# Wheels within wheels: an examination of the nature of psychological explanation via a theoretically oriented history of some mechanical models.

David M. Fryer.

Ph.D. Degree

University of Edinburgh

1978

CONTENTS.

"Forty years since, I found three hundred people assembled, to see, at a shilling each, a coach which went without horses; ... but a small paper of snuff, put into the wheel, soon convinced every person present, that it could not only move, but sneeze too, perfectly like a Christian. That machine was not a wheel within a wheel, but a Man within a wheel;"

- Philip Thicknesse (1784).

"Professor Bouillaud...confronted by Edison's phonograph, assaulted the technician; 'Wretch! We are not going to let ourselves be duped by a ventriloquist!'"

- Brian Inglis (1976).

ACKNOWLEDGEMENTS

I am grateful to Dr. J. C. Marshall and Dr. J. Beloff for their invaluable suggestions and encouragement during the preparation of this thesis.

# DECLARATION

Chapter 10 was jointly written by J. C. Marshall and D. M. Fryer and is to appear as: Speak Memory! An Introduction to some historic studies of remembering and forgetting (Marshall and Fryer, 1978).

Most of Chapter 2 was jointly written by D. M. Fryer and J. C. Marshall and is to appear as: The Motives of Jacques de Vaucanson (Fryer and Marshall, 1978).

Approximately half of Chapter 3 was jointly written by D. M. Fryer and J. C. Marshall and read as: A Timely Note on Nervous Ticks (Fryer and Marshall, 1977) at the XV International Congress of the History of Science, Edinburgh, 1977.

I herein declare that in each of these pieces of work my contribution has been substantial and that the remainder of the thesis has been composed entirely by myself.

## ABSTRACT

The aim of this thesis is to ask, and attempt to answer, some pertinent questions about that type of psychological explanation which proceeds by simulation, or model building. The method chosen is a detailed examination of some models, mostly 18th and 19th century mechanical ones, together with a theoretically motivated discussion of the relations between these models and the development of psychological theories contemporary with them.

Two types of model, formal and intimate, are distinguished, both by their aetiology and by the way they are used by working scientists, and several examples of each type are subjected to scrutiny, as are the intentions of their modellers in building or adopting them.

Four main foci of interest emerge: <u>the history of experimental psychology</u> (the myth that experimental psychology was born circa 1870 is exploded); <u>the sociology of science</u> (the impact of developing technology on psychological theory, via the proffering of models, is clearly demonstrated); <u>the philosophy of psychology</u> (issues such as the nature of explanation and the problem of representation are discussed); and, last but not least, <u>theoretical psychology</u> (the value of work in cognitive simulation, and of some work in Artificial Intelligence, is stressed and, partly, explained).

# INTRODUCTION

"the scientific materialists trusted that the world could be dissected into its parts, and that an understanding of the world was equal to a knowledge of how these parts fit together. Indeed this is the appeal of the mechanical model..."

Frederick Gregory (1977).

What is it to give a psychological explanation? This question is far from easy to answer, with the result that there is, in psychology, as Deutsch (1960) has pointed out, not only a lack of explanations, but even a lack of consensus as to what it is to explain! To ask a lesser but still formidable question, what are, if not the best, at least viable methods of extending our understanding of the cognitive capacities of intelligent beings?

One bold reply to the former challenge has been offered intermittently but persistently throughout the centuries: we atart to explain (and therefore understand) a cognitive capacity when we can say or show how it is systematically related to that entity which has the cognitive capacity and we complete the process when we can say or show how the cognitive capacity is systematically related to <u>any</u> entity which <u>can have</u> it, i.e. when we can show how the cognitive capacity is a product of a structure instantiated in organised matter.

A major inroad into the explanation of a human cognitive capacity is then made when we can say how some <u>other</u> entity e.g. a machine, which has the same capacity, works. Of course, the notion of 'same' needs some unpacking and explication, which will be attempted later, but, in essence, the answer offered is simple: we begin to understand an entity e with capacity c insofar as we understand a model of e which does the same as e in the same way.

If the desired end of research into a cognitive capacity is a systematic account of the relations between that cognitive capacity and any possible material instantiation of it, the implications for the means of research are clear and profound, and these provide a reply to the second challenge: we must avail ourselves of, and use, models.

A close look at the ways models have figured in psychological research reveals, however, two relatively distinct modes of scientific practise. In the one case, pre-existing entities or phenomena are seen to be similar in conceptually interesting ways to that which is to be explained and the status of model is bestowed upon this pre-existing entity by the modeller. In the other case, the model is actually constructed by the modeller to be similar in conceptually interesting ways to that which is to be explained.

Both of these ways of modelