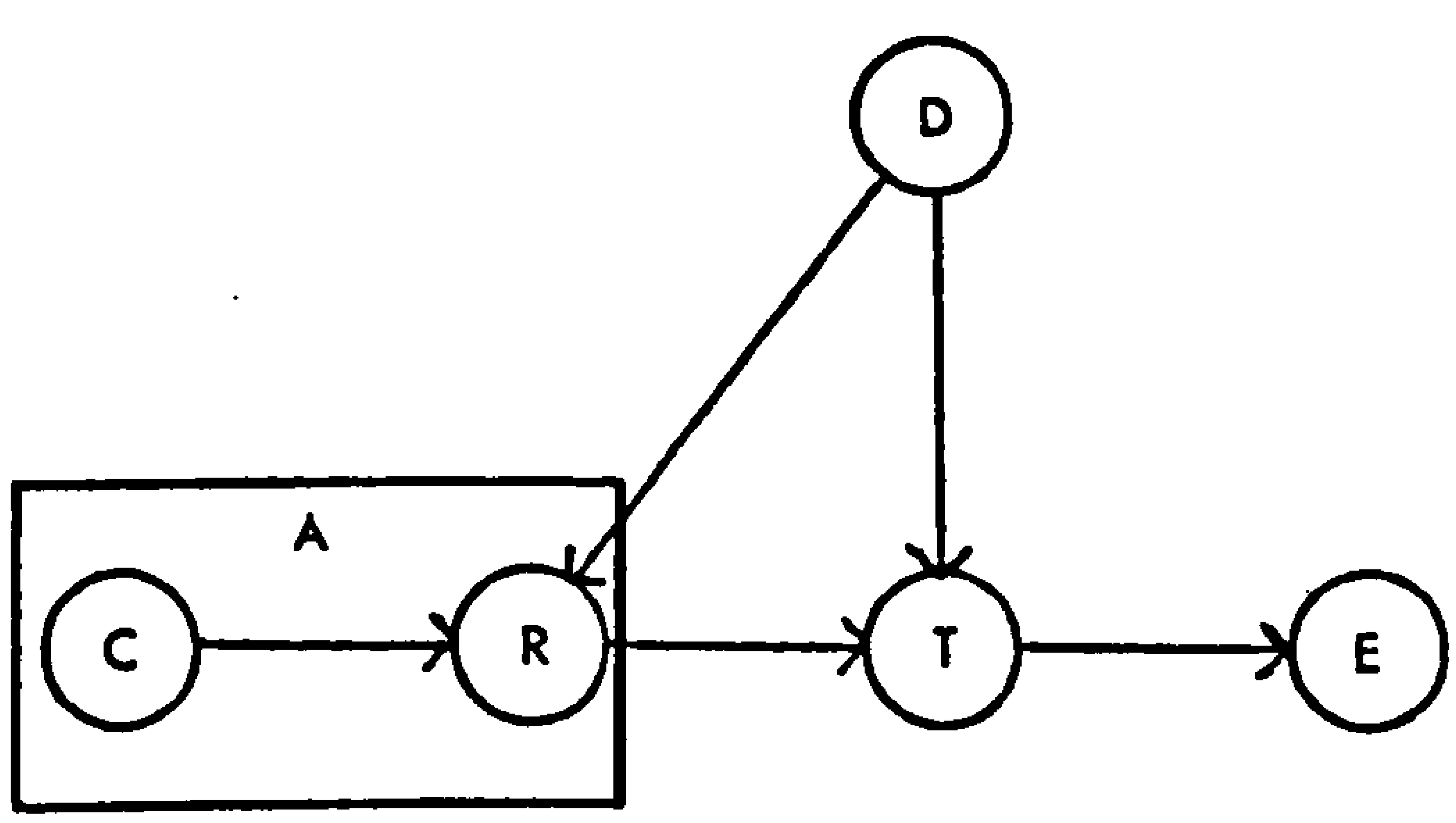


Figure 8.1.1.0

Such a representation would be incomplete, however, as it would provide only the goal-setting and execution systems, C and R respectively, without allowing for a system to protect against disturbance the goal that had been set. In our terms this system would be goal-seeking rather than purposeful. To be purposeful, A must be able to maintain invariant that part of its own internal state that represents its goals. Thus it must include a regulating system as part of itself. We might represent such a system as in figure 8.1.1.1

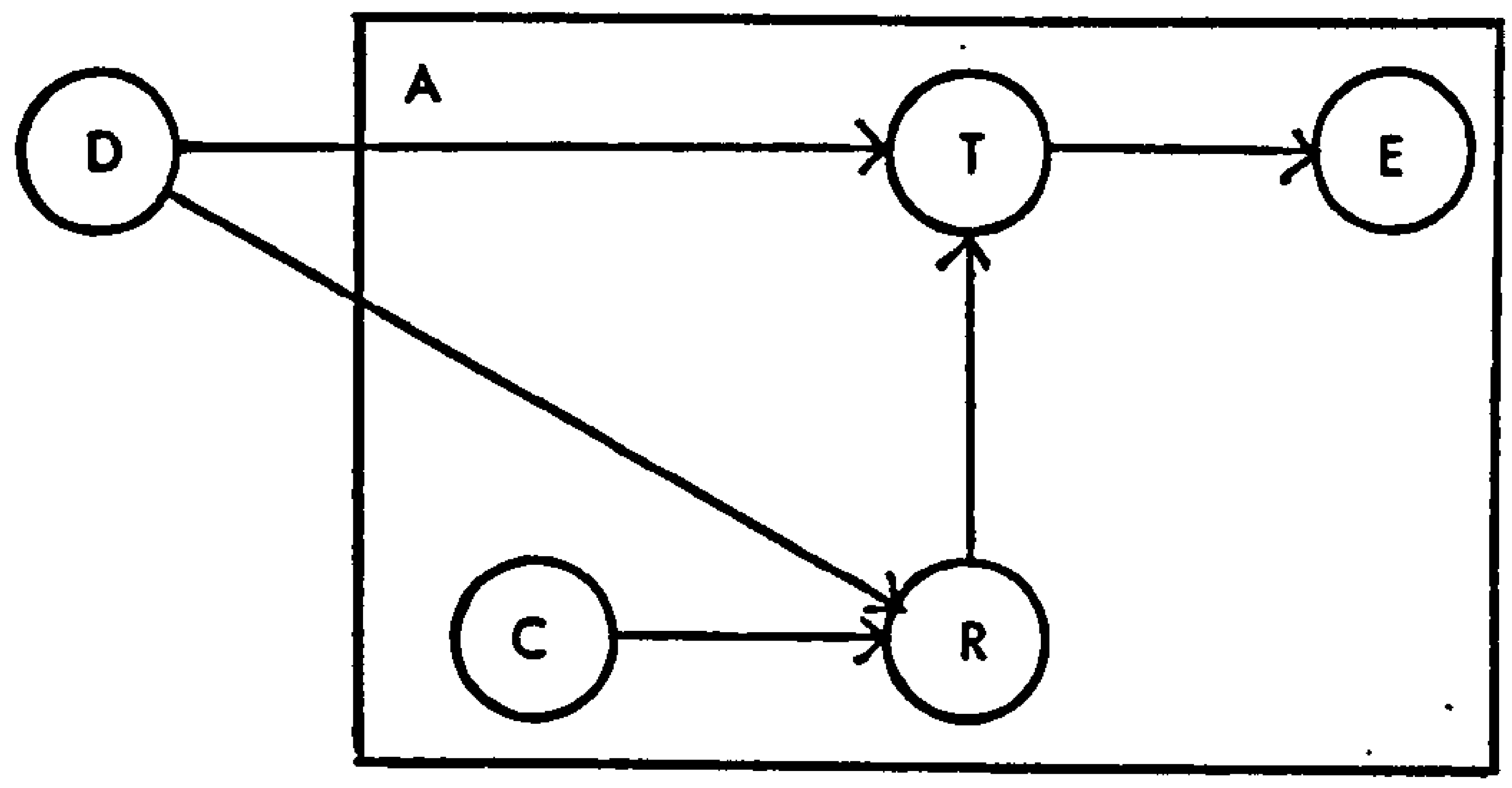


Figure 8.1.1.1

Here D represents inputs that A receives from its surroundings, internal and external; C acts as goal-setter by determining what E R will aim for; T refers specifically to the 'internal matters' mentioned in the previous paragraph and E is the representation of A's goal. Combining the representations of both these diagrams, 8.1.1.0 and 8.1.1.1, we obtain a full representation of A as a purposeful system, arriving at 8.1.1.2

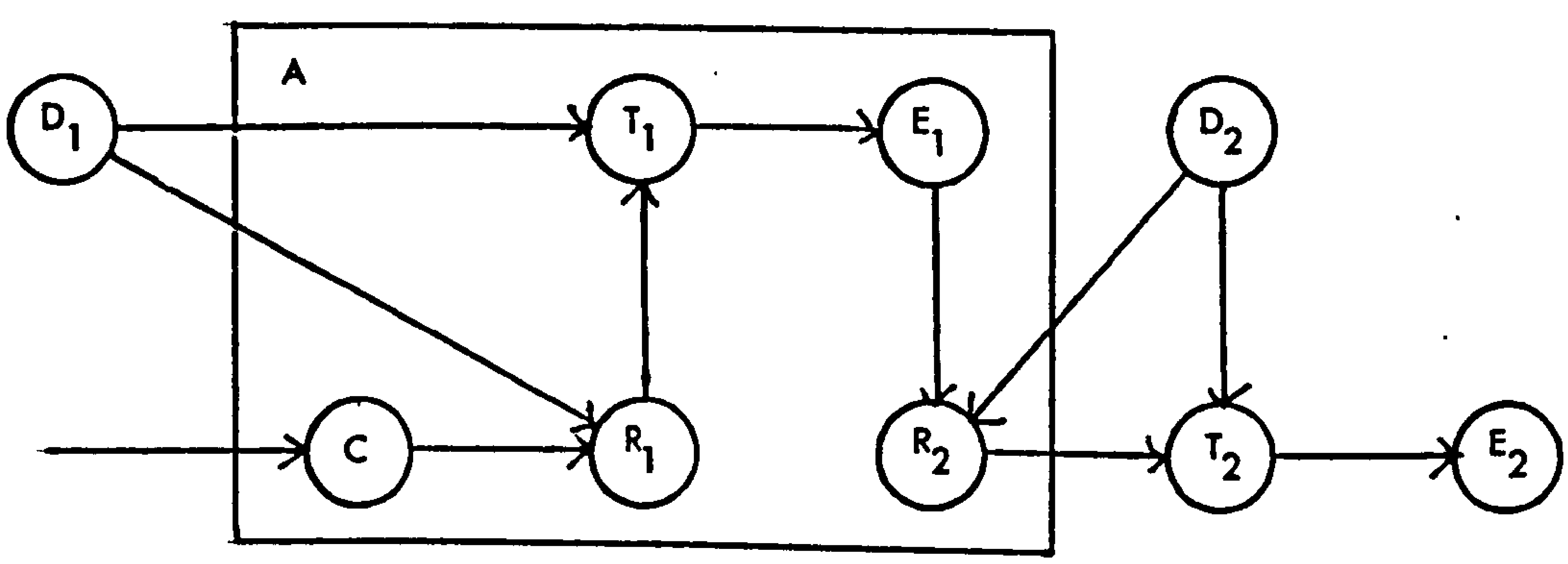


Figure 8.1.1.2

Regrouping the elements of this figure, we get figure 8.1.1.3, which reduces to a homomorph of figure 8.1.0, the figure we started with. That is we have not departed from Ashby's formulation, which we might interpret as follows:

C sets A's goals, with or without specifications received from elsewhere. We might say that C does A's order-extraction work, or is the creative or imaginative element in A; D_1 represents disturbances affecting A; T_1 represents the internal matters that Ashby mentions; E_1 is A's representation of its goal, and plays a role in A's effector system, which corresponds to C's role in A. Notice that C controls R_2. R_1 protects, as we have called it, the goal that C has set for A against the disturbances of D_1; E_1 may now be taken as A's goal representation and R_2 as the regulation of A's effector system against disturbances D_2 that act on it. T_2 represents the external world and E_2 represents W, the realization of A's goal.

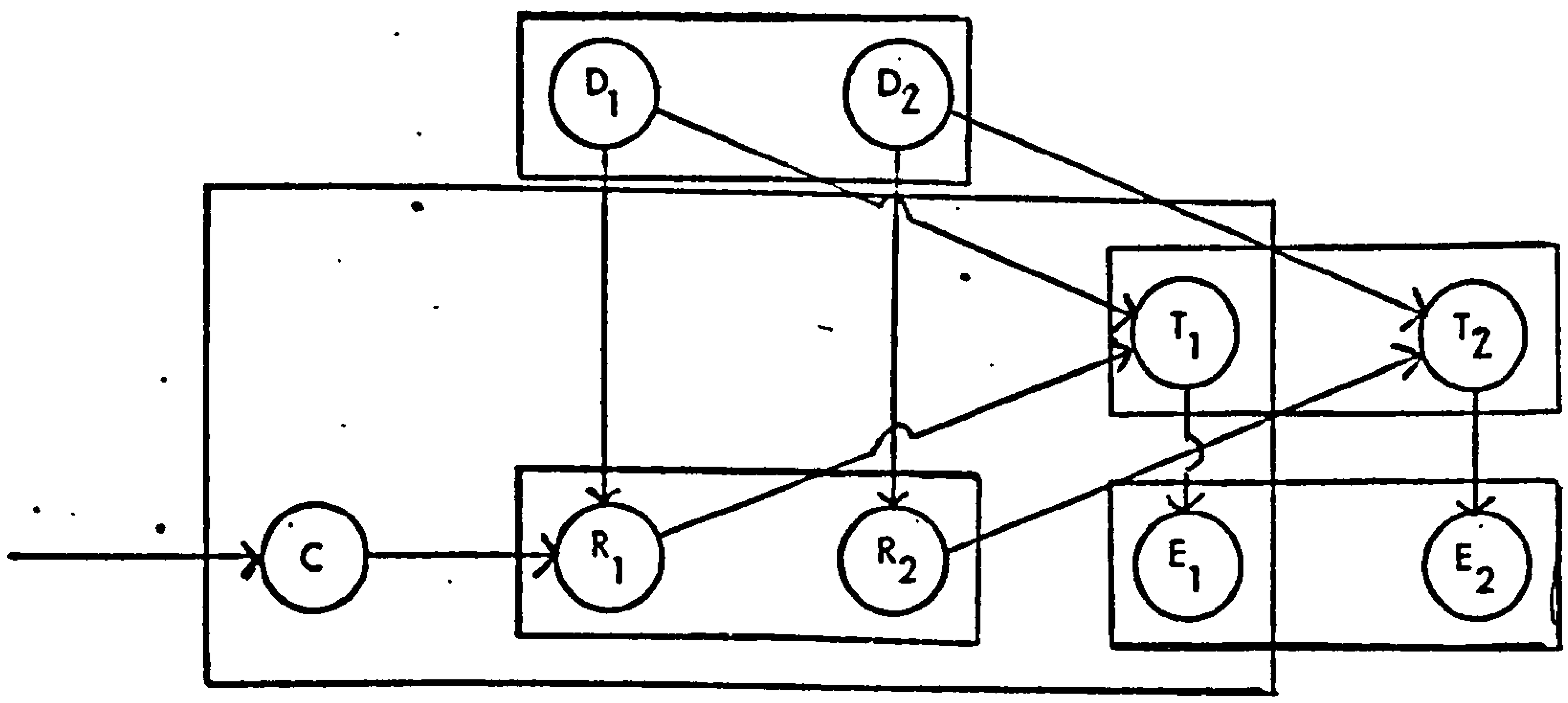


Figure 8.1.1.3

8.1.2 Notice that we might have achieved the requirements of a purposeful system more easily by a simple modification of the original paradigm, figure 8.1.0, by allowing a channel of information between E and C, as in figure 8.1.2

Purposeful system

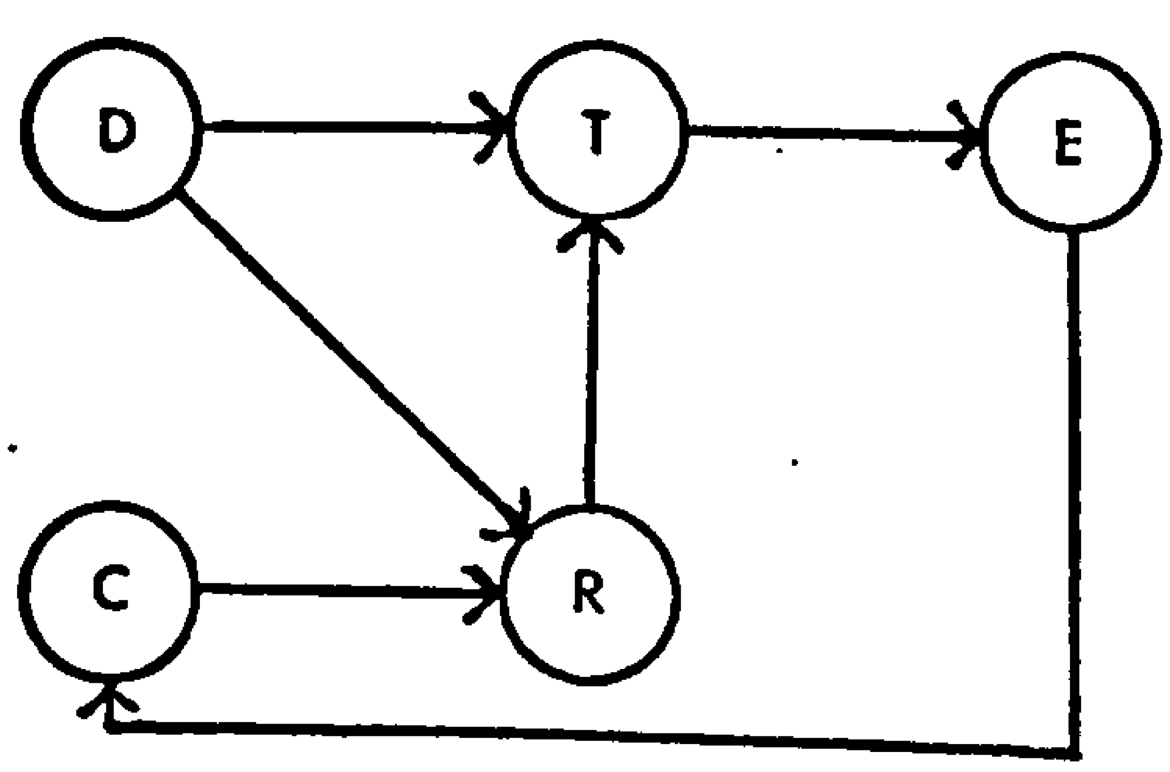


Figure 8.1.2

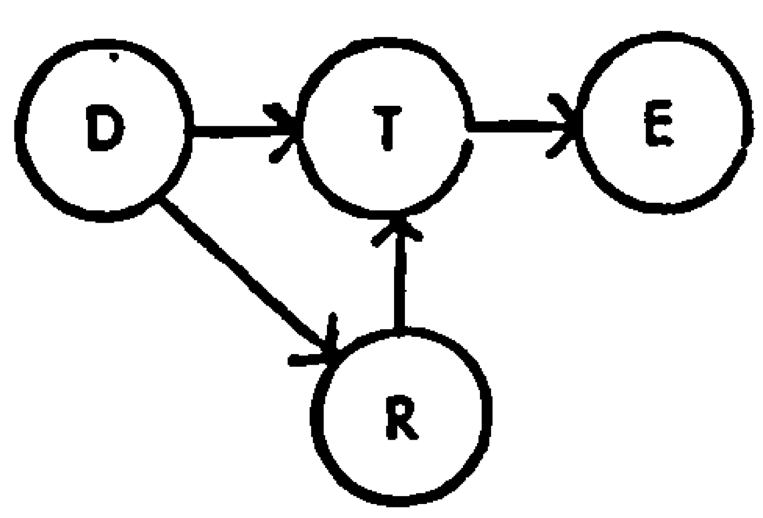
This system, by reviewing the success or failure of the regulation procedure it has determined may see to it that E remains appropriate. It is an ultra-stable system, in Ashby's (1965) sense, with C having the role of altering the parameters of the transducer R. If C operated on a step-function, this system would resemble Ashby's homeostat. However we have chosen to portray purposefulness in a form that emphasizes the complexity of the internal structure of the system, and to illustrate the symmetry between the regulating mechanisms directed respectively at the internal world of the system and the external world which is the system's primary concern.

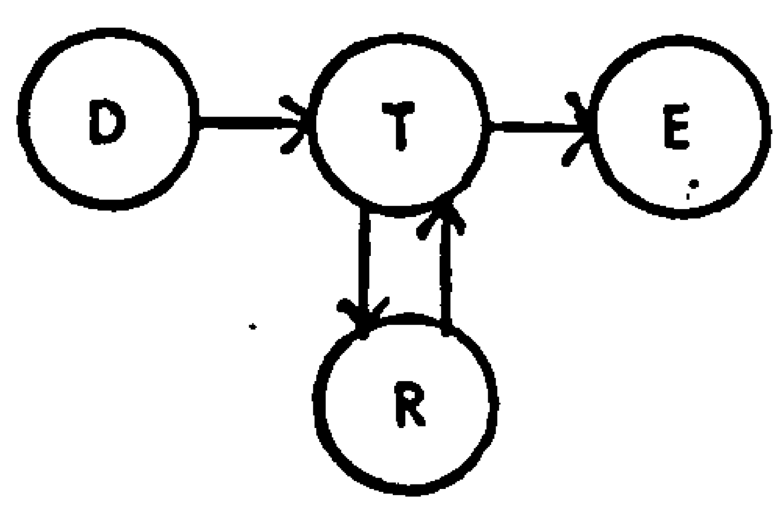
SECTION 2 TYPES OF REGULATION

8.2.0 Ashby distinguishes three varieties of regulators, those characterized by the diagram of immediate effects of 8.2.0.0, which correspond to the regulator part of the system of figure 8.1.1.3, and those systems that may be charac-

Information from D to R

terized as in figures 8.2.0.1 and 8.2.0.2, in which the information available to R concerning the disturbance D that R is required to block, is forced to take longer routes to R, passing respectively either through T or through both T and E.





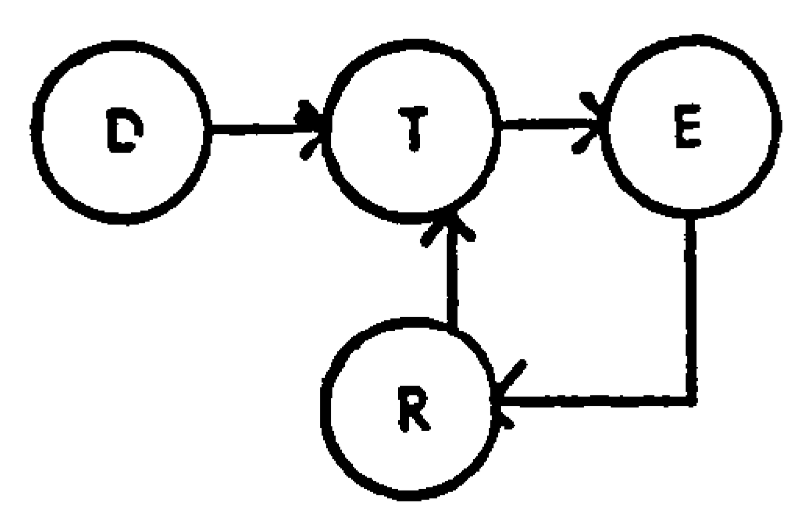


Figure 8.2.0.0 Figure 8.2.0.1 Figure 8.2.0.2

The last type of system (figure 8.2.0.2) has the 'basic form of the "error-controlled servo-mechanism" or "closed loop regulator", with its well-known feedback from E to R' (Ashby, 1964, p.223). The error-controlled regulator 'differs from that of the basic formulation [figure 8.2.0.0] only in that the information about D gets to R by the longer route $\boxed{D} \to \boxed{T} \to \boxed{E} \to \boxed{R}$ ' (p.223).

8.2.1 Conant gives the name 'cause-controller regulators' to regulators that receive their information about D by the shortest route, directly from D. Figure 8.2.1 portrays Conant's representation of cause-controlled regulation.

Cause-controlled regulation

R utilises information from the source that affects S to determine appropriate regulatory action aimed at preventing S from affecting Z. D is a source of 'primary disturbances' (p.338), a sequence of activity that affects both S and R. Without R's regulation, D's activity would be transmitted through the channel S to Z. R must co-ordinate its own actions with S's so that the resultant outcome of their joint activity is invariant. That is so that the channel from D to Z which R and S form jointly has 'minimal information capacity' (p.338).

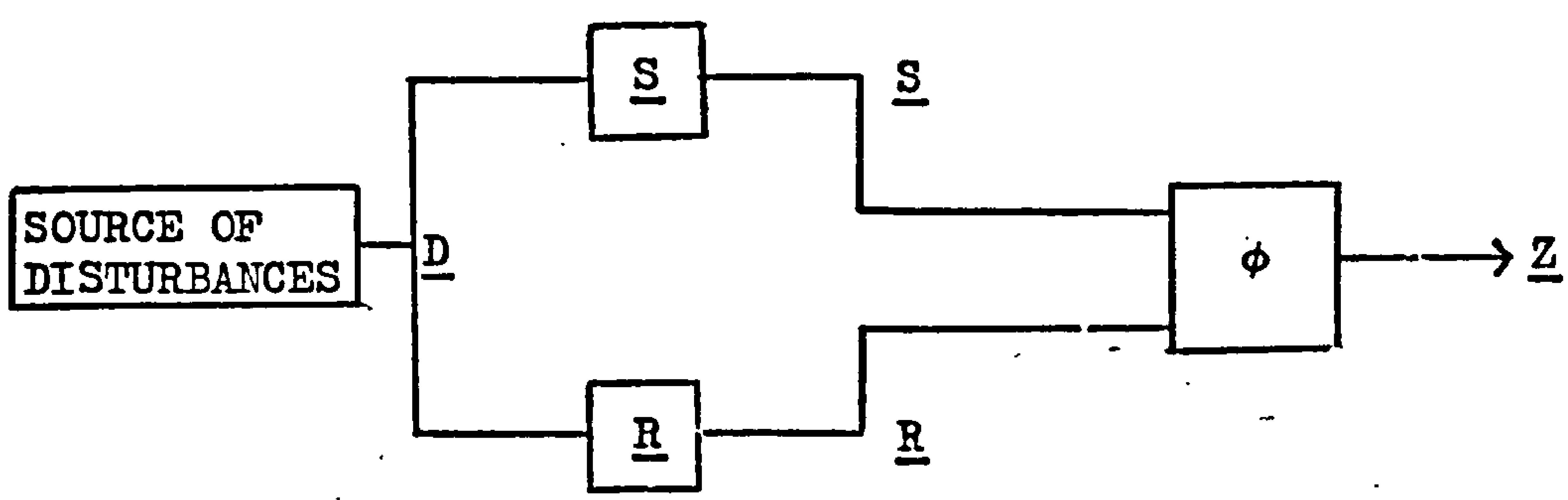


Figure 8.2.1

Cause-controlled Regulation (Reproduced from Conant, p.338)

Conant goes on to show that the amount of regulation achieved by a cause-controlled regulator will be maximised if R is or behaves 'isomorphically or perhaps homomorphically with respect to S', in which case R will be a perfect regulator. This constitutes 'information theoretic grounds for the observation that regulators are often isomorphs of what they are regulating' (cf. 2.7.2). Error-controlled regulators receive all their information about the errors they are required to regulate from the already-regulated signal. They therefore have more the role of what we might call error-attenuators, since their action relies on the survival of at least some error for its direction. Cause-controlled regulators are able to avoid the defect of being theoretically never quite able to eliminate error, by relying purely on information coming directly from the source of disturbance, without the intercession of any external agent. Notice that a model of D alone does not provide R with sufficient information for its regulatory task. This is because R forms a joint channel with S and it is therefore S's outputs that R must 'neutralize'. Conant's formulation is equivalent to Ashby's of figure 8.2.0.0, with Conant's S corresponding to Ashby's T and Conant's Z to Ashby's E.

8.2.2 For perfect regulation, R's and S's responses to D must reach their point of summation - φ in figure 8.2.1 - simultaneously. That is, the time it takes for the information about the disturbance to reach the regulator and the regulator's output to reach the point of summation must be the same as the time required for the disturbance to reach the point of summation, via the channel S.

Times of regulatory responses

8.2.3 Furnished with a model of S, R might determine its outputs in some quite simple and direct way, such as a coding procedure based on the model's outputs, or it might use its model as the basis for a surrogate error-controlled process, in which it pre-tested actions on the model, modified them, re-tested them and so on, before applying them in the form of an output. Observation might not always make it obvious which of these forms R was using. The reason for this may more easily be understood from an example. We recall the machine of Fogel & others (2.7.2) that was evolved in such a way that it pre-

Types of cause-controlled regulation

dicted the output states of its environment. These outputs consisted of a cyclically repeated string of randomly mixed zeros and ones. If instead the machine had been evolved to output ones simultaneously with the environment's zeros and zeros with its ones, the net output obtained by summation of its own outputs and the environment's would have been one invariantly. Such a machine, in the role of R, would perform perfectly in its environment S, againt disturbances D that produced the environment's emissions. If the regulator's environment consisted of more than one pattern of zeroes and ones, a more complex regulator would be called for, specifically one that included an 'anti-pattern' to neutralize each pattern the environment emitted, and in addition a device for selecting the appropriate 'anti-pattern' to use at a given instant. Such a device would need to incorporate some means of recognizing what pattern S was emitting and of selecting an appropriate 'anti-pattern' from whatever stock of them it might have. Put slightly differently, there might be a number of distinct Ss of which R had models, and R would therefore need a way of knowing which one was operating at a given moment. Clearly there is room for R to conduct tests using its models of S. What these tests achieve will depend upon the rate at which testing can be carried out on the model, relative to the time scale of the environment in which regulation is required. At worst this will reduce the likelihood of disasterous actions on the part of the regulator, more or less regardless of the relative time scales of S and R's model of it. At best it may promote relatively effective actions of R. Where regulation is not error-controlled, in the strict sense that the actual commission of error is vital for the regulator if it is not to block its only source of information about the disturbance it is to regulate, it might more properly be called regulation by error-anticipation. But error-anticipation must rely upon the availability of some kind of model of the environment, within which disturbance is to be regulated, so that regulatory actions may be tested, modified, re-tested and so on, before they are used; or which, through isomorphism or homomorphism may permit simultaneous regulation by some kind of matched reaction.

'That very intricate regulator, the brain, is successful because it contains, in some sense, a model of the environment which is an approximate isomorph or homomorph of that environment; on the basis of that model, the brain apparently calculates how S will respond to an observed disturbance and then formulates its own regulatory response which is coordinated with S.' (Conant, p.338)

SECTION 3 AW IN OPERATION

8.3.0 At last we are in a position to portray the productive process AW as it might operate to produce W, and to examine its goal-generating more closely. To permit goal-generation of sufficient interest to justify calling it creative,

Paradigm for AW

we shall suppose that A is provided with a facility for modelling its goals, and since these will all be concerned with producing W in E - by definition - we shall assume that the modelling facility extends also to E or, in other words, that A can model W-in-E. This implies that A must also incorporate an order-extraction facility to enable it to furnish its models. We propose the paradigm portrayed in figure 8.3.0.

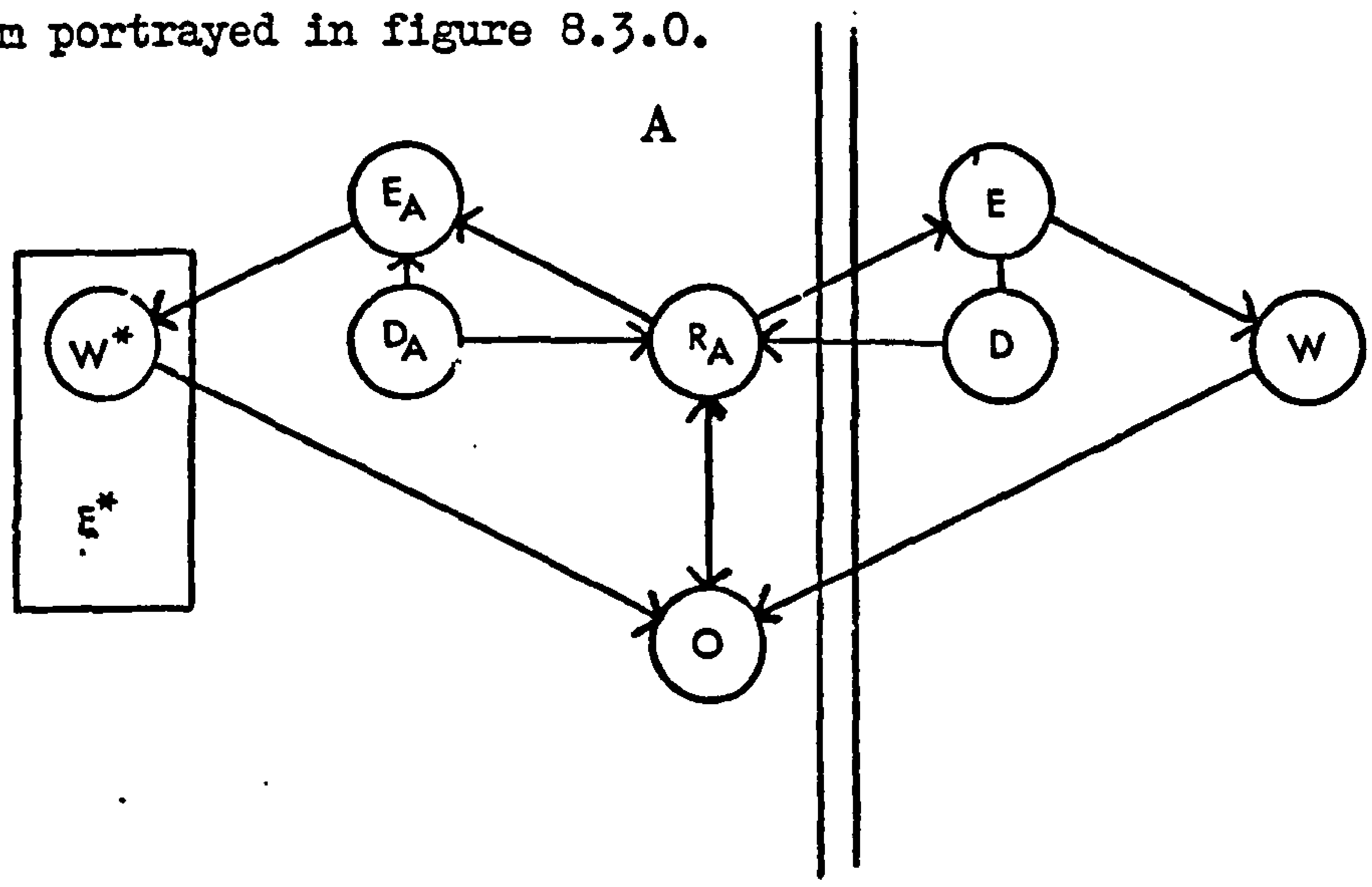


Figure 8.3.0

A disturbance D emanating from A's external environment E affects A - sets up a disequilibrium say. This is registered at O where there is a modelling facility to set up a goal for restoring equilibrium, possibly with test and modification procedures carried out on the model of the environment E*. Call the goal W*-in-E*, or simply W*. O sets R_A to regulate against internal and external disturbances, D and D_A respectively; and within A's internal environment E_A, to produce W*. R_A also acts via M - omitted from the figure for simplicity - to produce W symmetrically with W*. The process ends with the restoration of A'S equilibrium. The W that exists at that time is the realization of O's last goal. We mention O's 'last' goal, because O has been depicted as receiving information from both W* and W. The information channel from W to O makes the system purposeful in the sense of 8.1.2; what, by extension from the definitions of 8.2.0 we should call an error-controlled purposeful system. The information channel from W* to O provides cause-controlled regulation of the system's controller, so that we should describe it as a cause-controlled purposeful system too. Here is the justification for the paradigm of purposefulness portrayed in figure 8.1.1.3, which intimates the place and importance of a modelling facility.

8.3.1 Realizing a goal

W and W* will be subject to differences of variety and of conformability in their structures. Returning to the notation of the previous chapter, let us suppose A's goal g has in A the internal representation W*, provided by the set of specifications $\cap G$. W*'s variety may be greater than, less than or the same as the variety available for its realization, W; or its structure may be unattainable in reality, a difficulty that might easily come about through manipulation of E* not possible for E, or through a bad fit between E and E* in the first place. Suppose A commences its operation with an initial goal g_1, which AW begins to realize as W_1. In the absence of obstacles, AW continues to the completion of W_1. If, for one of the reasons suggested, AW either halts or detects discrepancies between W* and W, A will need either to find a way of removing the discrepancies or will be forced to alter or abandon g_1. If A is unable to avoid the discrepancies and wishes the process AW to continue - because the state of disequilibrium that gave rise to the process in the first place persists - then A will have to alter g_1 to g_2, causing M to begin executing g_2 in E as W_2 (either by modifying W_1 as it has come into existence so far, or by beginning again). Recurrence of these events will result in two sequences, g_i and W_i, $(i = 1, \ldots, n)$ say, of goals and their partial realizations respectively. AW halts at $n = k$, if W_k realizes g_k.

8.3.2 For this to occur, the sequences g_i and W_i must converge towards equilibrium points g_k and W_k respectively, and successive g_is must find on the whole closer approximations to their respective realizations in successive W_is. We picture two sequences g_i and W_i $(i = 1, \ldots, k)$ converging to the point at which W_k realizes g_k. In general we may relax the condition that makes A set out with a fully specified goal g, and assume instead that A's goal is under-specified, call it $\bar{G} = \cap G$, where $\bar{G}$ has more than a single member. Successive $\bar{G}$s will form what Wittgenstein calls a 'family resemblance group' (0.3.1.2). Similarly successive Ws.+

The development of W

8.3.3 The time scale for goal-setting procedures is not necessarily the same as for the procedures for realizing goals: goal-setting procedures operate according to the time scale appropriate for measuring 'events' in A, call it T; realization procedures, 'events' in E, call it t. That is, T instants in A correspond to t in E. The rate at which A can respond to information reaching it from E, has an upper bound determined by T, while the rate at which E can respond to operations on it, controlled by A, has an upper bound determined by t. If T is faster than t information from E* regarding W* may cause G_1 to alter to G_2 before G_1 can be realized, that is, before W_1 can be produced in E. It follows that to produce W in E it is necessary for G to remain fixed for at least some minimum time. This formulation suggests that AW will always terminate with W as the realization of g, or at any rate G. But this is obviously not the case. Apart from those goals that A will abandon,

Time scales

+ In practice W is an on-going enterprize. By 'successive' Ws we mean Ws delineated by changes in G - or g.

it will have to show tolerance in the cases of those with which it persists. A perfect match between W^* and W is unlikely ever to occur.

PART II AUTHORSHIP'S PRODUCTS

Part I set out to discover what conditions were necessary for authorship and more particularly for constructing a machine that would be able to act as an author. What emerged defined a category, including entities of every description displaying the characteristics of intelligence, purposefulness, and imagination or creativeness (the ability to generate their own goals). Moreover it appeared that if the products of any entity of this character were to have meaning or significance for humans, the entity's intelligence and so forth would have to make sense of the world in the same kind of way that humans do. Showing this might alone be considered useful, because it makes explicit what is frequently overlooked or neglected or simply taken for granted in the theory of art - and elsewhere - namely on the one hand, the close relationship between the various capacities that the production of art demands, and on the other the limits that confine innovatory power. The latter conclusion appears of particular interest for art theory, which sometimes tends to endow such faculties as imagination and creativeness with powers transcending such mundane matters as natural laws. Nothing said so far however suggests in what way, if any, an art machine might <u>differ</u> from other machines for authorship. Part II sets out to deal with this question by concentrating on W, with the object of discovering differences distinguishing WOAs within the wider class. For knowing these differences will indicate differences between an art machine and the wider class of author machines, or more simply between artists and a wider category of 'producers'. As previously, the objective is to arrive at plausible and useful conclusions - the use of plausibility as a criterion of truth is itself vindicated by Part I - in the least restrictive way we have been able to find. We therefore proceed as before, from abstract considerations, without appealing to any particular actual state of affairs. The argument begins by exploring W's value for A. From here it proceeds, establishing descriptive dimensions for W, which it finally uses to arrive at a definition of WOA and specific suggestions for an art machine, followed by a brief assessment of the future of the project of constructing one.

CHAPTER 9

OBJECTIVE KNOWLEDGE

SECTION 0 THE UTILITY OF W

9.0.0 The task Part I set itself was to determine the principles of authorship, which we defined in terms of a productive process AW. We deliberately avoided restricting W, explicitly allowing it to take the form, either of a performance or an object, as long as it left some trace - possibly only fleeting - in its surrounding, observable by some observer O. The reasons for insisting on such an unrestricted W will emerge more clearly in this chapter. Although our concern has been an artificial system, it is obviously the case that the only real producers, where the term in the sense that we use it connotes any acceptable degree of richness, are natural systems. Similarly, although we have discussed purposeful activity of all kinds, our interest is particularly in those purposeful performances that result in fairly permanent Ws that are objects, especially where these are of no obviously immediate utility. We are naturally led to ask why such Ws should be produced and what part they may play in furthering the - higher level - goals of the As that may produce them. This chapter enquires into the nature of Ws of this and other kinds and examines what use they may be to A.

Practical interest in richness

9.0.1 We begin by noting the characteristics of a W that satisfies Popper's requirements for 'objective knowledge'. Objective knowledge is not necessarily explicit nor does explicitness assure objectivity: empirical systems can never be fully explicit. Effective procedures are examined. If these are objective it is to an observer with a view of the whole procedure, not merely its step-by-step execution. Conscious and unconscious criticism are compared. Both conjectures and refutations by natural systems are likely to include unconscious elements; what differs between the two kinds of criticism is not so much A's procedure, as the Ws involved. Conjecture and refutation call for objectivity, but this may be conscious or unconscious. In either case A must have the structure of an author. The notion of 'unconscious' processes vanishes if we provide A with a stratified object language.

The meaning of objectivity

SECTION 1 W AS OBJECTIVE KNOWLEDGE

9.1.0 Recapitulating, we have concluded so far that an art machine must be an author, in particular a c-author, for which, we have said, it needs to be purposeful and in particular able to generate its own purposes; for which we showed it required an order-extracting capacity and a reflexive structure, to control its production. Furnished with these capabilities, a machine might produce Ws. Such Ws, in the machine's, A's say, intention would be partial self-reproductions of A in A's world E; in realization, they would be subject to constraints due to the structure of E - including the effector system A used in their execution - and the characteristics of A's control of E. The relationship of A to W, as we showed (8.2.3) is partly that of observer to object, where an object in this case is a

A's characteristics: recapitulation

distinguishable part of E, namely W. An immediate questions is, What will be the general nature of the Ws A produces?

9.1.1 As a first approximation, because our earlier conclusions suggest its possible suitability, we might characterize W as what Popper (1972) calls 'objective knowledge'. The idea is plausible because W is 'objective' relative to A, enjoying as it does a separate existence in E, where A may observe it. What are the characteristics of objective knowledge?

Objective knowledge

> 'Only objective knowledge is criticizable: subjective knowledge becomes criticizable only when it becomes objective. And it becomes objective when we _say_ what we think; and even more so when we _write_ it down, or _print_ it '. (p.25)

This description conceals certain difficulties. In the first place, unless we place severe restrictions on the meaning of Popper's phrase 'what we think', we shall be forced to allow objective status to any W that an A may produce. This is first because, as we have mentioned, W is certainly objective _vis-a-vis_ the entity that produced it, being an entity distinct from its producer; and secondly because, as a product of order-extraction work, - A's either by acquisition or 'programming' - it qualifies as the objective part of 'what we think', where 'we' refers to the producer, 'what' to W and 'think' to the process of authorship or c-authorship. In the second place there is a difficulty over the notion of the possibility of _degrees_ of objectivity that Popper implies, writing down what we think being in some sense more objective than saying it (cf N.8.1). It is not immediately obvious whether, in Popper's terms, it is necessary to say 'out loud' what we think, rather than say it 'to ourselves'. But as we may probably take all these qualifications to refer to practice rather than principle, we conclude that, on the face of it at any rate, there appears no reason in principle why an idea said out loud should be more objective than one said to oneself, given that both ideas are held with the same clarity. Presumably it is precisely the clarity that the practical expedients of print and speech are supposed to assist.

SECTION 2 OBJECTIVITY AND EXPLICITNESS

9.2.0 In addition to differing in objectivity, what we think - the more so if it is not said out loud but only to ourselves - may vary in explicitness,

Explicitness

from the degree of explicitness associated with formal systems, to what may be almost wholly inexplicit, such as the kind of heuristic model we may have of the world (cf 0.2.2). And even though such models might be regarded - following Popper (1972 and elsewhere) - as conjectures about the world that are subject to refutation, this in itself would not be enough to ensure their explicitness, or even to make us fully 'aware' of them, since many such models, although they are 'used', are never even consciously adopted, let alone made explicit. Explicitness does not imply objectivity, nor does objectivity necessarily lead to explicitness. In saying what we think, our statements may still carry implications that are quite likely to receive subjective interpretations until they are made explicit. For example though the axioms and inference rules of a formal theory imply all the true theorems of the theory, this alone does not tell us what the theorems will be. Barring subjective intuitions, one would not know what they were until they had been explicitly proved. Similarly the end tape generated from a given starting tape by a given Turing machine is no more than implicit or immanent in the starting tape and the Turing machine, until the mechanical procedures of generating it have been completed. 'Arithmetical propositions are ... always synthetical of which we may become more clearly convinced by trying large numbers.' - (Kant, 1934, p.33) (cf. Turing 7.2.3). The explicitness of the formal theory or the starting tape of the Turing machine lies in the furnishing of a procedure which, if mechanically followed, will lead to a particular result, though the result itself will be synthetic and may not be reached 'by mere analysis of our conceptions' (ibid. p.33). Saying what we think does not in itself necessarily make our thoughts explicit. It is a practical expedient that facilitates aligning the thoughts with the real world to which they refer, helping us to test them against this reality. But the testing itself will not necessarily make the thoughts any more explicit for whomever entertains them explicitly in the first place.

9.2.1 In actuality, explicitness may seldom be taken as an absolute. Like objectivity, what we think will probably possess it in varying 'degrees'. In general, implications will unavoidably underlie most explicit statements (Löfgren's unexplained explanans (2.3.2)). Only in a closed, formal system might one be certain of making all implications - the synthetic propositions immanent in the system's axioms and inference rules - explicit, and this only in systems in which it were possible to prove completeness. In complex, 'open', empirical systems, each new explicit statement is likely, in varying degrees, to give rise to new implications. Nearly all empirical propositions contain a mixture of explicit and implicit elements (the following chapter develops this view further). As an example we think of a partially-ordered system - a hierarchy or some kind of network say - representing a domain of possible propositions. Certain nodes, representing concepts say, are explicitly identified, thereby implying the identification of certain other nodes, namely those that lie on the various 'paths' that lead to the explicitly-identified nodes. The possibility of alternative paths interprets the notion of subjectiveness. For, if <u>A</u> and <u>B</u> use respectively paths <u>a</u> and

Degrees of explicitness

b to get to the same concept (node) W, they will each identify different sets of (implied) nodes on the way. So W will imply different things for A and B. (The operation of PLANNER provides a concrete example: two PLANNER procedures with different 'experience' might appeal to different back-up theorems in executing a goal.) In general in an open system, each new explicit idea will carry new implications, unless there is only one way that the idea may be inferred.

9.2.2 Summarizing these arguments, we should say that (1) every W is objective, in the sense that it is distinct from its producer A. Thus A is like Einstein in Popper's (1972) formulation:

> '... the main difference between Einstein and an amoeba ... is that Einstein consciously seeks for error elimination. He tries to kill his theories: he is consciously critical of his theories which, for this reason, he tries to formulate sharply rather than vaguely. But the amoeba cannot be critical vis-a vis its expectations or hypotheses; it cannot be critical because it cannot face its hypotheses: they are part of it'. (p.25)

Continuum of objectivity and explicitness

A can certainly 'face' W. (2) Objectivity, in the sense merely of something said or written down, is neither a necessary nor a sufficient condition for explicitness; and (3) while full explicitness may be possible, in general in empirical systems many explicit 'ideas' are likely to contain implications which, when made explicit, will contain further implications, and so on, in an explosively divergent way. Thus all Ws will be objective in the sense of (1), and in general will vary as to their degrees of explicitness. What distinguishes Einstein from the amoeba is his capacity for authorship, the ability to produce Ws. In the terms of the previous chapter, what the amoeba lacks is a world model distinct from its own overall structure. But the differences are not absolute. The two characters of Popper's distinction are no more than two markers on a continuum that we might trace from the explicit, complete, consistent, formal theory on the one hand, to the functioning of the amoeba, or of a much simpler entity such as the machine of Fogel & others (2.7.2), on the other. Einstein's reflective capacities, his ability to 'face' his hypotheses, are due to - and demand - a structure of a kind such as we have represented in figure 8.2.3, and shown to embody certain minimal necessary requirements for authorship.

SECTION 3 W AND EFFECTIVE PROCEDURES

9.3.0 Apart from these questions of principle are practical questions we have already mentioned concerning the heuristic value of objectifying a theory and the reasons why what is written down is somehow more objective than what is simply said or thought. In part the practical advantages of objectifying knowledge concern the peculiarities of the working of the human machinery and therefore fall broadly within the domain of psychology. To a large extent these advantages are the rewards that Descartes looked for in his 'method':

Objectivity's value and user's structure

> '... never to accept anything for true which I did not clearly know to be such; ... to divide each of the difficulties under examination into as many parts as possible, and as might be necessary for its adequate solution. ... to conduct my thoughts in such an order that, by commencing with objects the simplest and easiest to know, I might ascend little by little, and, as it were, step by step, to the knowledge of the more complex; ... in every case to make enumerations so complete, and reviews so general, that I might be assured that nothing was omitted.' (Descartes, 1949, p.16)

A procedure that satisfied these rules might provide an adequate starting formulation for a definition of a formal system. But, as we noted (9.1.1), nothing in principle demands that we literally say out loud or write down what we wish to objectify, nor a fortiori that some non-human entity do so. All that is needed is for the entity in question to have a structure of of the kind suggested in 8.2.3, namely one that allows the separation of two distinct functions, those respectively of producer and observer of what is produced. But notice that it is more than a quirk of the machinery of the human mind or psychology that the observer's function is facilitated by the heuristic aid of paper and pencil. Rather the reason is inherent in the structure of any entity that utilizes a model of its surrounding. For A's model E* of its environment E shares with E only such features as A may need it to. Isomorphism between E* and E would be impossible, if only because of the huge differences in the varieties of the domains in which the two structures occur. In effect E* corresponds to some 'aspect' of E that has relevance for A. On this account lacking E's constraints, E* may be manipulated by A in ways that E may not, with each manipulation altering E*'s structure and thereby degrading its correspondence to E. Given A's capacities, the use of paper and pencil combined with a method of the kind outlined by Descartes, opens the way to developing an effective procedure. Because of the structure of the entity that will use it, such a procedure may be taken to fulfil the necessary and sufficient requirements for objectivity, and not merely to act as an aid towards attaining it.

9.3.1 An effective procedure is explicit at each step. This is the criterion of its effectiveness. But it is objective, in Popper's sense, only for an <u>observer</u> able to <u>compare</u> it with something else. Step by step the inputs and outputs of a machine do not even tell O whether or not the machine is determinate. To find this out, O must be able to remember inputs and outputs as they occur to build up a model, which he may then test. If O is able to use the machine's procedure to objectify knowledge he has, he will have to do so in this overall rather than any step-by-step sense. Thus a given Turing machine's output is fully determined by the

Objectivity and effectiveness

machine's characteristics and its starting tape, though the machine has no objective knowledge of either. Like Popper's amoeba, its actions are part of it. But the machine's observer does have objective knowledge, because he can compare the Turing machine and its starting tape with its output. It follows from this that no essential difference distinguishes the knowledge of the Turing machine that its programmer has from the knowledge that the O has who understands its operation. Both need to know the machine's overall behaviour. Von Foerster (1972) observes that the scientific method 'rests on two fundamental pillars.' These are (1) the 'principle of the conservation of rules' according to which rules observed in the past shall apply to the future; and (2) the 'principle of necessary and sufficient cause', which requires that 'almost everything in the universe shall be irrelevant'. Of interest here is that 'relevance is a triadic relation', (p.86) which relates two sets of properties and the mind of whomever wishes to establish the relation; in our present terms for instance, the set of properties represented by the Turing machine's starting tape, that represented by its end tape and the O or programmer of the machine. In the terms of chapter 2, the observer must acquire the order that the machine's program represents, so that it can 'understand' the machine, while the programmer must have this order in advance.

9.3.2 Evidence for effectiveness determines degree of objectivity

In general an effective procedure need not in principle be sequential, though in practice the human observer will probably require any parallel process to be reduced to a sequential process in order to be able to check it step by step for effectiveness; that is to check its determinacy and to be sure that at every step a procedure exists for deciding what the next step must be. This practical requirement is no more than an heuristic, offering benefits of the same kind as saying what we think. But suppose a parallel process were reduced to sequential form by means of an effective procedure, checked for effectiveness and then restored to its parallel form by another effective procedure; or suppose more simply that a sequential effective procedure were operated too quickly to allow for step by step checking of its operation by an observer, then no observer could have objective knowledge of the process-in-operation, although this would in no way alter the way the operation occurred. Clearly if an observer O_1 had checked that a procedure P was effective, an observer O_2, who had been unable to check it, might, merely seeing it in operation, - comparing input with output say - suppose it to be quite haphazard, that is suppose he was witnessing a non-determinate process. Even O_1 would have no way of being sure that P operated in the same way when it was running at high and at low (checkable) speed, although in the former case he might devise <u>tests</u> to provide <u>evidence</u> about the way it was operating. Thus, though an effective procedure may operate without being observed, if it is to provide O with objective knowledge, O must be satisfied that it is effective. The degree of O's satisfaction will be determined by the extent or weightiness of the evidence to this effect he is able to assemble, and may be taken to correspond to the degree of objectivity of O's knowledge. In the limiting case O will have to be able to check the procedure at every step, and in doing this he may be aided, if he is human, by recording his findings by means of paper and pencil. The procedure has no value as objective knowledge, except to an O who has satisfied himself of its effectiveness by checking it.

SECTION 4 UNCONSCIOUS PROCESSES

9.4.0 Our argument so far implies that objective knowledge is a property of some particular O who is an observer by virtue of being distinguishable from the knowledge: he can write down the knowledge as a formal theory expressed as an effective procedure for example, and is thereby enabled to 'face' it in a 'consciously critical' way. But this image may be misleading. For if, instead of picturing Einstein seated at his desk with a pencil in his hand, writing on a page already partially covered with signs, we imagine successively a tennis player practising his forearm, a pianist his scales, a child learning to walk, an earthworm making an appropriate turn in a T-maze, the notion of an O facing his theory grows steadily weaker. The tennis player may make 'conscious' corrections to his strokes, telling himself to stiffen his wrist or follow through and, if his shots improve, this may be partly due to his conscious analysis. But part at least of any improvement is likely to be due simply to 'practice' the word psychologists use to describe situations in which typically numbers of repetitions of some action lead to a gradual 'improvement', usually some kind of assymptotic approach to a performance goal - activity resembling the amoeba's more than Einstein's. Traditionally psychology has viewed such 'practice effects' as the result of gradual strengthening of those components of action that mediate closer approach to the goal, and weakening of those components with the reverse effect. But even without adopting this traditional view, we shall still find that we are left with processes, similar to the consciously critical ones Einstein uses, only 'unconscious'. For instance Minsky & Papert (1972) take the view that 'The external appearance of slow improvement ... is an illusion due to our lack of discernment. Even practising scales, we would conjecture, involves distinct changes in ones strategies or plans

Unconscious conjecture

for linking the many motor acts to already existing sequential process-schema in different ways, or altering the internal structures of those schemas. 'Improvement' results from 'definite ... moments of conscious or unconscious analysis' that will include conjectures and 'structural experiments'. 'Thoughtless' trials are essentially wasted. The authors compare the process to debugging a computer program. 'It is not a matter of strengthening components already weakly present so much as proposing and testing new ones.' (Paragraph 4.3) If this is true of practising scales, it is presumably as true of the child learning to walk, with the possible difference that the child may turn more to unconscious analysis than the musician, though on the face of it there appears no obvious reason even for this. For the scientist, who Popper says advances his theory by series of conjectures or guesses, and refutations, it seems likely that the refutations may result from his 'consciously critical' attitude, but that the conjectures are more likely to arise, at least in part, 'intuitively', due to unconscious processes, however deeply these may be embedded in the appropriate structures that are part of the existing apparatus at the command of the scientist's thinking.

9.4.1 But it is for being 'criticizable' that Popper commends objective knowledge. Its objectivity is what facilitates the consciously critical method of one who seeks to 'kill' his theories. Refutation is what objectivity permits. But it appears that even refutations may come about through 'unconscious' processes. The child that improves its walking - even if it does so in the way Minsky & Papert suggest, <u>thoughtfully</u> - is probably satisfied with the improvement until new inadequacies appear <u>by chance</u>. The scientist on the other hand who improves his theory, sets out to find its inadequacies <u>by design</u>. But even this difference is not as sharp as the way we state it may make it seem. For on the one hand the child learning to walk will probably go in for at least some, more or less conscious experimentation, in which it tries out 'ideas' (conjectures) for new techniques in new situations; and, on the other, the scientist is often likely to miss obstacles to his theory until he stubs his toe against them. Furthermore the conscious criticism that the scientist directs at his guesses is itself very likely to rely to some extent on guesses.

9.5.0 Criticizing need not be conscious, and 'unconscious' criticism does not call for objectified knowledge. The foregoing argument suggests that Popper's assertion (9.1.1) should be understood to mean that only objective knowledge is consciously criticizable. This interpretation neither appears nor aims to violate the spirit of Popper's view, which we might paraphrase as a recommendation that knowledge should be put where it can be got at. Objectivity confers the practical advantage of fixing ideas so that they may be tested - 'What you put down stays there'* - which as we showed (9.3.0) is a facility indispensible to any entity capable of objectifying what it knows. Our object has been to show that, despite its advantages, objectivity should not be supposed the only way through which knowledge may be advanced. Conscious criticism demands objectivity, but unconscious criticism is also possible, and does not. It is to avoid such confusion that Pask (1975) introduces a stratified object language into the theoretical structure of his conversational and tutorial models. Thus instruction may occur by means of the higher level, tutorial language - English - called the L^1 language or by means of a lower level language L^0, which is a performance or 'task' language. The strat-

Higher and lower objectivity

---

* Remark made by David Hockney, during an interview on Omnibus, BBC Television, 8 May 1975.

ification offers an escape from the need for undesirable concepts, like notions of unconscious processes, to explain practice effects. This then permits the assumption that the objectivity of performance - at tennis say - is no less, for those processes governing it, than the objectivity of a theory for the process that produces it. Essential to both kinds of process is a structure that permits the process's controller to separate itself in the manner depicted in figure 8.2.3, from the models - of the world and itself - that it utilizes for producing the Ws that represent its 'knowledge'. Our interpretation of Popper's assertion to mean that objective knowledge should be consciously criticizable implies, in these terms, a preference for the higher level object language. But such a preference must be based, not on any claim for the greater objectivity of the higher level language compared to the lower, but to its wider applicability, which greatly enhances its capacity for representing order. Thus Ws that are tennis strokes will not easily connect with Ws that are piano-playing techniques, as long as both Ws exist at a performance level only; whereas a higher level representation may reveal connections (2.3.0) between them. Among the concerns of the following chapter will be the question of the language level of WOAs.

CHAPTER 10

EXPERIENCE

SECTION 0 SUBJECTS AND GOALS IN ART AND SCIENCE

10.0.0 Chapter 9 explored some of the general properties of W, especially in regard to their possible usefulness to A. Here we look for descriptive dimensions such as will allow us to distinguish Ws of different kinds. We argue that art's distinguishing characteristic is that its Ws are intentionally closely linked to experience, in contrast particularly to science, whose Ws are made deliberately to avoid such connections. Experience is individual, particular, unrepeatable and untranslatable. It is not merely what is called 'cerebral' but invokes the world of the senses and the emotions. The language of experience is a performance language (9.5.0). Art is its simulation. Chapter 0 cited views that assigned art the status of a kind of second best to science. Apart from Medawar's view of art's failure to conform to empirical reality (0.1.2), there was Von Mises's explicit contention that art belonged to 'areas of life not sufficiently explored by science' (0.1.3.2) which George (1970) echoes, assigning art 'prescisely the role that science is now taking over' (p.145). Here we contend that the goals of art and science - the subject matters or, more properly, what each comes to convey of the same subject matter - cluster at the extremes of the continuum that joins respectively the concrete and particular at one end to the abstract and general at the other, with less easily-classified works occupying the space between. This - as we shall see in the following chapter - dictates to art and science different modes of approach to their common subject matter. Here is the origin of much confusion and often contradiction in the views and attitudes about these two realms. For on the one hand, from the point of view of subject matter, especially with the development of psychology and the social sciences, it has become increasingly difficult to deny science and art common interests; while

Reproduction of experience

on the other, seen from the outlook of methodology, science and art have appeared - literally correctly in our terms - poles apart. Art's concern is experience, not its analysis but its reproduction. If play and exploration are the route to performance concepts in a performance language, art is their objectification and extension.

10.0.1 The chapter takes the problem of 'individuation' as its starting point, using a formulation of the problem, due to Russell (1948), that depends upon the concept of 'compresence'. This formulation is interpreted in terms of a machine with input, which is then used to define experience. Experience is shown to be particular and subjective. The formulation permits some comments on objectivity and subjectivity. Problems of derivation and the description of experiences are considered with a view to arriving at a notion of the dimensions of the concept experience. There is discussion of the characteristics of modelled experience. The chapter ends by characterizing art in terms of the reproduction of experience.

A formulation of experience

SECTION 1 COMPRESENCE

10.1.0 We shall base our definition of experience on a formulation that Russell uses in dealing with the problem of 'particulars'. Russell's formulation depends upon a notion that he calls 'compresence'.

Individuation

> 'There is a relation, which I call "compresence", which holds between two or more qualities when one person experiences them simultaneously - for example, between high C and vermilion when you hear one and see the other.'

We can form groups of qualities into single complex wholes, each of which has the two properties that '(a) given anything not a member of the group, there is at least one member of the group with which it is not compresent.' Giving its constituents defines such a complex whole but it is itself 'a unit not a class'. Its existence derives, not from its constituents' existence, but from their compresent existence in a single structure. When such a structure comprises 'mental constituents', Russell calls it a 'total momentary experience'. (p.315)+ The items of the complex may themselves occur frequently and 'are not essentially dated' (p.312), while the complexes, consisting as they do of many components that depend upon the spatio-temporal position of an observer, are empirically highly likely to be spatio-temporally unique and therefore time-ordered, in the sense of being non-recurrent, though repetition of a complete complex of compresence cannot be excluded on purely logical grounds.

10.1.1 Notice though that, although a 'total momentary experience' is almost certainly unique, nevertheless we may know and name a complete complex of compresence without having to know all its constituent qualities, in the sense that we may

Symbols

know - identify - a man by those of his qualities his passport enumerates, without the need to know anything else about him. Moles calls such identifying qualities 'symbols'. A complex signal E, consisting of the set of 'simultaneous' elements $\{E_i\}$ that separately create 'elementary perceptions' $\{P_i\}$ acquires a '<u>mnemonic symbolization</u>' when some 'well-chosen' element of E is sufficient to provoke the set P of elementary perceptions. Moles calls the well-chosen element the 'symbol of the set $\sum E_i$' (p.98). (cf. the observations of Bruner <u>et al</u>, 6.1.1).

10.1.2 In the physical world a 'complex of compresence' is constructed on the same principles as those employed in dealing with momentary experiences, merely by supposing an absence of percipients. Events may be defined in terms of 'in-

Events

complete' complexes, where an incomplete complex receives a topological de-

+ In reality a 'moment' has finite duration. Empirically, with humans for instance, there is a lower limit to the time intervals between which human subjects can discriminate, the length of time for which a musical tone is sounded for example. Moles calls this time 'the length of the present' (p.15).

finition in terms of a region of space-time. Thus an incomplete complex occupies a continuous region in space-time, if a continuous route, consisting entirely of points of the region of which the complex is part, connects any 2 space-time points of the complex. 'Such a complex may be called an "event". It has the property of non-recurrence but not that of occupying only one space-time point.' (p.322)

10.1.3 The distinction between a unit and a class parallels the distinction between a quality and an 'instance' of a quality. The latter 'is a complex of compresent qualities of which the quality in question is one' (p.3.6). Clearly

Time order

there can be any number of different instances of the same quality. But notice that instances of qualities are time-ordered, though the qualities themselves are not. Indeed it is wrong to speak of two occurrences of the same quality, since, being the same, they are indistinguishable and can therefore only be counted as one. 'When the same shade of colour exists in two places at once, it is one, not two' (p.316). The distinguishing characteristic of a complex of compresence depends upon its separability from other complexes, and in particular on its time-orderedness. Statements about purely logical structures such as classes, reduce to statements about their components, which is not possible in the case of time order.

SECTION 2 STATE DESCRIPTIONS

10.2.0 We may interpret Russell's formulation in terms of state descriptions. Suppose we take a complete complex of compresence to correspond to the state description of a system at a given instant. Such a state description might be repre-

Compresences and states

sented as a vector whose elements represented the 'qualities' compresent in the particular complex. A compresence composed of 'mental constituents', is termed by Russell a 'total momentary experience'. Such an 'experience' might be represented in the same way as any other complex of compresence. Russell indicates that a mental compresence is to be regarded as the result of a physical one. A mental compresence occurs in a 'percipient' who may be 'aware' or not of certain 'qualities', like redness for example, which are constituents of a physical complex. In our terms we might picture two systems, or more precisely a single system partitioned into two sub-systems, representing respectively the percipient and his world. Inter-connections between these two sub-systems permit the percipient to receive inputs from his world, which depend on the world's state. In this way the percipient's states come to correspond to correlative states of the world, and these states of the percipient's are called his experience. The states will themselves depend upon the internal state of the percipient at the time of receiving them as inputs. But this raises no difficulty because the percipient's internal mental states are simply correlates of his physical state, which is to say, included in the state description of the total system. The percipient's mental states depend upon his 'awareness', which in turn is related to the inter-connections between the sub-systems that the percipient and his world represent. For this reason 'We can never know that a given complex of compresence is complete, since there may always be something else, of which we are not aware, which is compresent with every part of the given complex.' (Russell, p.232) (cf. the views of Popper and Lofgren, cited 2.3.1; also Pask & Scott (1973),who identify mental states with the Cartesian product of inputs and outputs of the entity in which the states subsists).

10.2.1 More formally, if the percipient has internal states S, and input states I, then the produce state, I x S, will be mapped into S under some mapping f say, that depends upon the percipient's characteristics, what Ashby (1962) calls the 'dynamic drive of the system'. The mapping will correspond to the percipient's experience of I. In this formulation, the state description of

Vectors

I may be taken to be the physical complex of compresence and that of S its 'mental constituents'. Referring this formulation back to Russell's, we notice that if we remove the state description of S, we still are left with the state description of I. Presumably in the absence of a percipient, the physical complex of compresence would consist of the state description of I alone; and conversely, that the percipient's experience at a given instant is defined by the state description at that instant of S alone, irrespective of I.+ Events are definable in these terms as groups of transformations of vectors, and the experience of an event, consonantly with this definition and the definition of experience. The order of the vectors that occur will be finite, though high enough to make their recurrence exceedingly unlikely, thereby satisfying the requirement of particularity. In nearly all cases of experience, most of the variables that define a space-time point-instant will be unknown; and in practice a tiny proportion of the variables of a given vector will often be sufficient to distinguish the vector uniquely (in a non-recurring) way. (cf. Moles's formulation, 10.1.1) The 'qualities' of Russell's formulation are to be associated with the values of the individual variables that comprise the vectors.

SECTION 3 EXPERIENCE

10.3.0 The concept of compresence and our interpretation of it should not lead us to visualize a dichotomy, consisting of 'particulars' on one side and 'purely logical structures' on the other, no more in the world of experience than in the physical world. Complexes of compresence are not necessarily complete. Events are examples of incomplete complexes of a special kind, and as we observed in 10.2.0, we can never know whether a complex is complete because we

+ Notice that this accords with common usage that attributes experience to inaminate systems so that one may make statements for example about 'the world's experience to men'.

may be unaware of part of it. We may therefore regard particulars and classes as the extremes of a continuum, whose intervening space is inhabited by the events and other incomplete complexes of varying particularity. The more incomplete a complex, the less its particularity and thus the greater the empirical likelihood of its recurrence (where by recurrence of a complex we understand that, if given a complex and its recurrence we should be unable to say which had occurred at an earlier and which a later date). Corresponding to this view, we visualize a 'continuum' of vectors defining states that vary in generality roughly inversely with the order of the vector, from the completely general single variable at one extreme to the high order vector with all its variables defined and known at the other.

10.3.1 The greater the degree of particularity, in the sense in which we have been using the term, the more subjective the experience of it. The space-time point-instant is the total momentary experience of a single observer at a single instant. Conversely the experience of a quality, such as redness say, is not subjective or private, but may be experienced by any number of observers Only instances of qualities are subjective, that is qualities taken as components of complexes of compresence. Extending this argument, it follows that there will be degrees of subjectivity, which will vary in proportion to the size of the complexes of compresence, or the order of the vectors by which they are represented. In general experiences will be complexes sufficiently large to make the same experience for two observers highly unlikely. Thus we should say that almost all experience is necessarily wholly subjective, in the sense of being unique to a particular system. While conversely, 'experience' that can be shared by any number of observers - a quality say - is wholly objective. Notice that this formulation clarifies one of the difficulties of objective knowledge that was encountered in the previous chapter, namely that objective knowledge was likely to have a subjective and non-explicit content. This difficulty results from a failure to distinguish fully

Particularity and subjectivity

the difference between a quality and the instance of it, which is what is experienced. It is the problem of the observer that we shall meet in the following section. And we have come across it before under the heading 'knowledge by acquaintance and knowledge by description'. We recall a quotation from Russell cited 1.3.0 ' ... everything except myself is ... only known to me by description, not by acquaintance. And the description has to be in terms of my own experience', which is the conclusion to which our present argument also has led us. The ability to distinguish qualities at all from instances owes to the power to abstract, as psychologists sometimes call it. In our terms we may think of abstractions as logical <u>relations</u> between complexes of compresence that define instances of the same quality. Thus the join of a number of instances of red may contain only the single element red, which we should call the abstracted quality, redness.

10.3.2 As we have defined the experience of an individual as his internal state at a given instant, it follows that it will be impossible for him ever to compare directly his own experiences one with another, or with the experiences of another individual. Such comparisons would require changes in his state that would obliterate the experience he wished to compare. Yet similarities do occur between experiences. The comparison of two experiences is itself an experience that is different from either of the experiences being compared. Apparent similarities between experiences are in reality similarities between some kind of representation of them. Thus for example two experiences, each represented by a vector, could be compared in this represented form, by comparing the vectors by means of some form of mapping of one onto the other. If the vectors were both of a high order, the likelihood of discovering some kind of morphism or 'fuzzy' (7.2.3) similarity between them would presumably be lower than if they were of lower order. So the fact that similarities between the representations of experience do occur suggests that

Similar experiences

the representations used are of a simplified kind and possibly acquire a degree of uniformity from the way they are arrived at. Thus for two apparently similar experiences, the individual values of the elements of their constituent internal states - the values of the individual elements of the vectors defining the experiences - may be 'fuzzily' equal or homomorphic, because both occur as the result of similar physical complexes. A visual scene viewed by two observers for instance would furnish experiences to each that shared numbers of identical components as well as numbers that differed by a negligible or unimportant - in the context - amount. Or apparent similarities might be due to the effects of the internal states of the individuals or systems concerned in determining the experiences compared. For, as we have mentioned, it is not the experiences that are compared but descriptions of them.

SECTION 4 THE DESCRIPTION OF EXPERIENCE

10.4.0 Dewey (1934) distinguishes between experience and an experienc, 'with its own individualizing quality and self-sufficiency' (p.35). The concept 'experience' derives from descriptions of experience, of which it is itself a description.

Experience and observation

The definitions we have suggested give an indication as to how descriptions of experience might come about. In humans this is a matter for psychology, though broader, more abstract questions concerning description are, as we saw in Part I, of wide generality. To describe his experience an individual must observe it. Now, by definition, experience is subjective. Observation by contrast is not, at least in the same degree. For although an observer registers an observation as a change in his internal state brought about by what he observes, it is not the observer's internal state itself, but what it refers to, the state of the observed system, that is of interest. And the observer normally adopts expedients to minimize the effects of his own internal states and maximize those of the system he is observing. In the case of experience on the other hand it is the internal state of the system having the experience that is of primary interest, with any system to whose state it refers of secondary importance only.

10.4.1 Representation of experience

Ignoring for the moment difficulties of comparing the internal states of different systems, the above arguments will allow certain observations on subjectivity. In particular earlier discussions about perceptrons and procedures such as those of Guzman (6.1), Winston (6.2), and Winograd(6.3) have indicated some of the kinds of processing that the data of a complex may undergo. Perceptrons for example weight the 'evidence' supplied as the data of their inputs by the components of a complex, and arrive at a decision - expressed by 'naming' the complex - on the basis of this weighted evidence. Guzman's and Winston's procedures also utilize input data as evidence to which varying weights attach. Their procedures operate on already-structured data. They depend upon complexes consisting of relatively small numbers of components, with the 'other', in this case purely notional components - such as those that would accompany a human experience of Guzman's or Winston's blocks, for example - assigned a zero weighting by the outside agency of the designer of the procedure.+ Obviously the effect of such weighting of evidence will be to emphasize some - in effect a few - of the components of a complex while, usually greatly, reducing the importance of the rest. Clearly such compression of the complexes that make up experiences increases the likelihood of apparent identities between experiences, far beyond what we should expect from unprocessed data.

+ As we noted earlier (6.2) this, as well as the descriptions with which the procedures are provided, represent the greatest part of the order-extraction work associated with the success of these procedures. It is precisely the extent of this work that defeats the perceptron required to build up its own weights from a starting position in which every component of its 'experience' has equal weight. (And notice even here that a perceptron's experience rests on only a tiny fraction of the number of components that make up the experience of the experimenter say). We may therefore regard the experiences that such procedures represent, as having been drastically attenuated, by an amount proportional, we should say, to the order-extraction work of the designer of the procedure, compared to the total amount of order-extraction work that the procedure performs.

10.4.2 Clearly Russell's formulation was never intended as a practical guide for the observer, to show him how he might describe the abstract concept of experience - rather than individual experiences, from which the concept initially derives - so as to capture its quality in a way he might find intuitively familiar. But we should not take procedures such as Guzman's too literally either as providing descriptive models. A virtue of Russell's argument is that it emphasizes the transitory nature of experience, while in procedures such as Winston's or Winograd's, or more especially procedures such as those used by perceptrons, experience tends to become associated with fixed descriptions. What is lost in such fixed, 'ended-off' processes is precisely the quality of on-goingness or changefulness and, associated with it, what is described by words like subtlety and nuance, which, to our intuitive conceptions, characterize the very fibre of experience. It is mistaken to suppose that, simply because 'awareness' encompasses only a tiny fraction of the components to which it has access it follows that experience, stripped of these 'unimportant' components, would necessarily retain the quality of the <u>notion</u> of experience familiar to us, remain recognizably the same thing that the word 'experience' commonly denotes. What it would be like to <u>experience</u> the stripped world of Winograd's 'robot' is literally inconceivable to the human mind. Moreover awareness is not dichotomous, it varies in degrees. Changes in components of a complex of compresence that were not part of awareness may make them part of it. It is the case too that an experience can exist at varying 'levels' of completeness - number of components of a complex of compresence in awareness - and in a way which may make it seem that the levels represented separate experiences, following one another perhaps in rapidly changing sequence. For instance the experience an individual describes, the one he uses as the basis for some deduction, the one to which he gives a name and the one that reminds him of some other experience, may all consist of components due to the same physical complex, though each represented in experience in different combinations and strengths. It is by no

Characterizing the concept of experience means obvious what descriptions one might use to characterize the concept of experience in a way that would make it as recognizable as the lines and vertices used by Guzman's SEE programme make the shapes of its world. The more so because experience is ever-present and on-going: to be aware or conscious is to be experiencing; to cease experiencing is to be unconscious, unaware.

Compresence and the characteristics of experience

10.4.3 Compresence appears able to furnish the attributes that a description of experience requires. The concept captures at once the qualities both of simultaneity and time-orderedness, which introspective awareness distinguishes as characteristics of the abstraction 'experience'. And it is also able to describe other characteristics of experience. Thus the experience that the observer may describe or the experience that has some utility for him beyond itself and which he therefore strips of what, for his purpose at the time, are its inessential components, has less the characterizing qualities of experience than the fuller experience from which these derive. And progressively the more 'inessential' components that are discarded, the less the residual ones are able to embody these characterizing qualities. Yet the concept of compresence is quite as able to interpret these qualities as it is the more subjective, particular ones. That the irrevocably subjective character of experience as it is present to subjective intuition, as well as the non-subjective nature of abstractions, such as qualities and logical constructions, is captured by the single notion, compresence, and that moreover we can formulate the notion in terms of vectors and their products, familiar to cybernetics, appears a sufficient justification for adopting it as a basis for interpreting experience.

SECTION 5 ART

10.5.0 Chapter 2 interpreted the Ws that A produces in terms of order: as the outcome of order-extraction work. A's goals have the force of procedures whose execution reproduces order extracted in the form of Ws. Chapter 8 showed the necessity of having Ws of one kind or another for the operation of any interesting control procedure: the Ws furnish such procedures with the objective models of the world that the procedures require for predicting the world's behaviour. Chapter 9 noted the heuristic value for the extension of order extraction of being able to 'face' a W in the consciously critical way called for by the need to eliminate errors. The foregoing section of this chapter showed that, while experience is wholly subjective, descriptions of it are possible and permit objective operations with subjective experience as a basis. Without the possibility all objective knowledge or even communication would be impossible (cf. Russells observations on 'egocentric particulars' (1.3.0)). These descriptions are normally operated upon in a trimmed and generally more tractable form than the one that might portray the actual experiences that they describe. We showed how trimming was achieved, by reducing the characteristically high order vector by which we might describe the internal state constituting an experience of a given individual, to a vector or vectors of lower order, either closed within the same set of elements or homomorphic to them. It was suggested that these reduced homomorphs of experience steadily lost those qualities by which we should normally characterize them as experiences, becoming, as the order of the vectors representing them diminished, increasingly disembodied and abstract. We noted too that while experience was necessarily subjective, we might picture a 'continuum', with subjective experiences at one extreme, non-subjective+ abstractions at the other, and 'events' of varying degrees of abstraction somewhere

Characterization of art

+ In view of the problems associated with the notion of objectivity (10.3.2. et seq.), we prefer to avoid the term here and to use 'non-subjective' instead to denote what is not purely private, but enjoys public, shareable characteristics.

between. We shall define as WOAs those Ws concerned with experiences, as we have characterized them (10.3).

CHAPTER 11

THE METHOD OF ART

SECTION 0 SYNTHESIS AND SIMULATION

11.0.0 The previous chapter argued that the objectives of art and science differed. It was shown that while art aimed at producing Ws that were particular and therefore needed to be in some measure self-sufficient, science aimed at general Ws based on previous non-evident structure. Here we contend that this difference of aim dictates a difference of method which, more than differences of explicitness or objectivity or conformity to empirical reality distinguishes and characterizes art and the artistic content of Ws generally. The difference between art and science is broadly that between simulation and sythesis. The ontological status of WOAs is analagous to that of a simulation program for a computer. Ws that are scientific theories are the occasions for synthesis. This view appears to furnish a basis for a resolution of the conflict between art or science. Theoretically at any rate nothing can prevent science's taking over art's function, insofar as this is to provide Ws that constitute objective knowledge. On the other hand, nothing can take the place for O of the experience that a WOA affords. The success of a given synthesis or simulation represents a W's merit, which is variable, whichever way the particular W inclines. This chapter completes the answer to the problem we set ourselves initially, to determine the principles of an art machine. Part I showed what kind of machine would be needed. Thus far Part II has shown what the machine will have to be able to do. The argument here yields a number of conclusions which, though they follow from the arguments that have led to this point, represent, in some respects, a more speculative departure than any others so far.

Art's aim determines method

11.0.1 This chapter begins with a view of art as a model, of a model, capable of answering questions of a general nature about experience. It discusses the differences between models of reality and of experiences of it, and difficulties arising from the relatively greater variety of the real world than A. This leads to discussion of surrogate experience and the means of supplementing A's variety to facilitate its producing such surrogates. It appears that art produces surrogate experience by a mixture of simulation and synthesis. Differences between real and surrogate experience are discussed. Finally the chapter examines the ways in which A may approach producing surrogate experience. It emerges that WOAs have a dual existence, as surrogate and real experiences, and that each furnishes A with a language consisting of the same words but with different meanings. The chapter ends with a discussion of aesthetics.

Real and surrogate experience

SECTION 1 REPRODUCING EXPERIENCE

11.1.0 So far, in discussing W we have concentrated on the minimum requirements for a procedure that would produce W (c-authorship), the sort of structure the procedure would need and the relation of the procedure to the W it produced. Only at the end of Part I did we show explicitly how W might stand as a model of A's world or part of it, and we left for chapter 9 touching on some of the broader advantages (objectivity) of such models for advancing understanding of what is modelled. The reason for following this course has been, as we have previously observed, to avoid restricting unnecessarily the generality of the conclusions reached. Thus we have narrowed the area of speculation only where necessary. Recapitulating, we picture W as some kind of observable (to a properly constituted observer) object or set of objects, distinguishably (physically) separable from A and modelling some part of A

Art's function

which may include some part of the 'model' A has of the world. By 'model' we understand a broad notion, like the one Minsky (1965) describes. 'To an observer B an object A* is a model of an object A to the extent that B can use A* to answer questions that interest him about A' (p.47). In our terms W* is A's 'internal' model, which W is supposed to represent. Thus W is a model of W*, in Minsky's terms, the model, W**, of a model. From 0.2.2.0 we recall that such a model of a model, W**, answers questions about the kind of model W* is. Part II has dealt with aspects of this question in some detail, and has confirmed and in some respects extended Minsky's view. Thus the models art is able to furnish of particular experiences are the means by which to answer questions about the kinds of experiences they are. We shall devote the rest of our discussion to examining how this is accomplished and finding out what we can about the kind of procedures it requires.

11.1.1 The difficulty of modelling experience results from our wish to model the experiences themselves, not simply conceptual abstractions. What kind of model of an experience would provide a means of answering questions about that experience, without reference to the experience itself? (As we indicated (10.4.0), referring to an actual experience is anyway impossible, as the 'reference' would destroy the experience.) Such modelling is evidently possible; for, although experiences are by definition unique, unrepeatable and momentary, we can and do remember and reflect upon 'them'. Memories of experiences must be models of the kind we are discussing. They are not themselves the experiences - although they are experiences - but enable the recollection of the experiences, to some slight degree their re-enactment. Here we must notice an important distinction, namely that between the model of an experience and the model of what occasioned the experience, the experience's physical, publicly-observable correlate. An example will make the distinction clear. A visual scene affords a visual experience to the individual looking at it. The scene is something the individual observes. His

Models of reality and models of experience internal state that results from the observation is his experience of the scene, but the scene itself (cf. 10.2.1) represents the state of a system distinct from the individual. We shall call it the original of the experience. All a given individual can tell about the scene must come from his experience of it, which is to say that, for the same individual, a perfect model of the scene would not differ from a perfect model of his experience of it. In so far as the individual's models fall short of perfection, the purpose for which he intends them will, at least partly, determine their form. Thus the 'true' model of a scene and the model of an individual's experience of the scene are likely to differ. The best model of anything is one that provides answers to any question about the thing. Thus the best model of a physical object is an identical or fully isomorphic object. Similarly the best model of an experience. Were experience exclusively of the physical world, then modelling experience and modelling the physical world would, if perfect models were the aim, present identical tasks. In such circumstances, art and science would be indistinguishable. But from the definition of experience proposed in 10.3 it appears that this is not the case.

11.1.2 In (7.3.4) we cited Spencer-Brown's comments concerning description and injunction and suggested that descriptions referred to goals and injunctions to the paths that led to them. Or, as one might put it, that descriptions referred to ends, injunctions to means. Accordingly it appeared that descriptions were to be associated with heuristics, injunctions with algorithms, and it was shown that the division between the two was not entirely clear cut. The tacit assumption that underlay the discussion referred to was that possible goals were relatively fewer in number than possible starting points from which to set out for them. In the discussion of hill-climbing for instance (4.3.1) we supposed that the climber might begin from any of an in-

Simulation's building blocks

finity of points below the summit with the same aim, to arrive at the single point that was the summit. Modelling experiences faces us in one sense with the reverse situation. The same original may occasion inumerable experiences of it. In the formulation of 10.2.0, given an individual with internal states S and inputs - from the original - I, the experience will be represented by the mapping of I x S into S. Even supposing S's variety were very much smaller than I's, as it will in any event be large, the Cartesian product as it stands offers no sensible limits to S. To limit the variety of experiences that a given original may occasion, we rely upon the 'correlation' that we suppose exists between the individual and his world. That is, for the reproduction of what we shall call an <u>original experience</u>, we suppose that the individual has at his disposal descriptive components corresponding to the components of the particular original, such that these components enable the original in some sense to be computed from them in a manner such as that suggested in chapter 2. (cf. e.g. Craik's assertion that 'thought models or parallels reality' - 2.7.2 footnote.) These descriptive components are the descriptions of 2.3.0 or the descriptors of chapter 6, and thus fulfil a synthetic role. So that the reproduction of experience is evidently a mixed synthetic-simulatory procedure. The synthetic components may be taken as the building blocks of the simulation procedure that we call a WOA. These building blocks constitute the artist's prior structure that we discussed in chapters 4 to 6. If O lacks any of the descriptive components that A relies upon, he will be unable to reconstruct that part of the original to which they refer. Such lacks will appear as lacunae in the experience that the WOA occasions and may distort the entire simulation. Distortions will also occur if O and A synthesize A's descriptive components differently. Such problems are the subject of N.8.4.

SECTION 2 SURROGATE EXPERIENCE

11.2.0 Thus a WOA enjoys a dual role. It provides at once the occasion for what is an experience in its own right, while standing also as a surrogate for another experience for which it provides a simulation procedure. The previous paragraph referred to descriptive components that A uses as the building blocks of this procedure. If, ignoring possibly different hypothetical possibilities that are of no interest here, we suppose that A's variety is very much lower than E's, we may usefully look upon the descriptions A uses as ways that A employs to increase it's variety. For example, among A's descriptors are a class often called percepts, which, following the terms we have been using, we might think of as the building blocks of perception. The vertices of Guzman's SEE program are crude examples of what is meant. These vertices have a real existence in the world in which the program operates - as observers of the program in its world we are qualified to say so. Thus in a sense the program exploits its world as a kind of memory store to supplement its own limited one. In this interpretation A's percepts have the force of accessing sub-routines aimed at this memory. We should argue in favour of this view that an important part of W's role as we have depicted it is simply to be an extension of this accessing facility, one in which A actually builds the kind of memory store he wants in the physical world, rather than relying on what he can find there - A the cultivator, no longer simply A the hunter. Among the empirical evidence that might be taken to lend general support to this view of A's relation to and connectedness with its surroundings is the intuitive difference made between recognition and reconstruction. We may recognize a bird for instance from appropriate clues (6.1.1), as we may know a man by his passport (10.1.0). But to draw the bird with a pencil it would probably be necessary to have it in front of one for frequent checking of details whose absence from the mental picture of the bird pass unnoticed, like the blind spot passes unnoticed in normal vision. Instead of storing the bird in memory, a way of accessing

Supplementing variety

birds is stored instead, called a description of birds. The rapid loss of function of subjects in sensory deprivation experiments appears to offer further vindication of the view advanced here.

11.2.1 This view of A's use of descriptors furnishes a basis for a way of understanding the operation of the descriptive components A uses in simulation procedures. In the formulation of 10.3 we took experience to be defined by a complex of compresence consisting of mental constituents, and we showed that 'total momentary experiences' were highly unlikely to be repeated. Experiences consisting simply of the raw data of some complex might turn out impossible to compare at all, and therefore a fortiori impossible to judge as similar. The unfulfilled promise of perceptrons, as we saw in chapter 5, owes largely to a too simple belief in their power to process such raw data, without the intercession either of local or global descriptors. But remarkably, experiences are comparable and similarities between them commonplace, as long as descriptors are available to simplify descriptions. Descriptors' power derives (as in 2.7 and 2.8) from the way they permit shortening of descriptions. Or, put in terms more useful here, from the way they reduce descriptions' varieties. For instance to describe a grass lawn say by means of a numerical representation of some finite 'retinal' projection of it would require a variety proportional to the number of points on the retina, raised to higher powers according to the number of wave-lengths and intensities distinguished, a variety that 'lawn' reduces to a single word. So we reach the conclusion we have reached a number of times in different contexts, that without the structure, the heuristics, the descriptors or whatever, WOAs are impossible. And moreover, also as we have previously said, because 'there is no a priori unique information output in communications between individuals' A's audience will need to share most of his structure. 'The real information depends on the common knowledge of the transmitter and receptor'. (Moles p.52)

Descriptors and reduction in variety

11.2.2 With the description A has at his disposal he can do more than simply reproduce experience, he can create it. The idea of a WOA reproducing an experience is paradoxical. For what we have called the original, which is the occasion of the original experience, may bear little resemblance in its physical reality to the WOA, and O may thus in a sense be thought of as having an experience without having it. An analogous state of affairs exists in physiology, which distinguishes between adequate and inadequate stimuli. Thus light is the adequate stimulus for the cells of the retina; pressure on the eyeball, when it activates them, the inadequate stimulus. We may think of hallucinogenic drugs or electrical stimulation of the brain as displaying comparable inadequate-stimulus properties, defining the domain of imaginary, as distinct from real - 'adequately'-based - experience. In cases of drug-engendered hallucinations and even dreams, experiences may be difficult to describe or even retain. In support of the contention of 11.2.0 that A's perceptions are also a facility that permits using the physical world as a back-up memory, we might argue here that the difficulty of dream image retention resulted at least partly from the poor correspondence of the images with 'reality'. This would then lead to inadequate access (percepts) to the back-up memory, 'reality', with resultant difficulty of storing, leading to the kind of poor retention usually associated with dreams. In the case of other forms of hallucinatory experience - such as that reported by Penfield (1958) - due to direct electrical stimulation of the brain using micro-electrodes, what the patient experiences may be precise, explicit and even familiar. One is tempted to suppose what many theorists have had in mind, most explicitly possibly since Craik's (1949) exposition, that anyone knowing and able to operate the brain's equivalent of a computer's machine code, would have it in his power to induce in any given individual, any experience he wished by means of an appropriate inadequate stimulus. It appears that during hallucinations the descriptions themselves may become distorted, in the sense that the reality they generate does not resemble the 'normal'

Real and imaginery experience

reality of the physical world whose correlates they should be. In addition the constructions that are made out of these distorted building blocks may themselves be bizarre.

SECTION 3 CREATING SURROGATE EXPERIENCE

11.3.0 A's task is to find appropriate 'inadequate' bases with which to make such WOAs as will provide those imaginery experiences that he wishes them to occasion. Among such bases he must find the techniques he requires, whether they are merely the coding procedures necessary to produce the illusion of perspective on a flat surface, or the injunctions to bring about the simulations that will induce some particular mood or emotion. But we must qualify what has been said so far. The foregoing argument of this chapter has been put in a simplified form, as if it had been believed that O might be unable to distinguish an experience due to a WOA from one due to an original that the WOA represented. This was never supposed. It might be A's intention to make a WOA such as would delude O into believing that it were real, but this would not generally be the case. As we remarked at the beginning of this section, WOAs have a dual role, as the occasions of real experiences in their own right and as surrogates for other real experience due to originals that they represent as 'copies' - achieved as we have asserted by means of mixed simulatory and synthetic procedures. Both A and O are aware of this duality, which A is expected to and generally will use. Thus A's purpose may be illusion and he may achieve it with considerable success. Or he may wish the surrogate to provide for the detached contemplation of the real. Clearly W's capacity for 'fixing' in reality some more or less fluid condition of A's (8.3) is of particular value in WOAs that are able to apply this capacity to the transitory, moving aspect of experience, so that a single experience,

The duality of WOAs

or at least an aspect of it, is available for repeated review (See N.11.0). A may wish to utilize the immediate juxtaposition and alternating presence of the experience occasioned by a WOA and by the original of which it is a surrogate. Clearly A may have any among numbers of objects.

11.3.1 In 9.5.0 we raised the question of a stratified object language for A. The previous paragraph makes it clear that A commands not merely an object language but a meta-language too, in which he may discuss such matters as the relation of the WOA and its original. As we suggested, these meta-linguistic discussions may be - and often are - themselves included as aspects of WOAs. We should be clear what these terms mean when applied to a WOA. Notice first that the language that concerns us is the one in which the WOA speaks to O. Ignoring the question of a meta-language for the moment, it is clear that A may use a mixture of higher and lower level languages in a WOA. Because of the duality of the WOA, he may often, within a single statement, capture the force of both higher and lower levels at once. A speaks two languages at once, using the same words. In a sense this is the crux of artistic achievement and the examples of it are so numerous that choosing among them is necessarily highly arbitrary. But take the instance of Michelangelo's <u>pieta</u>. Jesus's figure <u>depicts</u>, in a higher level language, the limp body of a young man recently dead. But in addition its presentation offers O a set of injunctions in a lower level performance language that induce an <u>experience</u> of limpness, in his own body. Such an experience may often only be conveyed, or at any rate is usually in practice more faithfully conveyed, by such lower level injunctions that have the force of a simulation procedure. Injunctions in a higher level language, - in this case to 'be limp' - would probably fail to be effective, simply because O lacked the synthetic components necessary to carry them out. O simply does not know how to 'be limp' or cannot achieve limpness through his own devices as well as he can by means of

A's two languages

empathic control of his muscles.+ Considerations of a similar kind apply no less when A's 'medium' is language itself or even, in some cases, music. Notice that nothing compels A to use his two languages as in this example, to convey sympathetic or reinforcing messages, on the contrary he is free to seek effects by any number of means including the languages' antithetical use. The joint effect of A's dual messages, frequently commenting on one another, implies A's view of his WOA as it strikes him in his role as O. In this way his stratified language provides the vehicle for his meta-linguistic comments. Here again is language stratification. A's implied statement about a WOA, which emerges from the mutual commenting that takes place between the two levels of his object language, is itself like a performance language, which induces a 'frame of mind' in O without explicit statement. In certain circumstances - chiefly in literature - A may make explicit comments too. But one would not wish to stretch this notion of a stratified meta-language too far. For the division of levels is not clearly defined and the notion of no obvious use here.

11.3.2 Why make WOAs or Ws for that matter at all? A's complex purposeful structure ensures that the reason must form part of the overall 'teleonomic project' that Monod speaks of, whose main goal is survival; ' ... it is survival that

The WOA's utility for A

is the lynch-pin of our motives ... ', 'the fundamental feature' in the motivational hierarchy. (George, 1970, p.142). In several places we have mentioned W's broad utility for A, without attempting to decide whether it might be possible to distinguish WOAs as a sub-class of Ws, simply from knowing their utility to A. This chapter contends that WOAs deal with experience, extend experience, 'fix' it for contemplation and repetition. Paragraph 11.3.0 observed that WOAs generally were concerned with more than simply mimicking 'real' experience, and paragraph 11.3.1 showed A's capacity to stand back from real experience and observe the relation between it and the

+ This is not to deny that such effects might be achieved synthetically, by means of description. Techniques such as the Method of Mattheus Alexander aim at just this. But the descriptions this Method employs are complicated and lengthy, took years to evolve, and in the end are seen to depend more upon actual physical manipulation than verbal description.

experience occasioned by WOAs. This latter is clearly a necessary part of A's equipment. Without it no productive activity he began would ever end. For being trapped within an on-going experience, A's goal for producing W would be continuously changing, rendering its realization impossible. Among the advantages for A that might be associated with these and other of his characteristics, which he in turn imparts to his WOAs is one we mention specifically. It is the capacity that WOAs provide for allowing A gradually to move from simulation to synthesis, from imagination to description. The contemplation of experience repeatedly re-provided by a WOA facilitates order extraction. It provides hindsight, which Polya among others identifies as a powerful problem-solving heuristic. Understanding (order extraction) occurs when A is able to map a description (2.3.0) onto the WOA's set of injunctions. As a final observation, notice that A is self-sufficient. It requires no observers or audience. Once it exists communication with other entities that may exist also is incidental. In effect however, as a complex machine, A must be the product of a complex process controlled by some antecedent complex individual, and so on, back, and would therefore normally find itself one among many, with questions of audience arising naturally. Theoretically though, in so far as it might be possible to identify A as the end product of the continuous evolution of the same, single individual, it might exist and produce WOAs quite unaffected by the absence of an audience.

SECTION 4 AESTHETICS

11.4.0 It remains to say a word about aesthetics, - the appreciation and, in our case, also the production of beauty - though the theory of art, which has been our concern, and aesthetic theory, are largely separate interests. Art does not necessarily aim to create beauty nor is what is thought beautiful

Aesthetic value defined on a relation

necessarily art. Here, following the principle we have adopted from the outset, we shall examine only such aesthetic questions as might have a general application to art machines as well as human artists. That is we shall look for criteria that are not merely free from cultural or eductional or socio-economic influences, but that have a universal applicability, what Clive Bell (1914) showed must be a quality common to all objects that provoke it, though we should have to add to this the severe and possibly vitiating qualification that the quality need only be apparent to the objects' producers. The quality that we therefore shall be seeking will not be a quality of objects as Bell supposed, but, as chapter 1 suggested, a quality of a relation between objects and entities that asserted that the objects were WOAs.

11.4.1 Paragraph N.2.4 introduced the concept of 'information potential' and asserted the existence of a positive relationship between the degree of this quality and the aesthetic value of the object - description, image, or whatever - to which the quality was attributed. Information potential is part of what we have called the synthetic aspects of the WOA, and in general is not connected with the simulatory aspects - in human terms, the experience of emotions or sensations, which may be harrowing or grotesque at least as easily as they may be beautiful.+ We conjecture that this quality, together with the positive - albeit indirect and distant - associations that the WOA may have with A's - and his species - survival, whose overriding importance we have already noted, constitute the only objective criteria of aesthetic value. And this only because the survival criterion subsumes the criterion of information potential. For the accumulation of correlatives, descriptors, &c., having high information potential, must be of primary importance for survival. The assertion is that compression, economy, &c. has acquired intrinsic and

Objective criteria

+ There is a sense in which the simple power of a feeling that a WOA occasions is sometimes interpreted as being beautiful, irrespective of its content. The interpretation is partly related to the Aristotelean notion of catharsis and we shall not pursue it here.

overriding aesthetic value, over and above what it may convey, because its quality of economy or whatever, is important for survival at a higher more general level than any message it may be used to communicate.

11.4.2 There is a danger in any argument that attempts to deal with a subjective idea like aesthetics being pitched at too general a level, for it may easily end in meaningless abstractions. The alternative is an empiricism that may vitiate the aim of achieving objectivity. There is some excuse for applying to an art machine criteria that might be valid for humans, since we have shown that any art machine that we might make ourselves would be bound to turn out WOAs that resembled those of human artists. However when we turn to aesthetic theory we find little agreement among authors as to what such criteria might be. The indirect and distant associations with survival that we mentioned in the foregoing paragraph have to do with notions connected with physical and structural perception. Certain physical structural features and perceptual mechanisms in animals have been of very general importance in survival. The generality of a few of them extends throughout the living world and beyond. No more convincing empirical evidence could argue their importance. Symmetry is such a structural feature. One might regard symmetry as a morphism of gravity, against whose downward pull it offers a general protection, at least in regard to orientation. Symmetry is as much a feature of plants, especially the more primitive varieties, as it is of insects, fish and animals. With symmetry clearly such concepts - or sensations - as balance must be closely associated. Rhythm is another such structural feature and has predictive value, creating the expectation of its continuance. There are other phenomena in this class, such as for instance the sensitivity of the human retina to light of different wave-lengths which coincides closely with the distribution of intensity of the varying wave-lengths of the sun's radiation. Taylor (1962) expounds a theory of perception which is of part-

Aesthetic qualities and human characteristics icular interest in suggesting the closeness with which perception is tied to its physical correlates. According to this theory spatial co-ordinates are learned through the references of gravity and self - providing orientation and centrality respectively - and the differentiation of sensory modalities according to responses. His notion is that correct response must precede perception, rather than that perception permits response. Such a theory leads naturally to the idea of 'good' and 'bad' responses and so by association perceptions, and so provides a possible starting point for aesthetics, not far removed from the eighteenth century idea expressed by Hume that a certain natural relationship existed between aesthetic qualities and the constitution of the human mind. The difficulty about theories of this kind is that the primitive correspondences - supposing they exist - between structural features and so forth and aesthetic values are not more than the theory's building blocks. Elaborating these into a structure that will account for all the detailed problems of aesthetics moves the subject deeply into the realms of psychology in its more speculative avatar.

11.4.3 The more philosophically orientated approaches to aesthetics appear fraught with disagreement. Discussion ranges over questions as to what are the objects of aesthetic attention, the nature of aesthetic experience and the grounds of aesthetic judgement, with differing views as to which aesthetic theory should emphasize. Under the first head come questions as to what is a WOA, with suggestions, sometimes based on hedonistic assumptions, that a WOA should provoke aesthetic attention. This naturally leads to the next question, What is aesthetic attention, or, more broadly, aesthetic experience? Some characteristic is sought that renders objects able to sustain aesthetic attention. Without such a characteristic it is argued aesthetics becomes simply a study of subjective taste. The aesthetic attitude is supposed to lead to the apprehension of objects for their own sakes, rather

Problems of aesthetics

than for pleasure of the extension of understanding. What then are the grounds of aesthetic judgement? How does it differ from judgements of taste? Is natural beauty a legitimate object of aesthetic attention? The theory we have devloped suggests answers to numbers of these questions within the confines of its own framework, though it is not our purpose to propose them here. It appears more important that none of these pre-occupations of aesthetic philosophy suggest effective steps that we might take to ensure that an art machine would produce aesthetic objects. We are thrown back on our suggestion of the first paragraph of this section, that if an art machine produces WOAs as we have characterized them, they will have aesthetic value for those Os able to apprehend them. In other words we define aesthetic value as being a quality that belongs to WOAs, and content ourselves with defining these.

11.4.4 Moles attempts to define aesthetic value in terms of such a quality. He approaches the problem through information theory and identifies what he calls 'aesthetic information', though it never emerges quite clearly what he understands by the term. It 'refers to a repertoire of knowledge common to the particular transmitter and particular receptor' - a pragmatic component. 'One may liken it to the concept of personal information'. It 'has no goal properly speaking. It does not have the characteristic of intent; in fact it determines <u>internal states</u>' (pp.129,130). Unlike 'semantic information' it may not be translated - into a foreign language say - nor transposed into another medium. To an extent apparently it resembles what we call the simulatory component of WOAs. Moles suggests ways of measuring aesthetic information, based upon various kinds of methodical destruction of WOAs by known perceptible quantities, 'following the variation in aesthetic sensation, value and knowledge as a function of this destruction'

'Aesthetic information'

(p.201). Unfortunately such a method, while destroying one WOA would be creating another, due to the creative role of the O being tested, as he apprehends the WOA at each stage (N.8.4), and this would no doubt be a factor difficult to control. In another place Moles suggests that aesthetic information is to be associated with the individual A's preferences, expressed, for example, as a 'choice of certain frequencies and certain combinations, to compose spectral symbols, of certain lengths of phonemes, of certain phonemic combinations, etc.' (p.132), rather like what we described (N.7.3.1) as the creative contribution of the musical performer to the performance of some printed score. Similarly, he finds aesthetic information in the orator's tone, semantic information in his message. Overall the concept of aesthetic information appears to add nothing to our formulation, and itself clearly suffers from serious defects that would make it of little effective value in constructing an art machine. For instance, if as Moles suggests, a speaker's words convey semantic information and the timbre of his voice aesthetic, then to use this fact an art machine would need criteria for distinguishing what we might call 'positive' from 'negative' aesthetic information. And finding what such criteria might be is precisely the main problem of aesthetics.

CHAPTER 12

CONCLUSION: PROSPECTS

SECTION 0 THE PRACTICAL PROBLEM

12.0.0 The earliest and most primitive examples of art are cave paintings, many of which, like those of the Bushmen in Southern Africa - some relatively recent - appear to have a frankly practical aim, to promote success in the hunt. At a literal level the scenes portrayed are crude models - in the strict sense that an architect's plan is a model - and, if there were no more to these paintings, we should have to admit, at any rate for them, a second-rate status compared to science. But we may suppose that the paintings did more than merely provide a plan for the study of hunting strategy, that they also offered the occasion for a kind of enactment of the hunt, complete with its fears and dangers and excitements - all quite practical problems for the hunter, though not susceptible to the same kinds of study and planning as the manoeuvres of stalking say. Perhaps the best way to learn about hunting is by doing it, as the best way to learn tennis is by playing it. But instruction from experienced hunters, or 'textbooks' also have an important role in matters of technique and strategy. Only art however provides the laboratory where feelings and experiences may be studied, in forms similar to the ones in which they occur, yet 'in tranquility'. This of course is no more than the basis of a claim for a role for art and should not be taken as a suggestion that its aims are restricted to merely functional objectives, as if its cathartic qualities were its only justification.

Art as the laboratory

12.0.1 On the basis of the conclusions we have reached, one might approach the practical problem of building an art machine from either of two directions.

Simulating a human artist

On the one hand one might argue that significant art could only come from a machine that incorporated about as much structure as a human artist, since a much weaker device would be unable to conceive of art and so could never set out purposefully to produce it. This would imply that the 'best' art machine would take the form of a simulation of a human artist. Similar reasoning underlies the choice of approach in a number of problem-solving procedures, such as most chess-playing programs for instance, which prefer to incorporate human chess masters' heuristics, rather than build up their own from scratch; or financial investment programs that set out to simulate the activity of investment managers, rather than seek wholly novel patterns of share price movements say, to guide them. All such examples reflect not simply a tacit admission of the difficulties and slowness of order extraction, but go further in accepting that, without exploiting the humanly acquired structures that already exist, many procedures simply will not be got to work. But human simulations achieve their successes only by largely sacrificing the goal of novelty. They search for solutions to their problems only those parts of the tree which have been searched before, because it is to those parts that human mimicry leads them. Perhaps solutions from unfamiliar parts of the tree would be incomprehensible, and therefore unrecognizable as solutions to the human programmers. In any case we have shown sufficient reason for supposing, on theoretical grounds, that art of too novel a character would be rejected by its audience, a view art's history amply supports. So this is another reason for setting out to simulate the human artist. But chess is a limited activity, art a broad one. For a machine to play chess it is unnecessary for it to be able to conceive of playing it. The problem with the argument that led us to decide on simulation is that it is self-defeating: a true art machine must be capable of conceiving of producing art, and to do so would have to incorporate much of the apparatus of human intelligence which we are not at present able to furnish.

12.0.2 The alternative approach is more modest, and promises correspondingly less in the realm of art, though perhaps more for the art machine. It entails making the most of available devices, aiming to provide them with heuristics that will lead in the direction we want. We have already come across devices that in various ways model their worlds and their problems. Chapter 6 in particular dealt with examples that had the appearance of being the kind that might extend to meet the needs of an art machine. Winston's program constructed simple concepts and Winograd's robot incorporated an arrangement resembling in some respects what we have called the 'symbiotic' relationship between W (in this case statements in English), and its internal representation of W, W*, represented by PLANNER. The robot's statements do not have the characteristics of true Ws because they do not, once made, then assist the PLANNER language. The incorporation of a facility whereby the robot's statements were recorded, not merely as references for future statements, but as references that might advance PLANNER deductions as well, would remove this defect. But this would still permit production of only very simple and limited kinds of Ws. Greater complexity would demand not merely more output but also a more complex world than the 'blocks world' of the make-believe robot. When a child builds a tower out of wooden blocks we might argue that we were seeing an activity that included as part of itself embryonic creativity. The environment of the real blocks is vastly richer than that of Winograd's CRS robot. In fact the robot's blocks hav no environment, they are the environment. A real robot that constructed a real tower, if its construction took the place of Winograd's robot's statements, in the extended kind of Winograd robot that we have suggested, might be judged creative equally with the child. It is not too difficult to envisage, as a next phase in this development, a robot whose constructions not merely advanced its deductive capacities, but also initiated its next construction, and so on. Clearly error, an inability to handle the blocks perfectly, would advance such a machine's conceptualizing not hinder it.

Extending existing devices

For we may suppose that aiming for a goal not within too easy reach might be the best way of keeping some or other constructional activity going. A real robot with both a constructional and a language facility, able to advance its internal world by both, and able to use both to initiate further statements and constructions, where statements might be about construction and constructions suggested by statements, would be well on the way to being an art machine. One might call it a child art machine, because one might expect it to have about the same kind of relation to its statements and constructions as a child might have to constructions it made. But to turn the machine into an adult would require a further level of 'self-consciousness'. The machine would need not merely to be able to make block constructions and discuss them, but would require to be doing so as part of a wider project, entailing possibly some kind of model of its own internal system. A machine with these accomplishments could probably produce Ws that, for its purposes - which we have defined as a sufficient criterion - might have the characteristics of WOAs. Even such a machine would represent an extensive, if conceivable advance on anything at present available. But its WOAs would still be limited. To extend their quality, breadth and complexity so as to make them acceptable to a human audience would call for enriched contact with an enriched world, an elaborate sensori-motor sytem and, in particular a facility for acquiring socio-cultural experience. For either the machine must itself be able to have experiences for which it may attempt to produce simulation procedures, or it will have to be programmed with or acquire for itself - by order-extraction work - descriptions of experiences. In this latter case it would then face a task roughly the reverse of that of science. For while science attempts to describe what art may only be able to simulate, the machine would have to try to simulate what it had only been equipped to describe.

SECTION 1 PROSPECTS

12.1.0 Was Turing's (1950) prediction of a machine, by the turn of the century, that could succeed in his interrogation game too optimistic? Half the period is over but no clear indication has yet emerged. It is not even obvious how impressed one should be by achievements to date. From one point of view, Winograd's conversations with his robot are startling, from another, paltry, like Faustus and Mephistropheles, one marvels at the devilish power, dismayed at the triviality of what it is called to accomplish. It is clear that computer size and speed are not alone enough. And it is clear too that evolving order within the computer is a problem not to be lightly brushed aside with a mention of helping it along. As Ashby warns us, there is no getting selection for nothing; the Law of Requisite Variety will not be ignored. Perhaps the real shortcoming of Turing's argument is its failure to emphasize sufficiently a holistic view of the individual in his world. The individual is part of a total reactive system with a variety vastly greater than his personal variety. His receptor and response systems link him too intimately with his surroundings to allow total definitive importance to his own 10^{10} neurones. But the computer, locked securely in its cabinet, enjoys only the slenderest links with its world. Nor, until they are greatly strengthened, does it seem likely that increases in computer storage capacities will go far towards compensating for their lack.

Faustus and Mephistopheles

12.1.1 The problem is to estimate what point on the growth curve of artificial intelligence we have reached. Should we look for a growth rate for intelligence of an order comparable to that for the growth rates of so many other phenomena now? In the light of his analysis of animal intelligence, Sommerhof (1969) believes that 'we are compelled to the conclusion that true machine

A limited objective

intelligence is still a long way off' (p.20). This view must justify the wholly abstract nature of the undertaking here that is now concluded. No art machine was ever seriously envisaged as a true possibility at present. The machine was proposed simply as an exercise by which to clarify certain aspects of art theory and to close, within the single frame of organization, art along with all other enterprizes that call upon the resources of intelligence. If we could make a machine capable of producing art, it would probably produce it without the need to provide it with special features. If we were to look for special features, they would probably have to resemble those Beckett mentions (0.5.1), which, through negligence or inefficiency, incline art toward the omission of the duty of Habit, so opening for it 'a window on the real', for our machine, like the human artist, enlarging its world by increasing its responses to its conditions – even at the risk of its own extinction. The alternative, the adequate performance of the duty of Habit, is the prescription for Boredom, with attendant risks of its own, which, as the science of steersmanship, cybernetics has a peculiar duty to notice, if it is not to perish like the unfortunate Palinurus, who fell into the sea in his sleep. Palinurus

> 'was three days exposed to the tempests and waves of the sea and at last came safe to the seashore near Velia, where the cruel inhabitants of the place murdered him to obtain his clothes: his body was left unburied on the seashore.' (Lemprière, cited by Palinurus, p. ix)

NOTES

Beyond a formal understanding of art theory that one might expect from the methodological expedient of making an art machine is the promise of fresh intuitive revelations about the subject. The hope is that these will unfold bit by bit over the whole course of the argument. The object of the notes that follow is primarily to assist this revelatory process by means of examples and explanations taken, for the most part, from art and from the everyday world. Helpfulness towards this object has mainly dictated the choice of examples, and numbers of questions are raised to provoke intuitive insight, rather than with the intention of proposing firm answers to them.

The inclusion of a number of fairly lengthy quotations, many from practising artists, or more particularly, writers, is intended not merely to add corrobative comments on conclusions reached in the text, but also to illustrate directly the concern of practitioners with theoretical questions, and to convey the closeness of many of their notions to those of a theory derived from considerations seemingly quite different from their own. The reason for so far pursuing the ideas of a single artist has been to obtain a view of the completeness of the concern that occurs.

The division of these Notes into chapters corresponds to the text's chapter divisions, and in many cases individual paragraphs in the notes refer to specific paragraphs in the text. However, to allow the notes their own continuity, correspondences of this kind are not mentioned. So far as possible the notes have been written to permit reading them consecutively as a section of the text in their own right, one aimed at furnishing a novel insight, untrammelled by preconceptions and hidden assumptions, into the workings and character of both art and cybernetics.

CHAPTER 1

N.0.0 It should quickly become apparent that our art machine is no merely simple-minded device for arbitrarily turning out objects without interest for any audience beyond what the audience can itself invest in them. As this is the case, it is therefore hopefully superfluous to embark upon a catalogue of things the machine is not. Nevertheless it worth stressing explicitly at the outset that we do not propose to assume that art is necessarily a peculiarly human activity; it is precisely the unquestioned assumption that it is which is one we desire to escape. This wish would in itself justify avoiding at least initially, human-imitative procedures and arguments, like that advanced by Turing (1950). But there is the added reason that the imitation argument itself suffers from an inherent weakness that dissuades against its use,especially in connection with art. This is the common sense objection to it that it depends on 'mere' imitation. Vindication by rigorous cross-examination seems the least owing to 'other minds' by the egalitarian aspirations of non-humans - with evasions like '"I never could write poetry"' (Turing, 1950, p.434) strictly outlawed. Consider for example a machine capable of painting a recognizable scene in some accredited 'style'. Such an imitator, negligently approached, might seem deserving of classing as an artist. Yet its procedures might be quite simple, and easily within present technological capacities. One envisages for instance a mechanical paint brush depositing dots of pigment on a canvas, at points and using such colours as might be determined according to some extremely simple algorithm from a scanned visual image. Pointillist paintings of the image would result. Yet the device would be a trivial example of an art machine, quite lacking in ability to make choices of any kind, as say to style, subject matter, or even picture size. There is no important respect in which such a machine might claim to resemble an artist, in spite of its pointillist pictures, hardly more than, analagously, a punch-card reader might claim to resemble a seeing machine.

N.0.1 What should be grasped at the outset, because of what it reveals about a lot that is commonly taken for granted in discussions of art, is some of the problems that the notion of an art machine poses. Implicit in the concept of artificial intelligence is the idea of machines outstripping their makers. Beer (1972) for instance shows that a machine operating a suitably chosen heuristic programme is likely to take decisions beyond its programmer's comprehension. A point is imaginable, say when an artificial system first became able to construct a system like but 'better' than itself, after which intelligence might grow with increasing rapidity in successive generations. From a system with human-like intelligence, the evolution of systems vastly superior to humans is at least conceivable. Questions as to how such evolved systems might manifest their superiority do not appear, on the face of it at any rate, to present any particular difficulties, seeming to call for answers of the same kind as questions about human intelligence. Such systems one supposes would undertake the tasks of human intelligence with greater and steadily growing efficacy. But would it look like that to weaker systems? The question is obviously relevant in a wide range of contexts, particularly where there may be conflicts of strategies. (Answers could probably be developed as the basis of a justification for Leibniz's assertion that God had chosen to create the best of all 'possible' - in the sense of logically allowable - worlds.) Here the object of raising the question is simply to provide an analogy by means of which to interpret the equivalent question relating to highly artistic rather than highly intelligent systems. Bowden (1958) suggests that a sonnet written by a machine might 'appeal only to another machine' (p.321), though he makes little effort to understand why, or even what might be meant by 'appeal'. But it is still pertinent to ask what kind of output a highly evolved artistic system might have, and how lower order systems - humans say - might interpret it. At first sight it appears that the kinds of objective criteria suitable for recognizing and perhaps even for rating intelligence in artificial systems might be useless for dealing with art, where the grounds for evaluation are even less secure than they are for intelligence. But it is suggested and we hope to show in the course of the argument, that the divisions separating the two kinds of assessment are largely artificial, and positively obstruct proper understanding of the problem.

N.O.2 The strict separation of reason from imagination ('analysis' from 'originating') results in anomalies like for instance the expressionless speech attributed to the robots of science fiction. Vocal expression is supposed to be connected with 'feeling', which is distinguished from 'meaning' and, so popular belief goes, machines can deal with the one but not with the other. Insistence upon such differences would deny existence to the subject matter of pragmatics. Vocal expression itself conveys meaning, and part of a speech's meaning is what it tells of the speaker's feeling. The written word has to make good vocal expression's absence by means of greater precision, as well as by the use of various devices and conventions such as carry information about 'mood', 'tone of voice', and so on. A computer candidate hoping to compete successfully in Turing's (1950) 'interrogation game' would certainly need command of actual voice tone and expression, or at least the skill to use existing conventions for investing its written responses with these qualities. For human-like intelligence displays capacities to reason and to use imagination and intuition, as well as the ability to grasp meanings conveyed by others through the use of different tones of voice and the like, all exercised together. Meaning and feeling are not separable. Both belong to the realization of an intention.

N.O.3 In art it is clearly possible to appreciate merit, without knowing the artist's intention, as for example in the case of ethnological objects viewed by someone ignorant of the culture from which they sprang; or to 'see' in a WOA something more than, or different from what the artist had intended.

The creative role in such cases is clearly at least partly filled by whomever does the seeing, or exhibits the 'unintended' meaning to him. The case of 'found objects'* is another example of this kind. Such objects show no artistic intention. The object's finder fills the artist's role, providing an 'intention' by discovering its outcome in the object. But clearly these remarks apply in some degree to the audience's role in relation, to any WOA, and in particular to the artist as his own audience.

N.0.4 Relating to the artist as author - as distinct from audience - of his work are questions of the meaning of 'creativity' and the 'creative process'. At every stage in the making of a WOA the artist is likely to alternate between both roles. As audience, his concern is judgement, interpretation, meaning and so forth; which he may weigh against his intention as author. For this reason intention should not be taken to imply anything either explicit or fixed. Each time, during the process of making a WOA, that the artist reviews what he has done so far, he engages in two activities: he assesses his relative success or failure in realizing, by means of the object, which is called the WOA, a state of mind that may be called his intention; and he modifies or alters the intention in the light of the object so far actually produced.

N.0.5 Extending the formulation of the reflexive structure of creativity to an art machine raises interesting questions concerning the character of intention and judgement. This suggests the possibility of degenerate cases in which one or other was missing. In the case of the machine lacking judgement, would it be meaningful to speak of intention at all, when the machine would have no way of deciding how far, if at all, its efforts represented the intention's realization? An observer discovering an intention (N.0.3) might attribute it to the machine. More probably, if

* For a formal definition of this term see N.8.0

the machine appeared to the observer to be operating in a 'mechanical' way - which is likely, as such a machine would be acting without showing any 'awareness' of what it was doing - he would attribute the intention to the machine's maker. If such a 'mechanical' machine suddenly appeared to develop an intention, an observer would probably suppose that the intention had entered it by way of a controlled input, to which the intention properly belonged. The second degenerate case, that of the machine lacking intention with respect to its output and consisting simply of an evaluating facility, would be the equivalent of a WOA's audience.

N.0.6 A machine with outputs interpretable only by itself is conceivable as another degenerate case. Such a machine might simulate a dreamer rather than an artist, though much that will apply to mechanical artists is likely to apply equally to mechanical dreamers.

CHAPTER 1

N.1.0 The two concerns of chapter 1 are the paradigm for production and its interpretation, and the evaluation of its products. As these products may be altogether exotic, their assessment needs to adhere to an explicit method. Nothing supposed in the paradigm for the productive process AW implies any restriction on the characteristics of its elements A, W, &c., beyond the needs of A to be able to communicate with M, and M with E. Specifically there is no requirement for A or observers of W to be human. Provisionally W denotes any product whatever of AW, whether it is a poem, painting or musical recital, the solution of a problem, a theory, a strategy or an object or assemblage excluded from human observation or even understanding. Ultimately we wish to assign AW the characteristics of the 'artistic process', in which A is the artist; M his craftsmanship, the artist's technical skill at operating on his medium; E his world, including his medium; and W a WOA of which A is the author. Extending this interpretation, O becomes W's 'audience', and in conformity with the general properties of the other elements of AW, may thus be any suitable entity in U, human or not. Partly this suitability will depend upon W, since not every W will be open to human observation and some minimal conformability of the various elements is clearly called for.

> 'I realised that it is not only the material world that is different from the aspect in which we see it; that all reality is perhaps as dissimilar from what we think ourselves to be directly perceiving; that the trees, the sun and the sky would not be the same as what we see if they were apprehended by creatures having eyes differently constituted from ours, or, better still, endowed for that purpose with organs other than eyes which would furnish trees and sky and sun with equivalents though not visual.'
> (Proust, 1957, vol.5, p.83.)

N.1.1 The need for an objective procedure for deciding what qualifies as a WOA is not merely fanciful, nor a prollem confined to some shadowy boundary area dividing WOAs from the wider class of Ws. As long as art remained within the confines of readily apparent formal structures, such as the representationalism of painting and sculpture, melody and fixed formal patterns in music, prosodic rules in poetry, and so on, the search for some essential quality common to all forms of art that the right decision procedure would discover seemed reasonable. But the widening of artistic interest beyond the boundaries of Europe to the artistic achievements of other ages and cultures, and the relaxation of the hold exercised by the traditional forms, has steadily eroded the plausibility of prescriptive definition. We have seen (0.3.1) some of the difficulties that have re-resulted for art theory from the need for a changed approach. Adopting a method that relies on abstract entities implies that to deal with these difficulties, it is necessary to step back from too close an involvement in the subject matter. Such a view is of quite general application.

N.1.2 Problems of finding evaluative criteria are not confined to art. 'Value judgements' generally are subject to difficulties resulting from confusion between the two sets of procedures of O and $\bar{O}$. Moral judgements are peculiarly susceptible to these anomalies. Moral judgements are those made by a meta-observer $\bar{O}$ in which the Ws are ways of behaving or sequences of behaviour, but often fail to distinguish $\bar{O}$- from O-assessments. Failure of a similar kind leads to identification, especially among children, with the plots and characters of novels, plays and the like. Here $\bar{O}$, the reader or audience, slips into the role of the character in the novel or whatever, who may, in that context, himself play the part of O, vis-a-vis the other characters. Art, as we shall see, greatly relies upon this happening, and often deliberately sets out to make it happen. Conflicting decisions between O and $\bar{O}$ correspond in art to confusion in artistic judgements. Thus the critic's belief that others should accept his evaluation of a W is an example of $\bar{O}$ using as a criterion a decision reached as O. Being 'influenced by' others' judgements is an example of the reverse. A single entity - in practice always human - expresses his taste as O and his judgement as $\bar{O}$. Taste relates to O's experience of his acquaintance with W, judgement to $\bar{O}$'s, experience of this experience. Minsky's remarks (0.2.2.0) about the inadequacy of the notion of the relation 'contained in', when applied to information processing structures, may provide a basis for an understanding of the reasons for confusion and the difficulty of avoiding it.

CHAPTER 2

'Only connect' E.M. Forster

N.2.0 As will appear in the following chapter, the foundation for a more rigorous definition of authorship is now layed. Complex 'things' - which may or may not be WOAs - pre-suppose complex purposes, and at least equally complex agents to entertain the purposes. These agents in turn must thus also be the outcome of purposes, and so their agents and so on. Complexity does not spring into being at a creative stroke but evolves slowly. The paths of evolution and examination of some limits on its rate will form successively the subjects of the following three chapters. Here we wish to emphasize the implications for art of the connexion between structure, purpose and the notion of antecedents as causal explanations. First, it appears that any object interesting enough to be called art presupposes a structure at least as evolved, both in the artist and, by implication, in his audience. Secondly, explanation's evolutionary character, both its formal and intuitive aspects, and the other parallels that are seen to exist between these two modes - the latter which appeals to 'things' as explanatory principles, rather as formal explanations appeal to theories - point to similarities between art and science that are not always intuitively obvious.

N.2.1 (2.2) The notion of producing any work is seen from the outset to be bound to control, control to purpose and purpose to intelligence. Fogel & others emphasize that intelligence must involve a '<u>wide variety of goals under a wide range of environments</u>'. 'Certainly, without the existence of a goal, decision-making is pointless and the term "intelligence" has no meaning' (p.2). Ashby (1962) also insists on largeness of scale: an organism that evolves in an 'isolated' system must be able to go to an equilibrium involving only a 'small fraction' of its overall number of states, which must yet be of a sufficient number to permit 'a good deal of change and behaviour' (p.273).

N.2.2 (2.4) The parallels between organisms or theories evolving to completeness and consistency is obvious, and intuitively may often seem quite plausible and natural, though in the theory of art it is more often claimed than plausibly demonstrated that WOAs may have organismic features. One of the principal ideas of this chapter, consonant with its cybernetic aims, is to elucidate the correspondences between organisms, explanations and ideas, and their evolution. The connexion between explanation and WOAs will be taken up explicitly only in Part II. Here we merely observe the evolutionary process as it occurs in its several incarnations.

> 'Thus the empty spaces of my memory were covered by degrees with names which in taking order, in composing themselves with relations to one another, in linking themselves to one another by an increasingly numerous connexion, resembled those finished works of art in which there is not one touch that is isolated, in which every part in turn receives from the rest a justification which it confers on them.' *
> (Proust, 1957, Vol.6, p.314)

The 'finished' WOA is evidently a kind of 'knowledge structure', (cf. Gagné, 1962). What Proust describes is clearly similar in many respects to Pask's <u>entailment structures</u>, which are closed 'relational nets'. This is expected, as WOAs are simply a sub-class of the wider class of Ws, and knowledge structures are morphisms of effective explanations, procedures which themselves correspond to the processes AW from which Ws emerge.

N.2.3 (2.4.3) Less obviously it appears that value judgement is another feature common both to Ws and WOAs, for it is intrinsic to evolution. 'Good' and 'bad' - or their equivalents - are indispensible labels, distinguishing between mere 'self-connecting' and 'self-organizing', from which scientific theory is no more exempt than art.

N.2.4 (2.6) Explanations evolve towards shortness and simplicity.

'"Interesting" functions usually have features that permit enormous reductions in the sizes of the nets required to realize them.' (Minsky 1972, p.55). The nets, descriptions in Solomonoff's sense (2.3.0), starting tapes for

---

* Proust's observations concern a dinner at the Geurmantes when the names of the guests make links in his mind with the French and European history with which the names are connected.

Turing machines, may all be thought of as having information potential, which becomes an information flow when these are realized in operation. 'The overall scheme of a complex, multimolecular edifice is contained in posse in the structure of its constituent parts, but only comes into actual existence through their assembly' (cf. 'Step-by-step' effectiveness, chapter 9). 'The epigenetic building of a structure is not a creation; it is a revelation.' (Monod, p.87) (This suggests an answer to the question posed in 0.3.1.6 as to whether art is creative or revelatory.) Extending explanations (2.8) by 'joining' them, so that they may generate larger runs of information flow, might in these terms, be thought of as 'adding' potentials - perhaps in ways analagous to the composition of mathematical functions, the addition of vectors or procedures for incorporating computer programmes as sub-routines in larger programmes. Lofgren's definition of order (2.3.2) or Solomonoff's assignment of a priori probabilities to strings of symbols according to the way they might be produced by a universal Turing machine, suggest possible bases for information potential measures.

N.2.5 (2.7) Thus U-forms with lengths that were small compared to the lengths of the sequences they generated, or high a priori probabilities of strings generated by a Turing machine would measure high information potentials; and conversely. Of interest is the evident aesthetic value of information potential. Sensitivity (among human subjects) to form and pattern of high or low information potential may be extremely acute. Moles observes that an intelligible auditory message requires a noise level four to eight times higher than its fortissimi to destroy it. 'Intuition' and conjecture are frequently more sensitive detectors of pattern than available statistical techniques. High information potential assumes aesthetic value. Notions of elegance in mathematical proofs, as in design, are closely linked to brevity and economy. ' ... "beautiful" often means short in the sense that we have discovered some abstract concept of invariance, that permits a short recursive description.' (Löfgren, 1967, p.167) Images, including metaphors may have a high information potential. For they have the information potential of descriptions raised,as it were, to a higher power, possibly a number of times, by morphisms between the domains the images describe and other, analagous, but distinct domains. That is, an image is a description of more than one thing, by virtue of pointing to morphisms between the things. This 'evoluted' information potential provides greatly heightened aesthetic value. ' ... the image is not an idea. It is a radiant node or cluster; it is what I can, and must perforce, call a VORTEX, from which, and through which, and into which, ideas are constantly rushing.' (Pound, 1914, p.469) Pound's 'node' is an image, but Minsky's 'net' is closer to a literal reality. Though the value of the concept of a net in automata theory has probably been at least as great when it has been an image as it has when an actuality. A node with numbers of connections in a real net of the kind Minsky means, or in some conceptual net such as a knowledge structure for example, would in reality have qualities like those Pound mentions.

N.2.6 Very much the object of 2.7 is to indicate the stubbornness with which regularities persist. O does not believe in purpose arising spontaneously, nor in regularity existing in isolation, and part of the aim here has been to show that O has reasonable theoretical grounds for his prejudices. One might take as a simple model of O's views a chain of related regularities with purposes linking them. (For a more detailed view of this notion see the examination of chapter 7, which should provide a reasonable plausibility argument for present assertions.) A more complex model entails a net in which purposes join nodes of regularity.

> 'Since there are flowers whose fertilisation is impossible except by means of an insect, flowers which eat insects and therefore understand them, since so low and unconscious an order has these correspondences with the one above, may there not be animals and birds who make use of man and study his habits and if they do, why not insects and vegetables? What grape, to keep its place in the sun, taught our ancestors to make wine?'
> (Palinurus, 1961, p.13)

CHAPTER 3

N.3.0 The most important idea to grasp from this chapter is that a WOA is seldom likely to be a product purely of c-authorship. Indeed, by the arguments of the preceding chapter, this could be the case only if A, or that part of it responsible for W, had come into existence more or less spontaneously and without antecedents. As an actuality, such a state of affairs is next to inconceivable. What appears to occur generally in reality is almost the reverse of this. In the case of art, first and foremost the artist, as a human, is a product of millions, the culture in which he develops, of tens of thousands, of years of evolution. Thus, his art, his technique, his medium, the work of particular artists, all contribute to an overall structure to which he may add a little. What he adds will be his creative effort. Because of its relative smallness, it may be difficult to assess. This is not peculiar to art but is also the case in science. As in science, 'the order which the scientific discoveries reveal will be talked about not only by the scientists themselves but also by their colleagues in other fields, and eventually will be taught to a new generation of scientists. In this way, a potential is created for asking deeper questions about the systems revealed, and finally new or related orders and systems may be found.' (Löfgren, 1972, p.342). Science is frequently distinguished from art on the grounds of its progressive character. The view we have worked towards does not show this distinction. If science and art both rely on the use of developed forms, both must have similarly progressive natures.

> 'I was led to ask myself whether there was any truth in the distinctions which we are always making between art, which is not more advanced now than in Homer's day, and science with its continuous progress. Perhaps, on the contrary, art was in this respect like science; each new writer seemed to me to have advanced beyond the stage of his immediate predecessor; and how was I to know that in twenty years time, when I should be able to accompany without strain or effort the newcomer of today, [cf.461] another might not appear at whose approach he in turn would be packed off to ... limbo...?'
> (Proust, Vol.6, pp.22,23)

What gives science its apparently progressive character is the impression of the almost pre-destined course of its development; an impression gained from events like the very nearly simultaneous but independent invention of the calculus by Leibniz and Newton for example - though such coincidences might be explained on the basis of Spencer-Brown's observation that proofs are found for those theorems that lead somewhere. The review of scientific development commonly tends to emphasize the convergence of whole domains of knowledge into some single new 'discovery' (heuristic; explanation) and to neglect the huge new vistas that such discoveries open up. Among great discoveries one would include Mendelyev's Periodic Table, the Theory of Evolution, Bragg's Law, Maxwell's synthesis of magnetism and electricity. After such discoveries, the problem of finding alternative explanations for the knowledge they draw together is even more difficult than that of coming by the explanation offered in the first place. An alternative to the Theory of Evolution for example, which would 'explain' the same set of data as the Theory, would need the same order-extraction effort as was needed to create the Theory, plus an additional effort to overcome the 'blinkering' effect (6.2.2) the Theory has, once created. This is one reason for the tendency to emphasize the convergent aspects of scientific development.

But what leads away from discoveries is highly divergent (see N.5.0). Without Bragg's Law for instance the discovery of DNA would probably have been impossible, yet this discovery would have been hard to predict, even in approximate form at the time the Law was first enunciated. Art does not take along all its knowledge at each new departure and is often contrasted with science in this. But with science too, much becomes useless and is discarded. As we shall see in the following chapter, there is no escape from the difficulty that results from having to build new order upon old. In art or science, order extraction is limited by the relatively low variety of the processor that does the order-extraction work, compared with the high variety of the surrounding from which the order is to be extracted.

N.3.1 This conclusion raises a question as to the interpretation of W, in particular of a WOA; namely how much difference is tolerable between the order accumulated in A and in O (in his role as audience), before the W A produces ceases to be a WOA, or even distinguishable as a W by O. How much creativity is an artist allowed? Chapter 1 dealt in part with this question, though on the assumption that W would be distinguishable. When this is not the case, the argument of chapter 1 fails, but then the question never arises. But even in the absence of bizarre differences between A and O - when both are human say - what are the chances that the WOA A produces will receive greatly varying interpretations from the two individuals? The question is clearly wide and outside ths scope of the present argument. Here we restrict ourselves to observations on the effects of O-A differences on taste, interpretation and expression.

N.3.1.0 Moles asserts that the size of the public interested in a concert varies inversely with the concert's originality, (where he defines the programme's originality H as $H = -5\ p_i \log_2 p_i$; -5 derives from five categories of composers grouped in terms of frequency of performance; and $p_i = p_w \times p_c$, with p_c being the composer category and p_w the composer's work category). Implicitly this defines public taste in terms of frequency of performance. In a sense such a definition is circular, since it defines taste in terms of audience size and makes audience size depend on taste. But the definition also suggests that people like what they know. Knowing the music increases the audience's capacity as a receptor, so raising what we have called the information potential of, in this case, the concert. And we have already noted (N.2.4) the aesthetic value of information potential. 'Any study of the value or quality of a message' - taste - 'must be based on the capacity of the <u>ultimate</u> receptor' (Moles, p.19).

N.3.1.1 Clearly this approach raises questions of the artist's and his audience's education, socio-cultural milieu, and so on. Misconstruction is possible at many levels. Intelligibility at one level does not ensure it at another. Karen Blixen (1964) records telling the story of <u>The Merchant of Venice</u> to her Somali servant in Kenya. The Jew should have used a red-hot knife, the servant argued, it draws no blood. What about the pound of flesh, neither more nor less? He could have taken a little at a time.

> 'But in the story the Jew gave it up.'
>
> 'Yes, that was a great pity, Memsahib.' (p.279)

N.3.1.2 For the artist such difficulties raise problems of expression: how much creativity will destroy intelligibility? Many artists, especially in this century, have been content to ignore such considerations. Others, without either the intention or the wish to do so, have fallen foul of their own creative achievements, only to be vindicated by later generations. In any event, the problem of the balance between novelty and banality remains interesting. According to Moles, intelligibility varies inversely as originality. The difficulty of transmission of a message increases with its information content. '... a proper balance between banality and originality is now a cornerstone of aesthetic perception' (Moles, p.vi) To pursue this question (5.3.6) we shall want a more detailed understanding of what choices the artist has. Montaigne quotes Horace: 'I strive to be concise/And grow obscure'. Here we observe that these problems vanish only if A and O are identical individuals, that is if O and A are subsets within the same set of procedures, both drawing upon the same data base. In chapter 8 we shall show that a structure of this kind is not simply desirable but actually necessary for any art machine.

N.3.2 If for the present we dispose in this way of problems of audience and interpretation that results from differences in the accumulated order of A and O, the question of A's 'creativity' remains. Newell & others define creativity within the context of problem-solving. This is not in itself restrictive, as any purposeful activity might be defined as problem solving in which the problem was to achieve the purpose. However the four conditions they list as sufficient to characterize creativity in general are wholly based on experience of human creativity. The conditions are: (1) 'novelty and value' (for the thinker or his culture); (2) unconventionality, achieved by modifying or restricting previously accepted notions; (3) the need for high motivation or persistence 'over a considerable time' or 'at high intensity' (p.4); (4) part of the problem consists in formulating the problem itself. The notion of order extraction covers the first two of these criteria. The way the authors stipulate them simply adds to our terms a vague and arbitrary requirement as to the quantity of order that must be acquired before the acquisition qualifies to be called 'creative'. The third criterion is psychological. While it is likely to be a feature of human creativity one would not regard it as a necessary one. The creative act might be quick and depend entirely upon previously programmed (in our terms) order that, by definition, did not depend upon prior activity. Nor would one suppose the condition sufficient. Neither is the fourth criterion sufficient to justify giving the name creative to thinking that required it, since re-formulating a problem does not necessarily solve it. But it is probably a necessary condition of creative problem solution; for a problem stated in such a way that it needed no re-formulation would contain its solution <u>in posse</u>, rather as the starting tape of a Turing machine contains <u>in posse</u> the end tape it generates. Fogel & others liken 'creativity' and 'imagination' to artificial intelligence and suggest that, given these qualities, 'it is possible to seriously consider programs that would provide the machine with self-awareness and, ultimately, with an ability to select its own goals' (p.123). So, if this were the case, we should argue from our proposed definition of authorship that self-awareness would be one of the characteristics it determined.

N.3.3 Our definition's chief advantages are, that it is based upon a rigorous conceptual foundation, and that it recognizes the connexion between structure and purpose, and so, the inevitably self-reproductive aspects of purposeful productive activity, and the overriding homeostatic purpose concerned with equilibrial states and the invariance of identity. For art theory the definition is satisfactory. It provides a framework within which to examine the qualities 'we demand in our sensations' from an 'object of art', which

'will be order, without which our sensations will be troubled and perplexed, and the other quality will be variety, without which they will not be fully stimulated ... there is something more - there is the consciousness of purpose ...'
(Fry, 1937, pp. 32,33).

CHAPTER 4

N.4.0 Chapter 4 takes up the theme of the difficulty and slowness of order extraction from a point of view that differs from the two chapters that precede it. The implications for an art machine of the chapter's conclusions remove some of the difficulties that earlier chapters had revealed, though at the cost of limiting the scope of possibilities of what we may hope to construct. The results of order extraction may be used as heuristics to advance further order extraction: c-authorship fuels itself. But finding heuristics is slower than it often may seem, simply because habituation conceals how much structure many common heuristics - like Beer's dog-training heuristic (4.4.1) for instance - may incorporate. Moreover, because of the slowness of order extraction, any art machine will have to depend for its initial structure upon what its maker is able to provide. As the maker will have to rely for this on his own structure - what he has been 'programmed' with and what he has acquired (N.3.0) - the art machine is bound to resemble its maker in what it produces. Chapters 5 and 6 explore this further.

N.4.1 ' ... the codification of a proof procedure, or any other directive process, although at first useful, can later stand as a threat to further progress.' (Spencer-Brown, p.10) In other words it is not just error in the normal sense to which heuristics may lead but, more broadly, to determining the path of evolutionary development. (See notes to chapter 7) The action of the mechanism that leads to this determination is what also conceals it from us, precluding awareness of the part our received structures play in determining what structures we discover and, more particularly, what we do not. Pask, Scott & Kalliakourdis (1973) examine the dependence of problem solutions on problem structures. In heuristic problem-solving programs the problem is largely solved before the outset of the 'solution' by the program's operation, either because the problems themselves are embedded in highly structured domains - chess, theorem-proving or the like - or because the author has already selected what he considers relevant to the solution in specifying the problem. And his selection may entail a huge reduction in variety between the actual problem and its specification. The alternative, a system that had 'not evolved to match the world into which it was born' (McCulloch, p.221), would have to acquire its own heuristics and face the slow task of reducing the world's variety to manageable proportions.

N.4.2 The author is bound to feel ambivalent about heuristics, which increase his power as a c-author on one hand while they diminish it on the other. The author cannot escape from old, long-developed structures, his human form for example. But at the frontiers of authorship there is some opportunity for him to choose his own forms (heuristics). Here one of his problems might be communication with an audience. For c-authorship is innovatory by definition, and therefore a departure from, a 'creation' added to, the world of the author - and so, a fortiori, his audience. The question is, What degree of departure is permitted an author before his message becomes unintelligible to an audience? But the author's problem is only secondarily one of communication (Notes to chapter 7). As heuristics are really variety reducers, his deeper problem remains that of choosing them, in so far as choice is possible. The author must strike his own balance between doing more with less and less with more.

N.4.3 Art (including music and literature) provides examples both of authors drawn towards forms and turning away from them. With the modern abandonment of traditional forms (see notes to chapter 7), artists have sought new ways of reducing variety. 'Mandarinism' in literature, pre-occupation with the medium over the subject in painting, are possibly examples of attempts to escape from variety, as for instance are such common new expedients as the use by artists of photographs to provide preliminary studies. Overall the 'space' for choice is narrow. According to Moles the gross redundancy of French is forty-five percent; redundancy based on letters is nineteen percent. 'The difference is due to the overall structure of the language ... the impact of thought on language takes place at this interstice'. (p.46)

CHAPTER 5

'... one of those perfect wholes
that it takes centuries of time
to produce ... (Proust).

N.5.0 The considerations of this chapter impose strict limites on the scope of any art machine to be a c-author. For it follows from our argument that an art machine capable of producing an output such as we should be able to distinguish (from randomness) would have to rely heavily on structures with which we provided it. Unfortunately, this is the only kind of machine we should be able to make with an output rich enough to be of interest. We should be mistaken though to over-emphasize the importance of this restriction. First by definition it is no more severe than the restrictions on the c-authorship of every human artist - or on every form of human activity. Theoretically at least the restrictions on the machine should represent a relaxation of human limits, depending as they do on systemic considerations and not on physiological or physical ones, connected with the materials of which the machine might be built.* Secondly, the restrictions that the need to 'pass on' structure implies, are themselves flexible, standing as points of departure from which there is a wide choice of directions in which further exploration may embark. Chapter 6 examines the operation of a few structures in greater detail.

N.5.1 (5.3.6) Restrictions on c-authorship are a problem for any artist (cf. N.3.1.2): how to make use of others' efforts without stifling originality. On the other hand 'power flows through the forms, [in this case, of poetry, though the reference could equally have been to music or painting] so that to study them is to plug oneself in to a source of imaginative strength'. (Wain, 1975, p.50) Form 'is a source of strength, of subtlety, of awareness' (ibid. p.56). But form uninvigorated by imagination becomes banality, habit, which 'Of all human plants, ... requires the least fostering, and ... is the first to appear on the seeming desolation of the most barren rock' (<u>Remembrance of Things Past</u>, Marcel Proust, cited and translated by Beckett, 1970, p.28) (cf. 0.1.5). Wain's words are a plea for forms that comes late in a cycle which began with a rebellion against them: ' ... poets in our civilization, as it exists at present, must be <u>difficult</u> ... The poet must become more and more

---

* We must remember that systemic structure, in the sense in which we have been discussing it, is bound to be largely inseparable, in a quite fundamental way, from physical structure. This is so because conceivably there could turn out to be important systemic structures, attainable <u>only</u> by way of certain physical (atomic) events. Altered, these physical events might produce corresponding alterations in the logic of the system of a kind that could be neither avoided nor nullified by other atomic events - because of the structure that is already a part of physical atoms. Theoretical equivalents between McCulloch-Pitts neurones and natural nervous systems do not furnish grounds for supposing that natural neural logic might not have differed had natural nervous systems developed in some other form.

comprehensive, more allusive, more indirect, in order to force, to dislocate if necessary, language into its meaning' (Eliot, 1951, p.289). 'Difficult' poets will rely less on form, and it is an easy step that brings a shift emphasis from Eliot's positive injunction to a negative abjuration of form itself; which, in the end, Wain argues, is impoverishing, as our own conclusions would make us expect. '... it is not given to any human artist to be so original that he breaks free of <u>all</u> influences and still remains an artist...' (Wain, p.54) In other words, the choice, as we have suggested, is between retaining much of the received structure - most is in any event too deep to reject - and suffering a degradation of creative power. Complete individualism is impossible, there is only 'a choice of different categories of alignment' (ibid p.50). (See also Notes to Chapter 10)

CHAPTER 6

N.6.0.1 (6.1.0) Artists' use of descriptors, comparable to the ones discussed in this chapter, may be explicitly called for: di San Lazzaro (1949) cites Cézanne's precept, '<u>Traitez la nature par le cylindre, la sphère, le cône</u>'. More often they will be explicitly present in WOAs, without verbal vindication. Examples are numerous: the match-stick figures of Klee; the shapes of Kandinsky & Miro; the <u>descriptive</u> rather than representational figures of 'primitive' artists and of children: painting in which 'importance' rather than perspective dictates size, as in Japanese landscape painting. But to some extent the very possibility to 'depict' with paint or, more obviously, by drawing, depends upon descriptors, more complex and numerous perhaps, but of a kind similar to those Guzman uses. Representationalism depends in the first instance on knowing how to do it, having the right heuristics and descriptors; Byzantine painting is without perspective. Giotto 'discovered' perspective but much of his work shows that he did not fully understand its use, perfecting it came to fascinate renaissance painters. Later European painters take it for granted and appear to lose awareness that it is dispensible. Habit has concealed the effort that discovered it in the first place. The modern movement seeks to throw off its constraints, but vestiges cling until after cubism, until in fact Clement Greenberg (1947) announces that Flatness is painting's aim and the 'integrity of the picture plane' comes to dominate modern painting (cf. N.5.1)

N.6.0.2 Descriptors in literature and music are less obvious. Music is an 'abstract' form, which means that it is not in general part of its aim to establish direct links between itself and the 'real' world. Literature uses words, which are already descriptors of a kind. Their relation to other kinds of descriptors, such as those of the SEE programme emerges in the use Winograd makes of them. But literary forms are not restricted to words alone. Symbols, in the literary sense, may be descriptors of a kind. Moles calls them forms which 'may be normalized and repertoried' (p.59). Pound (1914) likens them to images (cf. N.2.4) only where 'symbols have a fixed value, like numbers in arithmetic, like 1, 2, and 7 ... images have a variable significance, like the signs a, b, and x in algebra' (p.463.). Prosody and poetic forms are not descriptors in our present sense. Nor is form in music. Discussion of the function of these forms must wait for Part II.

N.6.1 (6.3.2) A's problems of expression and communication, or O's of interpretation are partly alleviated by O's position, 'between' A and W, somewhat analagously to PLANNER's position 'between' the 'blocks world' and the syntax of Winograd's language program. Just as Winograd's program is able to economize on syntactical rules by importing structure from the 'physical' world of the blocks, so O is able to interpolate knowledge he may have of A to fill gaps in his understanding of W. Such a procedure enables O to distinguish A's intention from his execution, simply by comparing his knowledge of A, or possibly even the techniques A has employed, with W. O progresses like a man climbing a rope using first his feet then his hands in alternate extension. Chapter 2 intimated some such process when it showed (2.7.3) how O distinguished W or A and then used it to discover respectively a matching A or W. The effect is analagous to that of a ratchet: O fixes himself to A as he scales W as far as he can reach, then he fixes himself to W and scales A. In chapter 9 we shall discover similarities between this technique and the conjectures and refutations described by Popper (1963) by which, he argues, science progresses. In Winograd's programme, W is the 'blocks world', in scientific progress, W is what Popper calls 'objective knowledge'. Chapter 10 will argue that WOAs are Ws with a similar rôle, though in a different sphere.

CHAPTER 7

N.7.0 At best any art machine we can make will resemble the human artist; that much has emerged. Osborne remarks that 'it is always necessary to regard the work of art as an intentional object, created by a human being within a cultural milieu and subject to specific social conventions' (p.23). The cultural milieu and the social conventions are a part of the WOA's structure because the artist has deliberately used them as part of his heuristic repertoire. He musters whatever structure he can from himself and his history - specific and personal - in order to lengthen the strides of his own creative progress. But doing so he pays a price in the narrowing the limits of the kind of creativity he may achieve; and incidentally, the audience he can reach: Moles notes that meaning is not 'transmitted'; it pre-exists the message, resting on a set of conventions, which receptor and transmitter hold a priori in common. (cf. what Monod says about 'creation' and 'revelation', N.2.4) The art machine will not escape the human artist's dilemma; it will be no more able to do without cultural milieu and social conventions than the human, not unless it reduces its own capacities. The artist's and the art machine's goals and goal-generating potential - creativity - extend both backward and forwards in time. The relevance of chapter 7 is in tracing some of the aspects of this extension from the goal's inception, or conception, back towards its origins and forward towards its execution.

N.7.1 In the creative process AW we may characterize the generation of A's goal as imagination, inspiration, creativity. For A the artist, the goal is the idea, which is part of him, distinguished from the thing he makes, which exists separately from him. His skill at manipulating his medium enables him to 'translate' his idea into something not himself, when it becomes the object of his own and others' contemplation. (cf. Fry, 1961) As a paradigm, AW clarifies the distinction between the artist's conceptual and imaginative skills and his technical skills, which the single term artist confounds. The artist's technical skills are the means by which he makes an observable thing, those skills denoted by the word craftsmanship. A's goal corresponds to what Collingwood (1963) calls the 'work of art proper', which exists 'in the head' of the artist. Whatever the merits of this idea, it is clear in its emphasis of the importance of the artistic creativity, the c-authorship content of the WOA. But Fry's demand appears no less important. However secondary or 'incidental' Collingwood may believe the 'bodily or perceptible thing' that we call the WOA, without it there can be no way of knowing, even for the artist himself, whether the contents of his head amount to art, or simply some trivial banality, heightened in his estimation by euphoria. As we shall show in chapter 10, the WOA represents 'objective knowledge' without which no A can be more than a dreamer.

N.7.2 Presented with a WOA, we look for the artist, for evidence of c-authorship, and will even withold or assign the name WOA together with its associated qualities according to the degree of c-authorship discernible. In the limiting case for instance a painter copies someone else's canvass. Making an exact copy requires him to use the same kind and mixture of paints, the same quality and size of brush, the same kind of canvas, of brush strokes, &c. Indeed it is probably to gain understanding of these features of another painter's work that an artist decides to copy his canvas in the first place. A perfect copy may lead us to admire his skill, though we should not think of him as the artist any more than we should a computer that reproduced a picture according to some relatively simple process, by scanning an existing one say, and translating its light intensities into a pattern on a CRS or some kind of print-out. (Yet pictures made in just this way have been described as 'computer generated'! Harmon & Knowlton, 1968). The WOA must exhibit the artist. We wish to recognize the painter in his portraits of different sitters, the actor in the character he plays, the performer in his rendering of a piece of music.

N.7.3.0 (7.4, 7.5) But the credit we accord an author also depends on our knowledge of his problem, of what he set out to accomplish. A move on a chess board may win aesthetic approval though only from one who understands chess. O assesses the solution against the problem and takes both into account in arriving at a judgement of the author. In cases of shared authorship this acts to diminish the degree of c-authorship that O is prepared to assign to the various partial authors. This is what happens in practice. Paragraphs 7.1.1 and 7.1.2 provide the practical situation with a theoretical framework, while suggesting grounds on which c-authorship might be apportioned, roughly according to the scope for c-authorship left to successive partial authors acting within the confines laid down for them by their antecedents.

N.7.3.1 Thus a musical score shows a high degree of structure (information content), being a selection of one from all possible musical scores. Whereas the degree of structure of the performer's rendering of it shows a lower degree of structure, being a selection from the narrower range of variety that is left to the performer, between the injunction of the score with its limited capacities to designate gradations of amplitude, tempo &c. and the finer discriminative capacities of the individual's auditory apparatus. C-authorship is credited both to composer and performer though generally, if not invariably, a higher degree to the composer. The performer is the composer's instrument as much as the performer's may be the piano or the violin.

N.7.3.2 The architect A*, the builder A and the building W provide another example of shared authorship. In the case of this example, O apportions c-authorship in different shares. A* sets A's goals by realizing a sub-goal in the form of his plans and specifications for the building, but receives most of the credit as author, not simply for the plans, which are quite likely to be forgotten once the building has been completed, but for the building itself; while the builder receives an author's credit, only for such a limited interprative role as architect may have allowed him.

N.7.3.3 Other co-authors are the computer and its programmer - what conditions 'allow us to assert that the programs problem-solve'? (Newell, 1962) - and, at the other end of the scale, the patron and the artist he employs. The patron may commission a work from an artist, specify the subject and even make stipulations as to its treatment but, unless he were to take a substantial role in at least planning the details of its execution, he would receive little credit for playing a creative part in the finished work. In all these cases the information-content test for authorship applies, at least in principle. These cases are all sequential situations, in which purposeful systems operate in linear series. Other possibilities may be more complicated, such as for example the production of a play or an opera, a film or a television programme, though such cases reduce to sequential forms, as we mentioned in 7.5.3, if the various participants in the productions are assigned priorities. Indeed this is almost certainly a necessary practical condition for the achievement of such productions at all.

N.7.4.0 (7.6) What about group production processes in which the participants may not be supposed - as hitherto - to have the same goals, but may share no more than certain sub-goals? Such processes are the milieu of designers rather than artists, though the distinction, as we shall see, is somewhat blurred. We may assess the designer as c-author according to the same principles we used for apportioning c-authorship among groups of partial authors, namely by the variety within which his order-extraction occurs. Suppose a productive process consists of a chain comprising a number of purposeful systems to each of which priorities have been assigned, so that the process is sequential, as above. Suppose A has voluntarily (7.3.3) accepted a relatively low priority in the chain. Then A's freedom to generate goals will be circumscribed by the goals of all the systems in the process that have higher priorities. For instance A might be an industrial designer, required to design an object for mass production, conforming to specifications laid down by various different departments of a manufacturing firm, variables occurring under such heads as accounts, tooling, material costs and availability, sales and so forth. Clearly these variables show a high degree of inter-dependence: A more expensive material might increase sales but put up costs; or a popular material might need special manufacturing equipment or have unreliable suppliers. Now each department would have its own main goal to which A would have to conform. The tooling department might wish to use available machinery; the sales department to compete with some particular item on the market, and so on. Or A might be a 'craftsman' making a table, say. In this case his own sub-goals will take the place of the main goals of the various departments of the industrial designer's firm. The cases are clearly equivalent. In terms of the formulation of 7.1.1, A faces a double task, first to form an explicit goal set, made up of elements that comply with all priority goals and sub-goals, and then to select from this residual set a complete set of compatible, non-redundant specifications of his own goal, - or in the craftsman's case, his own artistic, goal. This is a problem whose solution, to an observer without knowledge of it, displays little if any evidence of how it has been reached. Clearly the variety of A's residual goal set (what the constraints due to prior goals have not eliminated) will be greatly reduced compared to those of the artist on whom constraints of comparable severity seldom result from goals of higher priority than the making of a WOA. Credit for c-authorship of a design normally goes to those for whom the design has been the main goal. But in evaluating the quantity of c-authorship, the character of the designer's task inevitably emphasizes his problem-solving ability in circumstances that frequently conceal from the design's judges, the problem he has solved.

N.7.4.1 The reduced variety of the world within which the designer may be a c-author is seen to detract from his achievement and to assign him and his works a position of importance inferior to that of the artist. Similar considerations, which detract from the assessed artistic merit of Ws, applies in all cases in which the object of making a WOA has not been paramount. Examples are propagandist writing or painting, advertizing, or any form of art in which the communication of a 'message' or some other didactic purpose has taken precedence. We should emphasize that this view derives from the conclusions of our theoretical arguments, though various empirical arguments support it. Two in particular seem important: (1) the notion that art should be didactic ('committed' as it is sometimes called) has been present to a greater or lesser extent at most times, within Western culture at any rate. Ethnological evidence suggests that probably most art originates in solving quasi-didactic problems. But the art whose merit has survived the judgement of successive generations of critics has tended, in modern times, to be non-didactic; and of the art - including ethnological artefacts - of earlier, more explicitly and generally didactic periods, has tended to survive despite its didactic intent rather than because of it. (2) All the arguments we have used so far have applied equally to art and science. Science at least casts no empirical doubts upon the conclusions they have led to. Scientific theory or experiment whose main goal is anything other than that of extracting order from its data, or of adding to the data by discoveries, is judged worse in so far as it is seen to be the result of pre-conceptions. Alchemy, and Lysenkean genetics are among the more obvious examples of its consequences.*

*Monod draws attention to Engels's (1935) applications of Dialectical Materialism to science. In elementary algebra for example this reveals that,

> 'stripped of the veil of mystery in which it was wrapped by the old idealist philosophy and in which it is to the advantage of helpless metaphysicians ... to keep it enveloped', (p.152)

for any algebraic magnitude whatever, call it a,

> 'negated, we get -a (minus a). If we negate that negation, by multiplying -a by -a, we get $+a^2$, i.e. the original positive magnitude, but at a higher degree, raised to its second power. In this case also it makes no difference that we can reach the same a^2 by multiplying the positive a by itself, thus also getting a^2. For the negated negation is so securely entrenched in a^2 that the latter always has two square roots, namely a and -a'.(p.153)

N.7.4.2 In modern times the movement against ulterior purposes in art, flourished partly in reaction to earlier didacticism, but was also associated with notions of individual liberty. Williams (1961) traces the rise of the idea of the individual to the idea of the individual soul, which nourished the Reformation, 'man's direct and individual relation with God', rather than his 'destiny within an ordered structure' (p.91). Artists frequently had had to fill the diminished role of designers and had been wholly dependent on patronage. Duchamp's answer when he was offered a number of commissions after his success in 1913 expresses a widely held attitude: 'No thanks. I prefer my liberty.' (Cited by Lake & Maillard, 1956, p.84.) Though by this time the attitude was probably hardly more than what Tom Wolfe (1976) refers to as 'the anti-bourgeois sing-along of bohemia, standard since the 1940's, as natural as breathing by now [the 1920s], and quite marvellously devoid of any rational content'(p.74).

N.7.5 In addition to constraints on A due to other purposeful systems, among which we should include A's culture and the history, and traditions that affect his chosen field, are constraints due to the 'real' world. On the one hand, a human A owes his own structure very largely to the long, evolutionary order-extraction process that has permitted the development and survival of the human species. As we showed in chapter 2, this process has led to correlations between A and those aspects of the world with which A's antecedents have survived contact. A's perceptions further these correlations, reflecting, not simply <u>some</u> 'real' world, but the one with which A, both specifically and individually has a relationship. For creatures to survive, 'the self image must be in close correspondence with the reality of themselves and their environment.' (Fogel & others, p.122) If A is a machine then, as chapters 4 & 5 showed, much of its structure will derive from its human maker. To this extent, any A is also a W. It is the product of a process in which both purposeful and non-purposeful processes have played a part and, in this, is constrained in the goals it can generate and thus the Ws of which it can be the c-author. But for A the controller, the constraints that make up the structure of M and E also restrict the Ws that AW can produce. These limits being 'physical' naturally are most confining for As working in the 'physical' world, like painters and sculptors, or those who rely on a physical medium, such as musicians, whose music depends on the phenomena of sound and the characteristics of musical instruments. But no A is free of physical constraints. Commonly A will turn constraints to his purpose. They do not 'make the work any less individual, and whether it be that of an architect, a cabinet-maker or a composer, [they reflect] no less minutely the most subtle shades of the artist's personality'. (Proust, Vol.4, p.292) These constraints do not detract from A's achievement. Unlike the designer who may be compelled to work in oak or stainless steel when he might have chosen walnut or pewter, the artist is normally freer to make his own choice, and uses the choice to achieve the effect he wants. The painter cannot determine the position of every particle of paint. Instead he relies on choosing oil paint, water-colour or whatever as his 'medium', picks his brushes and the kind of surface on which to work and allows these to 'arrange' the paint particles for him. The musician chooses his instrument and relies upon his choice for achieving what he would be unable to do in any other way. Simulating the attack and decay of sound patterns for different instruments, using a computer, has so far proved an intractable problem. Writers, especially poets, deliberately draw upon 'physical' qualities when they might have been supposed relatively free from the restrictions these impose. Their repertoire consists of rhythms, rhymes, the tonality of words and imagery appealing to the senses - including the kinaesthetic, what William Empsom calls 'muscular imagery'. Indeed the artist uses the constraints of his <u>chosen</u> medium to extend his own variety into the wider variety of the world.*

---

*Aspects of this notion, especially in the form in which we have considered it here are treated by Kepes (1965), Rudofsky (1965) and Whyte (1961).

CHAPTER 8

N.8.0 Chapter 8 ends our examination of authorship. It argues that A must include among its various features one that resembles O; such a member is a necessary part of its control and regulatory apparatus, without which it would be unable to realize any goal in E. This conclusion provides a basis for an answer to an important question that we have so far neglected; What is a WOA? Beyond asserting that a WOA must be the product of authorship, we have nowhere so far been explicit about its distinguishing characteristics. Paragraphs 0.4.2 dealt with some of the difficulties of arriving at a satisfactory definition of a WOA. Either the definitions turned out to be too wide, admitting all comers without providing sound reasons why, as in the case of Wittgenstein's 'family resemblance group' (0.4.2.2); or they tended to depend upon other notions, as difficult to define as the idea of a WOA, for which they had been invoked. The breadth and evident arbitrariness of what legitimately claim to be WOAs has increasingly, in this century, particularly since the Dada movement, made the task of definition impossible. But the problem partly vanishes if we adopt a paradigm for an author that acknowledges the controller-observer's role, which, as we have seen, other reasons demand in any event. For if A asserts that W is a WOA, then, if no hoax is intended, it is - for A. We mention in passing that knowledge of brain structure makes our present paradigm by no means far-fetched. George (1965) for example, observes that 'the total motivational system operates on a conceptual basis' (p.120) and that 'there is ... good reason to accept curiosity as a drive' (p.121). This notion of conceptualization, which George bases on brain models, implies at least the possibility of a natural model of the kind we have proposed. The problem of possible hoaxes, which might have been serious, turns out not to be, because any claims A makes for W will have to satisfy the individual judgement of each particular O whose assessment may be sought. This is the case whether A claims that W is a WOA or that it is not. Thus ethnological objects are not intended as WOAs by their makers or even their users, yet may be judged to be WOAs by certain Os. Indeed 'found objects' may become surrogate WOAs simply on the judgement of their finders. Mayer (1969) defines a 'found object' as 'An object which is found selected and exhibited by an artist, usually without being altered in any way ... The artist's role in its presentation is creative only in that it points out aesthetic values that the object already possesses but that were not deliberately considered in its construction' (p.153). How this comes about will appear in the following paragraph. But before advancing any explanation we should claim that the overall assertion of our argument is not implausible. Agreement between art critics and art's wider audiences is not so great that we should feel a need to seek an explanation of it in some unifying principle of WOAs. On the contrary, taste is so disparate among the different groups, distinguished by history, culture, class, education and so forth that it appears to be the variations more than the agreements that call for understanding. In any event we have seen in a number of contexts, beginning in 2.2, that any order that appears does sc relative to some particular observer. The present contention is merely a corollary to this assertion.

N.8.1 Of particular interest in AW, as chapter 8 newly depicts it, is the relationship between W and W*. How do the two entities affect one another and where does A stand between them? Paragraph N.6.1 intimated that O's position 'between' W and its producer enabled O to use each as a source of information about the other, and we likened this procedure to the use of PLANNER in Winograd's language programme. The present paradigm clarifies these earlier ideas. Some O, as we now see, must be part of A and, though W may have other Os, considering them rather than the O that belongs to A, introduces separate and, in this context, extraneous problems. Thus, in the greater detail that chapter 8 furnishes, O appears symmetrically between W* and W. Looking forward to the discussion of objective knowledge that follows, at the beginning of Part II, we anticipate the main idea that the argument there advances, that all Ws which AW may produce represent objectifications of W*s, and that therefore WOAs, as members of the class of Ws, represent objective knowledge. We shall look to our present argument to support this view. For our present conclusion imparts to W, from consideration of AW alone, irrespective of the kind of W it produces, the status of objective knowledge. Since, as we have observed, the value of the objective knowledge that any W provides is to be assessed relative to the O that uses it, this value will depend upon similarities and differences between the user-O and the particular O that was part of the process that produced it.

N.8.2 (8.3) According to 8.3 the creative process is by no means a unitary event, but proceeds rather by successive approximations, during which not just W but W* too, the model of W that A's goal-setting procedure uses, alter, if AW succeeds, in gradual convergence, until the correspondence between them is sufficient to permit the process to halt. This occurs in the manner N.6.1 describes, with the O in A - that is A in its capacity as O, write A(O) - alternating between two production processes, AW*, the construction of W*, and AW. W provides a test facility for <u>ideas</u> that develop in W*, according to the results of which W* may be modified and the ideas further developed. Such a procedure is possible provided the time scale T of events in W* is more rapid than the time scale t of events in the real world, which is the case for the human A. We recognize in this process a resemblance to a 'scientific method', though lacking formal proof procedures that enable the connecting together or ordering of ideas, in the manner described in chapter 2. The symbiotic relationship between

W and W* depends upon the differences in variety and structure of E*, A's representation of the environment, and E, the environment itself. In general correspondences between E and E* will be both loose and changeable. E*'s variety is lower than E's with the result that A will use E as a stimulus for its own creativity. A's initial goal is likely to be quite inexplicit. T S Eliot has observed that a poem may begin from no more than the idea, in the sense of an experience, of a rhythm, say. Perhaps it may be less explicit still, no more than a sense of restlessness for instance, sufficient to set AW in motion. From here A may seek to narrow his goal's specifications through recourse to E. This may even take the form of an appeal to the characteristics of the medium itself. Thus artists will allow colours to 'run' and generally to use physical constraints in the manner we have noted (N.7.5). Such procedures have become highly explicit this century. Arnheim (1971) cites works of Abstract Expressionism, particularly Jackson Pollock's paintings of the late 1940s which 'show a random distribution of sprinkled and splashed pigment controlled by the artist's sense of visual order' (p.23). The Dadaists cut up printed poems with scissors and drew the cut up words out of a hat to make new poems. William Burroughs (1970) reports something similar. Francis Bacon reports (Sunday Times Colour Supplement, 24 March 1975) making use of the artistically suggestive qualities of partially destroyed photographs. Less randomly, though still providing explicit examples of the nourishment afforded W* by W, Steiner (1972) shows how writers who write in more than one language are able to cross-fertilize their linguistic styles. He cites the example of Samuel Beckett, who has translated his own writing in French into English, showing how he will often develop the words of an expression when he translates into English, from expressions suggested by the French written previously

> 'No outside translator would have found the equivalents chosen by Beckett for the famous crescendo of mutual flyting in Act II of Waiting for Godot ... The English version springs from the French not by translation but by intimate recreation; [italics added] Beckett seems capable of reliving in either French or English the poetic associative processes that produced his initial text' (p.18).

Most finished WOAs conceal the developmental process that underlies them, as much or more than a scientific paper or the proof of a theorem conceals the intuitions and heuristic stratagems that led to it - more, because it is often the object of the artist to conceal his way. (See for example Medawar, 1969) Moles observes that 'The artistic deed is autonomous, independent of its technique of construction. It may be accessible through its structures, but nothing a priori indicates that these are connected with the technique of construction' (p.114). Pentimento or original manuscripts of music or writing reveal what the finished WOA does not, the deletions, additions and changes of mind of partial goals partially realized.

N.8.3 Notice that A's goal, whether fully specified or not, is reached only if and when W satisfies A as sufficiently realizing it. In general A is likely to be unable to give objective specification of its goal, except by means of W. (Chapter 10 takes up this subject in detail.) So what A's goal is may never be more explicit even to A than W makes it. In practice A may employ various criteria for determining whether or not W realizes a goal. A may seek justification for the development of W by making explicit reference to the received criteria of his art, rather as a mathematician may seek proofs to justify his intuitions. But what indicates to A whether or not a goal has been attained, may equally remain quite vague. Nadezhda Mandelstam (1971) for example reports that her husband, the poet Ossip Mandelstam would experience a sense of buzzing in his head during the composition of a poem, which would persist, evidently outside his control, until the poem assumed the form that he then took to be 'right'. The goal is known only when it is reached. It is worth stressing this because the reverse is often supposed: unless failure occurs, A's goal - final goal - is what W, more or less, realizes; and although alternative goal descriptions may be offered - such as a verbal description of a painting given by an artist - W still furnishes the only means for fully revealing the description's meaning. Thus a young child, asked to draw a man, may be satisfied with lines resembling a spider. One would look for reasons for this, connected with the child's repertoire of descriptors, with his 'understanding' of the instruction, and with the state of his perceptual and motor development. Klee's figures that sometimes resemble children's naturally call for quite different explanations, reflecting as they do the realization of different goals. Before modern times, art, especially in the West, has often been deceptively clear in its objectives, which is why modern art, which often deliberately sets out to destroy the subtleties that veil the goals WOAs express, provides many illustrations in which it is easier to see the operation of the processes that are our concern than earlier works. What goals for instance are realized in Morgenstern's poem Nightsong of the Fish, a wordless composition consisting merely of prosodic markings? We have mentioned these various questions mainly because the theory so far developed appears to offer a richer source of answers than approaches that ignore artistic purpose, such as for instance the information theoretic approach of Moles, or the Gestalt psychological one of Arnheim (1967)

> 'We must have imagination awakened by the uncertainty of being able to attain our object, to create a goal which hides our other goal from us...
>
> 'There must be between us and the fish which, if we saw it for the first time cooked and served on the table, would not appear worth the endless trouble, craft and strategem that are necessary if we are to catch it, interposed, during our afternoons with the rod, the ripple to whose surface come wavering, without our quite knowing what we are intended to do with them, the burnished gleam of flesh, the indefiniteness of a form, in the fluidity of a transparent and glowing azure.' (Proust, Vol.4, p.133)

N.8.4 The argument of N.8.0, that every WOA is particular to the O that claims it is one - an assertion whose truth only the general agreement that sometimes occurs among groups of Os conceals - implies that to apprehend a WOA, O must undertake a process similar to the creative process on which A produced the WOA in the first place. Analagously with the way we have portrayed A's creative process, we may picture O's role as audience. We envisage a state of disequilibrium induced in O by W - or more properly, by O's addressing himself to the task of apprehending W. This, O will gradually reduce by means of a series of references back and forth between W-in-E and O's representation of E, E*(O) say. The disequilibrium vanishes when O has extracted order from W, which he can represent as a W*(O) that matches W. Two immediate consequences of this portrayal of A's audience are, first that, if W is to cause O to enter a state similar to A's at the completion of W, then O and A must share similar kinds of structure, and more particularly, the variety of O's representation of E must be at least as great as A's. Osborne mentions R K Elliott (1972) among a number of theorists who have insisted that aesthetic apprehension of a great work must have 'full imaginative commitment to its evocative force'. Secondly, O may need, in addition to the indicators W provides, evidence, such as W does not display, of A's intention in producing W.

N.8.5 Various indications support such a view of O's rôle. In N.8.0 we mentioned the 'artistic' rôle of the finder of 'found objects', or of the O who viewed some ethnological artefact as a WOA, even though its maker had never intended it as one. We may take this argument further. For O must settle questions of A's intention generally. O is not passively manipulated by W, like a computer by a simple program, he assumes meta-rôles, and reflects on his own reactions and the relation between A's goals and product, using evidence that is likely to come from a wide variety of sources besides any particular WOA that may be the subject of his attention at the time. O does not differ from A in any of this. Questions of A's intention in a WOA touch on problems of 'unconscious' intentions; multiplicity of interpretations - such as of Shakespeare's plays, especially in production, for example; how A's intention may be known if he has failed to express it in a WOA; how knowledge of A's intention from an external source might influence interpretation or judgement of a WOA by O; what the relation is between O and a WOA. Answers will come from the criteria and conventions governing the use in the judgement of WOAs of what we have elsewhere referred to as knowledge of the problem; in our terms, through a study of the structures by which A may have been previously programmed (3.2.0), with the object by means of this study, of gaining an understanding of the kinds of goal structures that A might have been trying - perhaps unsuccessfully - to realizo by means of some given W. The problem of which knowledge is sought here includes the full contextual specifications of A's task. There is the context of 'art' itself: Nightsong of a Fish is a poem only because it is proclaimed to be one by inclusion in a poetry book; just as a heap of leaves is a 'sculpture' only when an art gallery, say, exhibits it as one. There is the historical context, the particular tradition and development of which W forms a part, with which it may consciously associate itself by alluding to some specific aspect of the tradition - allusions whose force will be lost on an audience ignorant of their objects. There is A's biography and his development as an artist, his sketches, cartoons and first draughts, his critical evaluations. And there are wider contexts known by names such as 'Zeitgeist', 'Weltangschaung' and 'ethos', which may involve many aspects of the world in which W is produced. All this and much more besides makes up O's evidence of A's intention. This is why, beyond A's intention, an artistic intention may be internally apparent in the work itself, although this intention may not correspond with any conscious or unconscious intention in the biography of the artist'. Nevertheless, 'it is possible to judge how far a work of art succeeds in realizing the intention implicit in it'. (Osborne, p.22).

N.8.6 So A may betray an emotion as well as express one, or generally endow a WOA with content, without being aware of it simply, because he has been 'programmed' with it. And O, by discovering A's programming may uncover what the content is. A acknowledges O's creative role. We have already given the example (N.8.3) of <u>Nightsong of the Fish</u>. We turn once more to other modern examples because of their caricaturing way of deliberately revealing the tricks they use. 'Minimal' art is an example of A explicitly calling upon O's creativity. Similarly, the simple act on A's part of placing W in a context associated with art - publishing <u>Nightsong of the Fish</u> in a book of poems, or hanging a plain black canvass or placing a heap of leaves in a picture gallery - may often be sufficient to embark O on his creative role. Osborne cites Cioffi: 'A conviction that a poet stands in a certain relation to his words conditions our response to them'. With this relation established, O may at once undertake his creative task. Notice how the theory of the mutual convergence of W* and W is what enables us to assign O his creative role. The creative process of A and that of the independent O differ only in that O's carries the restriction which requires W to remain physically fixed; so that, the direction of O's creative process is the reverse of A's, with O seeking a W* to match W rather than the other way about. Thus Duchamp may present his 'ready-mades',* or the Dadaists their picture painted by a brush, tied to a

*'Ready-made': 'A man-made object, usually mass-produced, that was not made with any artistic consideration in mind but is mounted or displayed as an aesthetically significant structure; a form of found object.' (Mayer, p.322)

donkey's tail, and Os will emerge to create WOAs out of them. And when they have done so, the full received critical apparatus will be at their disposal to reinforce their creative achievement with critical acclaim. The critic's role is the last line of argument we cite. The critic, as a particular O, apprehends a W in a particular way. He then proceeds to try to induce the same apprehension in others; and, by drawing attention to particular properties and presenting particular kinds of descriptions of the W, to induce others to adopt them. His selection of properties and presentation of his description mimics A's own operations.

N.8.7 Presumably it is the separate creative roles of A and each individual O that gives art a bad name with science (chapter O). Though nothing we have assumed in arriving at our view makes it apply only to WOAs and not to other Ws. If there is a distinction to be mentioned here - Part II will deal with W in greater detail - it must be connected with the relative degrees of agreement between the judgements of different Os that Ws in science and Ws in art command. A scientific theory for instance such as the Theory of Evolution brings together into a connected whole a large body of knowledge, about which there may already have been a wide measure of agreement; a new theorem attaches to the axioms and inference rules within which it is proved and, unless some O disputes some notion with a very wide acceptance - such as the nature of proof for instance - there will be little ground for major differences between different O's assessments of it. Clearly the case of WOAs is different, though the differences are exaggerated for several reasons. First, it is true that a random collection of smudges on a canvass may be pronounced a WOA by some and dismissed as a random collection of smudges by others. But, in practice, discrepancies will generally be less bizarre. For the rule of connectibility, which calls for each W to fit into some existing structure, applies to art too; and wide discrepancies of interpretation are unlikely to persist for very long, as each particular WOA or group of WOAs, is required, over a period of time, to connect to a wider and wider base. Secondly, the extent of agreement between Os in science may easily be over-estimated, if only because of the unifying power of sciences 'successful' concepts, which allows 'failures' to be neglected and quickly to vanish almost without trace. Thirdly the explicit technicality of science successfully excludes from among its judges many of those whom ignorance of its problems properly disqualifies from that role; while art, lacking any comparable technical barrier, attracts judges with too little knowledge of art's problems to permit them meaningful assessments. In short, scientific Os are screened by the filter of science's technicality, while art must accept the disparate opinions of all comers.

N.8.8 We note that, although we have approached the problem quite differently from Collingwood (1963), the WOAs that AW will produce within our formulation satisfy all Collingwood's criteria for works of art, namely, (1) that art is not a means to an end - goals extraneous to a WOA detract from it (N.7.4). (2) in art there need be no distinction between planning and execution, - W* alters along with W; (3) following from (1) and (2), no reversal of order is possible from ends to means or execution to planning; (4) there is no distinction in art, between raw material and finished product (cf. Moles's observation, N.8.2). A uses the constraints of his medium to make W's production progress (N.7.5) and W* moves therefore with the medium; (5) by (4), there is no distinction in art between form and matter as there is in craft , the product of design (N.7.5); (6) Art lacks the hierarchy that crafts show, each dictating ends to the one below it, and providing either means, or parts or raw materials to the one above (N.7.4).

CHATPER 9

N.9.0 Here, explicitly, begins a drawing together of art and science. The attitudes and arguments that chapter 0 cites of mutual antagonism between the arts and the sciences needs to be differently founded. Because we have allowed W such full generality, all that has been said of authorship is equally true whether its products are scientific or artistic. In N.3.0 we noted similarities between scientific and artistic progress. Subsequently (N.8.5) we suggested, using similar arguments, likenesses between scientific and artistic 'justification' (Reichenbach, 4.1.2). Here we explore the uses of W. Distinguishing features of scientific and artistic Ws are seen to be less sharply defined than they are usually represented to be. Objectivity and explicitness and their opposites are not dichotomous features but the poles of continua, or more properly of continuous mixes. Effectiveness depends on knowledge of it. Conjectures and refutations will have unconscious as well as conscious elements. Medawar's objection to literature (0.1.2), that though it is required to be internally consistent it is not expected to be empirically true, might apply equally to an axiomatic system. But a correspondence might become apparent between the latter kind of system and the real world, in which case the system might provide the basis for an empirical theory; and if the correspondence were 'close', the theory might extend the notion of reality beyond the realm of experience, as for example in the case of modern physics. No reason is immediately apparent why an internally consistent piece of literature or other work of art should not be put into a similar kind of correspondence with reality, though art might prove more suitable as a model of a somewhat different 'aspect' of reality from that for which we should use an axiomatic system. The different aspects of reality that the axiomatic system and the work of art might be used to model could for example be those corresponding to the two parts of the bi-partite structure of the heuristic that Minsky mentions (0.2.2.1). In other respects the conclusions of this chapter suggest no reason in principle, on grounds of explicitness, objectivity or effectiveness, for denying artistic Ws equal status with scientific ones.

CHAPTER 10

N.10.0 In his essay Politics and the English Language George Orwell enunciated a number of rules for writers. What he advised included recommendations to prefer the active voice to the passive, the particular to the general and the concrete to the abstract.* Tom Wolfe (1976) writing about modern painting in an essay entitled The Painted Word argues that, as art turned away from realism towards abstraction, the theory underlying it became more and more important until now he can quote Hilton Kramer, chief art critic of The New York Times as saying, 'frankly, these days, without a theory to go with it, I can't *see* a painting.' Juxtaposing these views makes an important point about art, especially as it differs from science. Approaching the question from different sides, both Orwell's and Wolfe's arguments - Orwell's concern here is writing and particularly writing with a purpose, but there is nothing in the spirit of what he says to preclude its extension to the wider field of art - imply a need for self-sufficiency in art, a belief that art's concern is with the particular, the concrete, the actively observable, the thing that exists in its own right, accessible without the aid of special knowledge or analytical apparatus. Without this self-sufficiency, art ceases to be art. No such demands are made

*Orwell argued roughly that unclear writing reflected unclear thinking, and that if a writer could not express what he wanted to say, it was because he was not sure of it himself. This is a view that positivists would accept, and which extends naturally to the notion of effective procedures.

upon science. On the contrary: science's concern is with the abstract, the general, with conclusions based upon the broadest possible theoretical foundation. No scientist would expect to be able to understand the papers in a scientific journal without a grounding in the theory on which they are based. Wolfe's view must not be pressed too far. We have argued throughout for the importance of forms and structures for art. Nevertheless we should interpret the development that Wolfe complains of in painting, as vindicating the ideas we have advanced. For, as we observed (N.5.1, N.7.4.2) the movement towards abstraction in painting began as an attempt to break with what was seen as the tyranny of received forms in art. And now

> 'How far we've come! How religiously we've cut away the fat! In the beginning we got rid of nineteenth-century storybook realism. Then we got rid of representational objects. Then we got rid of the third dimension altogether and got really flat (Abstract Expressionism). Then we got rid of airiness, brushstrokes, most of the paint, and the last viruses of drawing and complicated designs (Hard Edge, Colour Field, Washington School).
>
> '<u>Enough</u>? Hardly, said the Minimalists ...' (Wolfe, p.79)

But in some degree at any rate these forms had all been readily accessible to almost anyone in Western culture and well beyond. Abandoned as tyrannies, the forms now appear to have been art's indispensible framework, which surrendered, had to be replaced, if art were to be possible at all. If structure does not exist, it is necessary to invent it. Though replacement by private theories of art's publicly accessible tradition is a movement towards a subjectivity that denies art the general value that it otherwise enjoys as a W. And though the theories are individual, they tend nevertheless to be unoriginal, - as we should expect from what we have said (especially in chapter 4) about the slowness and difficulty of invention - ecclectic assemblages, heavily dependent on other orthodoxies.

N.10.1 '... what it is that the artist, as such and essentially, produces ... is two things. Primarily it is an "internal" or "mental" thing, ... something of the kind we commonly call an experience.' (Collingwood, p.37) To this 'mental' thing Collingwood gives the title, the 'work of art proper'. A WOA is one owing only to its relation to this 'work of art proper'. Within the theory developed in chapter 8, this view is untenable as it stands. It requires the relaxation of the condition that demands that the 'work of art proper' should exist in the artist's head alone. As chapter 8 showed, the WOA is only apprehended and given the name by an observer, when he has constructed his own 'mental' thing to match what he observes. The problem is related to questions of knowledge by acquaintance and knowledge by description. Experience represents knowledge by acquaintance (10.3). If a WOA represents experience, it does so by description. George (1965) compares the two kinds of knowledge respectively to the machine languages and the automatic programming languages of general purpost digital computers. How art's automatic programming language operates is the subject of the following chapter. Here our concern is with the idea of transmitting experience, which results from the interpretation. of art we have arrived at.

N.10.2 In a certain respect it is obviously true that art is its own justification, sufficient for those who want it and unconcerned about detractors. But is there a defence against the charge (N.10.0) that art is merely doing - in a second-rate kind of way - a job that science will eventually take over? Is there anything that art can do which science cannot? 'Experience' must provide the answers to these questions, in particular its uniquenes. Paragraph N.8.3 looked at the way A's goals develop. We have mentioned (N.8.2) the gradual convergence, during the progress of AW, of W and A's representation of it, W*. And we have argued that the finished W is the most specific description of A's goal that is possible. Suppose though that this were not the case, that it were possible to describe in words say, the goal that some particular A had realized in some given WOA, a painting for instance. In order to replace the description that the painting provided, clearly the words would have to convey the same experience as the painting. It follows that, though the alternative to the painting might be possible, it would be (by 10.1.0) highly unlikely.+ In the final instance, the meaning of a painting is the painting; the meaning of the poem, the poem; the symphony, itself. Other descriptions - usually, though not necessarily verbal - can be no more than approximations. Pictures, novels, poems, music are not illustrations of ideas, they are the programs of experience. And though it may be possible to abstract ideas from them - especially their verbal forms, in which ideas specifically may feature as part of a wider ensemble - or describe in words - or some other form - what they are, it is next to impossible to find substitutes for them. The chief reason for this is the time-orderedness of experience (10.1.0). A poem for instance may include images, symbols, rhythm, sound, associations, &c., all occurring simultaneously or in a fixed pattern, as complexes of compresence, giving rise to sequences of 'total momentary experiences' in some fixed order.

N.10.3 But WOAs are no mere surrogates for the reality that is experience's usual source, they are sources in their own right. They do more than simply 'help us extend our personal experience' (George, 1970, p.130) in the sense of multiplying it, providing surrogates of experiences we might not have had. Though WOA's may resemble recognizable realities they also extend them. This view is widely accepted and has been associated with creativity. Thus Osborne observes that, 'in so far as it is creative',

+We avoid cluttering the argument by not introducing the complications of A-O and between-O variations, and variations in the experiences occasioned for a single O by the same complex at repeated presentations. These questions are not germane here.

poetry has been supposed to embody imagination and meaning, in the sense of being 'something added to all reality outside the poem and ... not to be gauged by its correspondence with that reality' (p.2). Moles notes that 'the peculiarity of a work of art is that its richness transcends the individual's perceptual capacity' (p.166), and that its value is 'as a creator of sensations' (p.2) 'A painting by Kandinsky gives no image of earthly life - it is life itself. If one painter deserves the name "creator" it is he'. (Diego Rivera cited by Guy Brett in The Times, 17 April 1973). 'Only a work of art can say with validity and force, as Anna Karenina does, "This is life"' (Leavis, 1967, p.13.). The notion is always the same, that the WOA has brought something new into being:

> 'He had shown in this water-colour the appearance of the roses which he had seen, and which, but for him, no one would ever have known; so that one might say that they were a new variety, with which this painter, like a skilful gardener, had enriched the family of roses.' (Proust, Vol.8., p.123)

The precision here resolves nicely the conflict between this view of art as creative and the idea of art's purely revelatory role. It is clear that A's revelations result from order extraction that he has carried out, which, as we have shown, is his creative effort.

N.10.4 When the question of aesthetics comes up (11.4) it will appear that aesthetic value is often associated with the 'uniqueness' of aesthetic experience (cf. N.10.2), what our terms associate with the order of size of the complex of compresence that a WOA represents, the closeness of the given W to the extreme of particularity, on the continuum that joins it to generality (10.3.0). Science, as we have seen, aims for the opposite extreme. The continuum therefore possesses an evaluative as well as a definitive function. The region between the extremes contains Ws of varying scientific and artistic merit. At the extreme of generality one looks for the 'best' science and expects zero aesthetic value. Notice that there is no conflict in supposing this while simultaneously admitting the high aesthetic appeal of the best science. To anyone lacking

knowledge of the background of 'facts' and theory that some new scientific formula draws together into a relatively short description - one with a high 'information potential' (N.2.4) - the theory will be devoid of aesthetic interest, missing the self-sufficiency (N.10.0) art demands.
But there is still an anomaly when it comes to scaling the aesthetic dimension. Ws that fall at the zero point of the particular-general axis are, in the terms here, neither art nor science. Here questions of aesthetics do not arise. This, not the generality extreme, appears the appropriate point at which to assign zero aesthetic value. Then the generality extreme would acquire a negative rating of magnitude equal to the positively rated particularity extreme. Thus we should be assigning high negative aesthetic values to non-self-sufficient Ws of high information potential, and high positive values to self-sufficient ones. These ratings might also act as 'accessibility' indicators. In terms of the compresence formulation, the world of 'events' occupies the neutral zone, which accords with what we find in practice. Novels usually receive lower aesthetic ratings than poems, genre painting than landscapes, and so on.

CHAPTER 11

N.11.0 Notions about how art 'works' and the beginnings of the idea of artists producing causes to bring about somewhat different effects are one of the chief pre-occupations of art theory. 'Peindre, non la chose, mais l'effet qu'elle produit', was the way Mallarmé (Hartley, 1970) expressed his discovery of this phenomenon, as he saw it. The difference between having an idea of something and having an experience of it is a matter of what is often called feeling. Normal usage applies the word impartially to sensations and emotions. And artists - including as usual here writers, musicians and so forth - and art theorists have proposed numbers of explanations as to how art should set about capturing 'feeling'. Underlying many of these explanations and more or less explicit in some, is a view shared with the one this chapter advances, of WOAs engendering simulated experiences, but in such a way that the experience engendered enjoys more the status not simply of a surrogate experience but of a comment on it or something to be contemplated from a position of detachment (11.3.1). Nineteenth century theories developed early in this century, 'to formulate the view that the artist embodies or symbolizes in the art work an emotion or feeling in such a way that the observer savours and enjoys it without experiencing it in the full sense' (Osborne, p.18) This is what 11.3 referred to as art's dual rôle, and accounts for the conflicting views of those on the one hand who hold that art is no more than revelatory and those others who hold that it is creative.

Langer (1953) adopts a view similar to Pound's notion of images (N.2.4). For Pound an image is at once something to be grasped in an instant and something which may yet be finely and lengthily elaborated into minute detail. Pound cites the <u>Comedia Divina</u> of Dante as an example of an image. The notion is not contradictory. The work forms a single unitary whole. (N.2.2) For Langer art is not a language, but each WOA reproduces by means of its own structure some pattern of feeling. Bullough (1957) stresses the notion of the WOA's dual status as a surrogate experience, able at once to evoke and comment upon some real experience. He proposes the idea of the artist's 'Psychic Distance' from his subject (cf. 11.3.2), which may be seen as an extension of Wordsworth's view in the <u>Preface to the Lyrical Ballads</u> that poetry 'takes its origin from emotion recollected in tranquility.' Tolstoy (1929) emphasizes art's rôle in extending experience (cf. George, N.10.3).

> 'To evoke in oneself a sensation which one has experienced before, and having evoked it in oneself, to communicate the sensation in such a way that others may experience the same sensation ... so that other men are infected by these sensations and pass through them; in this does the activity of art consist'. (p.123)

Roger Fry calls art 'the chief organ of imaginative life; it is by art that it is stimulated and controlled' (p.29). Freud (1966a) thought of literature as a controlling and controlled mode, reaching to the private depths of the poet's consciousness and connecting with universal human problems.

N.11.1 How does the artist achieve these objects? He

> 'has gradually evolved the law, the formula of his unconscious gift. He knows what situations, should he be a novelist - if a painter what scenes furnish him with the subject matter, which may be anything in the world but, whatever it is, is as essential to his researches as a laboratory might be or a workshop' (Proust, Vol.4, p.210).

The artist learns gradually the performance language that provides the characterizing feature of his WOAs. By means of this performance language, 'Genuine poetry can communicate before it is understood' (Eliot, 1951b, p.238). This is because its basic language is injunctive and has the capacity to induce states in O by simulation that O could not induce in himself, lacking synthetic understanding of them. The communication guides him to the understanding, and so produces new synthetic components, new building blocks out of which to construct new WOAs and simulation procedures, in a continuing chain. Pound reports on incident that provides an example. He gets out of a 'metro' in Paris and sees 'suddenly a beautiful face, and then another and another, and then a beautiful child's face, and then another beautiful woman.' He tries all day to find words for what this had meant to him but can find none 'worthy, or as lovely as that sudden emotion'. But later he suddenly finds 'the expression', not 'words, but ... an equation ... not in speech, but in little splotches of colour ... a "pattern "' (p.465).*

---

*Notice the hallucinatory quality of the anecdote and compare the report (N.8.3) of Ossip Mandelstam at work composing poetry. In these cases at least the comparison (11.2.2) between the hallucinatory and creative aspects of imagination do not appear too far-fetched.

Art is not self-defeating, constantly allowing new synthetic knowledge to erode its own preserve. For, every new synthetic description it furnishes is immediately available as a new module for further building. Thus, Pound believed that his discovery provided the basis for a new 'school ... of non-representative painting, a painting that would speak only by arrangements in colour.' (p.465) In this sense art progresses no less than science (cf. N.3.0, N.9.0). Its traditions, the evolution of its forms, the development of individual artists, coalesce - though not necessarily to enhance art's uniquely characterizing powers. These are always to do with the experience it can provide. And this is the realm of the artist's performance language.

> 'The only way of expressing emotion in the form of art is by finding an "objective correlative"; in other words, a situation, a chain of events which shall be the formula for that particular emotion; such that when the external facts, which must terminate in sensory experience, are given, the emotion is immediately evoked' (Eliot, 1951b p.145).

N.11.2 This procedure confers considerable freedom on A. It frees his creations, his WOAs, from the confinement of physical reality. A may 'extend' reality either by relaxing the constraints that bind his world picture or by securing new constraints or by both procedures, with his only restriction the possible consideration of an audience, which conceivably might lose itself in the unfamiliar landscape his actions created. Similarly he may loosen the binding strength of form or seek escapes from the limitations his medium imposes, or he may invent new forms and new media. But whether relaxing constraints or inventing new ones, his actions require order extraction - are creative. To relax a constraint he must first discover that it exists. And, in any case it is not always obvious how a particular 'extension' of reality has been achieved, whether by relaxing or adding constraints. The devils and hobgoblins that decorate Gothic cathedrals for example, are they the result of relaxing reality's constraints or of strengthening them? And if both, which of their characters if due to which procedure? The more that enters A's work that is private to him alone, the less an audience will be able to apprehend of it. However neither A nor the audience need fear too greatly on this account. For, structure being hard to come by, if a WOA were truly complex - highly structured - it could not contain more than a relatively small quantity of 'new' order, due to A's creative effort. And if O found it unintelligible, apparently due to A's excessive inventiveness, he could reassure himself that his failure to understand must be due to a degradation of structure, which from his point of view it would be, not to the reverse, and to offer himself the comfort that what sense he made of the work was attributable to his own efforts.

N.11.3 A's objects in producing 'distortions' of reality might be, and perhaps usually are, part of an attempt to achieve the opposite effect, that of rendering reality more sharply. We have seen (10.4.2) how A will arrive at descriptions of experiences that suit his purposes. His habituations will strip the experience of reality of such 'inessential' constituents as fail to further A's aims. In memory, for instance, Proust says, 'Habit weakens every impression'. The 'better part of our memory exists outside ourself ... wherever ... we happen upon what our mind, having no use for it, had rejected'. (Vol.3, p.308) What we store in our memories is what is useful to us. It is recalled by the cues most commonly associated with it. Suppose the memory is of a thing or of a person, and suppose, whichever it is, vanishes or ceases to exist. The common cues will vanish too. But other, more 'loosely' associated cues may still occur, being not part of the vanished object itself, but part of some experience in which the object was present - an instance of the object (10.1.3). Then these cues may re-evoke the memory of the object when it is neither expected nor desired. Without a goal 'memory is blind and even survival hinges on pure chance.' (Fogel & others p.122). The shock of memory evoked without a purpose, or of reality otherwise surprised, are effects A may deliberately seek. For this he may *blind memory* or twist reality to force some aspect of one or other upon O, in this respect the manipulable and passive subject of chance's experiments, that A simulates.

LIST OF REFERENCES FOR THE TEXT

Ackoff, R.L. & Emery, F.E. (1972). On Purposeful Systems. London, Tavistock Publications Ltd.

Ashby, W. Ross (1962). Principle of the Self-Organizing System. In Principle of the Self-Organizing System (eds. H. von Foerster & G.W. Zopf Jnr.), pp. 259-278. London, Pergamon Press.

Ashby, W. Ross (1962a). The Self-Reproducing System. In Aspects of the Theory of Artificial Intelligence (ed. C.A. Muses), pp. 9 - 18. New York, Plenum Press.

Ashby, W. Ross (1964). An Introduction to Cybernetics. London, Methuen & Co. Ltd.

Ashby, W. Ross (1965). Design for a Brain. London, Chapman & Hall

Ashby, W. Ross (1967). The Set Theory of Mechanism and Homeostasis. In Automaton Theory and Learning Systems (ed. D.J. Stewart), pp. 23 - 52. London, Academic Press Ltd.

Ashby, W. Ross (1972). Systems and Their Informational Measures. In Trends in General Systems Theory (ed. J. Klir), pp. 78 - 97. New York, John Wiley & Sons Inc.

Beckett, S. (1970). Proust. London, Calder & Boyars.

Beer, S. (1972) Brain of the Firm. Middlesex, Allen Lane.

Bowden, B.V. (1953). Faster than Thought. London, Sir Isaac Pitman & Sons Ltd.

Bremermann, H.J. (1962). Optimization through Evolution and Recombination. In Self-Organizing Systems (eds. M.C. Yovitz, G.T. Jacobi & G.D. Goldstein), pp. 93 - 106. Washington DC, Spartan Books.

J.S. Bruner, J.J. Goodnow & G.A. Austin (1956). A Study of Thinking. New York, John Wiley & Sons Inc.

Collingwood, R.G. (1963). The Principles of Art. Oxford, Clarendon Press.

Conant, R.C. (1969). The Information Transfer Required in Regulatory Processes. IEEE Transactions on Systems Science. SSC-5, 4.

Craik, K. (1967). The Nature of Explanation. London, Cambridge University Press.

Dewey, J. (1958). Art as Experience. New York, Capricorn Books, G.P. Putnam & Sons.

Descartes, R. (1949). A Discourse on Method. (Translated by J. Veitch). London, J.M. Dent & Sons Ltd.

Fowler, H.W. (1926). A Dictionary of Modern English Usage. Oxford, Clarendon Press.

Fogel, L.J., Owens, A.J., & Walsh, M.J. (1966). Artificial Intelligence Through Simulated Evolution, New York, John Wiley & Sons Inc.

Freud, S. (1933). New Introductory Lectures on Psycho-Analysis. (Authorized translation by W.H.J. Sprott). London, The Hogarth Press.

Freud, S. (1966). Creative Writers and Day-Dreaming. In The Standard Edition of the Complete Psychological Works of Sigmund Freud. (Translated from the German under the general editorship of James Strachey in collaboration with Anna Freud, assisted by Alix Strachey and Alan Tyson), Vol. 9 pp.141-155. London, The Hogarth Press.

Freud, S. (1966a). Civilization and Its Discontents. In The Standard Edition of the Complete Psychological Works of Sigmund Freud. Vol. 21 pp. 59 - 246.

Fry, R. (1937). Vision and Design. Middlesex, Penguin Books Ltd.

Gagné, R.M. (1962). The Acquisition of Knowledge. Psychological Review, 69, 4, pp. 355 - 365.

Gagné, R.M. (1964). Problem Solving. In Categories of Human Learning (ed. A.W. Melton), London, Academic Press Ltd.

George, F.H. (1970). Science and the Crisis in Society. London, Wiley Interscience.

Hewitt, C. (1969). PLANNER: A Language for Proving Theorems in Robots. Proceedings of the International Joint Conference on Artificial Intelligence 9, pp. 295 - 301. Pedford, Mass., Mitre Corp.

Honour, H. (1968). Neo-Classicism. Middlesex, Penguin Books Ltd.

Kant, E. (1934). Critique of Pure Reason. (Translated by J.M.D. Meikeljohn). London, J.M. Dent & Sons Ltd.

Klir, J. & Valach, M. (1965). Cybernetic Modelling. (Translated by P. Dolan). London, Iliffe Books Ltd.

Lettvin, J.Y., Maturana, H.R., McCulloch, W.S. & Pitts, W.H. (1965). What the Frog's Eye Tells the Frog's Brain. In The Embodiments of Mind (ed. W.S. McCulloch), pp. 230 - 256. Cambridge, MIT Press.

Lindsay, A.D. (Translator) (1954). Plato's Republic. London, J.M. Dent & Sons Ltd.

Löfgren, L. (1967). Recognition of Order and Evolutionary Systems. In Computer and Information Sciences Vol. II. (ed. J. Tou), pp. 165-175. New York, Academic Press.

Löfgren, L. (1969). Relative Recursiveness of Randomization and Law Recognition. Notices of the American Mathematical Society, 16, 4, p. 685.

Löfgren, L. (1972). The Relative Explanation of Systems. In Trends in General Systems Theory (ed. J. Klir), pp. 340 - 406. New York, John Wiley & Sons Inc.

McCarthy, J. (1958). Programs with Common Sense. In The Mechanization of Thought Processes, Vol. I, National Physical Laboratory, Symposium No. 10, pp. 75-84. London, Her Majesty's Stationery Office.

MIT Press.

McCulloch, W.S. (1965). The Embodiments of Mind. Cambridge MIT Press.

Medawar, P.B. (1972). The Hope of Progress. London, Methuen & Co Ltd.

Minsky, M. (1958). Some Methods of Artificial Intelligence and Heuristic Programming. In The Mechanization of Thought Processes. Vol. I, National Physical Laboratory Symposium No. 10, pp. 3-27. London, Her Majesty's Stationery Office.

Minsky, M. (1965). Matter, Mind and Models. Proc. IFIP Congress, 1, pp. 45 - 51, Washington DC, Spartan Books.

Minsky, M. (1972). Computation: Finite and Infinite Machines. London, Prentice Hall International, Inc.

Minsky, M. & Papert, S. (1969). Perceptrons. Cambridge, MIT Press.

Minsky, M. & Papert, S. (1972). Artificial Intelligence, Memo No. 252. Cambridge, MIT Press.

Moles, A. (1966). Information Theory and Esthetic Perception. (Translated by J. Cohen), Urbana, University of Illinois Press.

Monod, J. (1974). Chance and Necessity. (Translated by A. Wainhouse). Glasgow, William Collins Sons & Co Ltd., Fontana Books.

Newell, A. (1962). Some Problems of Basic Organization in Problem-Solving Programs. In Self-Organizing Systems (ed. M.C. Yovitz, G.T. Jacobi & G.D. Goldstein), Washington DC, Spartan Books.

Newell, A. & Simon, H.A. (1956). The Logic Theory Machine. IRE Transactions on Information Theory, IT - 2, No. 3.

Newell, A. Shaw J.C. & Simon, H.A. (1958). The Processes of Creative Thinking. Rand Corp. Memo. P. 1320.

Newell, A., Shaw J.C. & Simon, H.A. (1963). G.P.S. A Programme that Simulates Human Thought. In Computers and Thought (eds. E.A. Feigenbaum & J. Feldman). pp. 279-297. McGraw-Hill Books Co. Inc.

Osborne, H. (1956). Theory of Beauty. London, Oxford University Press.

Osborne, H. (ed.) (1972). Aesthetics. London, Oxford University Press.

Pask, A.G.S. (1959). Physical Analogues to the Growth of a Concept. In The Mechanization of Thought Processes, Vol. II, National Physical Laboratory Symposium, No. 10, pp.877-922. London, Her Majesty's Stationery Office.

Pask, A.G.S. (1968). An Approach to Cybernetics. London, Hutchinson & Co (Publishers) Ltd.

Pask, A.G.S. (1969). The Computer-Simulated Development of Populations of Automata. Mathematical Biosciences, 4, pp. 101 - 127.

Pask, A.G.S. (1974). Analogy. In Cybernetics of Cybernetics (Compiled by H. von Foerster). Urbana, University of Illinois, Biological Computer Laboratory Report, No. 73.38.

Pask, A.G.S. (1975). Conversation, Cognition and Learning. Amsterdam, Elsevier Scientific Publishing Co.

Pask, A.G.S. & Scott, B.C.E. (1973). CASTE: A System for Exhibiting Learning Strategies and Regulating Uncertainties. Int. J. Man-Machine Studies, 5, pp. 17 - 52.

Pask, A.G.S., Scott, B.C.E. & Kalliakourdis, D. (1973). A Theory of Conversations and Individuals. Int. J. Man-Machine Studies, 5, pp. 443 - 566.

Penfield, W. (1958) The Role of the Temporal Cortex in Recall of Past Experience and Interpretation of the Present. In The Neurological Basis of Behaviour (eds. G.E.W. Wolstenholme & C.M. O'Connor), A CIBA Foundation Symposium, pp. 149 - 174. London, J. & A. Churchill Ltd.

Polya, G. (1948). How to Solve It. New Jersey, Princeton University Press.

Popper, K.R. (1959). The Logic of Scientific Discovery. London, Hutchinson & Co (Publishers) Ltd.

Popper, K.R. (1963). Conjectures and Refutations. London, Routledge & Kegan Paul Ltd.

Popper, K.R. (1972). Objective Knowledge. London, Oxford University Press.

Reichenbach, H. (1968). The Rise of Scientific Philosophy. Los Angeles, University of California Press.

Rosenblatt, F. (1958). Two Theorems of Statistical Separability in the Perceptron. In The Mechanization of Thought Processes. National Physical Laboratory, Symposium No. 10, pp. 419 - 456. London, Her Majesty's Stationery Office.

Russell, B. (1948). Human Knowledge: Its Scope and Limits. London, George Allen & Unwin Ltd.

Shannon, E. & Weaver, W. (1949). The Mathematical Theory of Communication Urbana, University of Illinois Press.

Schrodinger, E. (1947). What is Life? New York, McMillan.

Snow, C.P. (1971). The Two Cultures and the Scientific Revolution in Public Affairs. London, McMillan Ltd.

Solomonoff, R.J. (1964). A Formal Theory of Inductive Inference. Part 1. Information and Control, 7, pp. 1-22.

Sommerhof, G. (1969). The Abstract Characteristics of Living Systems. In Thinking Systems (ed. F.E. Emery). Middlesex, Penguin Books Ltd.

Spencer-Brown, G. (1969). Laws of Form. London, George Allen & Unwin Ltd.

Sutherland, N.S. & MacKintosh, N.J. (1971). Mechanisms of Animal Discrimination. New York, Academic Press Inc.

Taylor, J.G. (1962). A Behavioural Basis of Perception. New Haven, Yale University Press.

Turing, A.M. (1937). On Computable Numbers with an Application to the Enscheidungsproblem. Proc. London Math. Soc., (2) 42, pp. 230 - 265.

Turing, A.M. (1950). Computing Machinery and Intelligence. Mind, 51, No. 236, pp. 433 - 460.

Uhr, L. (ed.) (1966). Pattern Recognition, New York, John Wiley & Sons Inc.

Von Foerster, H. (1960). On Self-Organizing Systems and Their Environments. In Self-Organizing Systems (eds. M. Yovits & S. Cameron), pp. 31 - 50. London, Pergamon Press.

Von Foerster, H. (1972). The Perception of the Future and the Future of Perception. Instructional Science, 1, No. 1.

Winograd, T. (1972) Understanding Natural Language. Edinburgh, University of Edinburgh Press.

Wittgenstein, L. (1966). Lectures and Conversations (ed. C. Barrett). London, Oxford University Press.

Zadeh, L.A. (1965). Fuzzy Sets. Information and Control, 8, pp. 338 - 353.

LIST OF REFERENCES FOR THE NOTES

Arnheim, R. (1967). Art and Visual Perception, A Psychology of the Creative Eye. London, Faber & Faber.

Arnheim, R. (1971). Entropy and Art. Berkeley, University of California Press.

Blixen, K. (1964). Out of Africa. London, Jonathan Cape Ltd.

Bullough, E. (1957). 'Psychical Distance' as a Factor in Art and an Aesthetic Principle. In Aesthetics: Lectures and Essays by Edward Bullou (ed. E.M. Wilkinson). Cited by H. Osborne (ed.) (1972), Aesthetics. London, Oxford University Press.

Burroughs, W. (1970). The Job, Interview with William Burroughs by Daniel Odier. London, Jonathan Cape.

Cioffi, F. (1963). Intention and Interpretation in Criticism. Proc. Arist. Soc. 1963 - 1964, 64, pp. 85 - 107.

Collingwood, R.G. (1963). The Principles of Art. Oxford, Clarendon Press

Di San Lazzaro, G. (1949). Painting in France 1895 - 1949. London, The Harvill Press Ltd.

Eliot, T.S. (1951). The Metaphysical Poets. In Selected Essays (by T.S. Eliot), pp. 281 - 291. London, Faber & Faber Ltd.

Eliot, T.S. (1951a). Dante. In Selected Essays (by T.S. Eliot), pp. 237 - 286. London, Faber & Faber Ltd.

Eliot, T.S. (1951b). Hamlet. In Selected Essays (by T.S. Eliot), pp. 141 - 146. London, Faber & Faber Ltd.

Elliott, R.K. (1972). Aesthetic Theory and the Experience of Art. In Aesthetics (ed. H. Osborne), pp. 145 - 158. London, Oxford University Press.

Engels, F. (1935). Herr Eugen Dühring's Revolution in Science (Anti-Dühring). (Translated by E. Burns). London, Lawrence & Wishart Ltd.

Fry, R. (1937). Vision and Design. Middlesex, Penguin Books Ltd.

George, F.H. (1965). Cybernetics and Biology. Edinburgh, Oliver & Boyd.

Greenberg, C. (1947). The Present Prospects of American Painting and Sculpture. Horizon (ed. C. Connolly), No. 93 - 94, pp. 20 - 30.

Harmon, L.D. & Knowlton, K.C. (1968). Computer Generated Pictures. In Cybernetic Serendipity (the computer and the arts). (ed. Jasia Reichardt), pp. 86 - 87. London, Studio International.

Hartley, A. (ed.)(1970) Mallarmé. Middlesex, Penguin Books Ltd.

Kepes, G. (ed.)(1965). Structure in Art and Science. London, Studio Vista.

Lake, C. & Maillard, R. (1956). A Dictionary of Modern Painting. London, Methuen & Co. Ltd.

Langer, S. (1953). Feeling and Form: a Theory of Art. New York, Scribner.

Leavis, F.R. (1967). Thought and Significance in a Great Creative Work. In Anna Karenina and Other Essays (by F.R. Leavis). pp. 9 - 32. London, Chatto & Windus Ltd.

Mandelstam, N. (1971). Hope Against Hope. A Memoir. London, Collins & Harvill Press.

Mayer, R. (1969). A Dictionary of Art Terms and Techniques. London, Adam & Charles Black.

Medawar, P.B. (1969). Hypothesis and Imagination. In The Art of the Soluble: Creativity and Originality in Science (by P.B. Medawar). Middlesex, Penguin Books Ltd.

Palimurus (C. Connolly) (1961). The Unquiet Grave. London, Arrow Books Ltd.

Pound, E. (1914). Vorticism. The Fortnightly Review, 46, No. 573. New Series.

Proust, M. (1957). Remembrance of Things Past. (translated by C.K. Scott-Moncrieff). London, Chatto & Windus Ltd.

Rudofsky, B. (ed.) (1965). Architecture Without Architects. New York, The Museum of Modern Art.

Steiner, G. (1972). Of Nuance and Scruple. In Extraterritorial. Papers on Literature and the Language Revolution (by G. Steiner). pp. 12 - 21. London, Faber & Faber Ltd.

Tolstoy, L. (1929). What is Art? (translated by A. Maude). London, Oxford University Press.

Wain, J. (1975). The Breaking of Forms. Encounter, 45, No. 2, pp. 49 - 56.

Whyte, L.L. (ed.) (1951). Aspects of Form. London, Lund Humphries & Co. Ltd.

Wolfe, T. The Painted Word. Harper and Queen, Feb. 1976, pp 70 - 81. National Magazine Co. Ltd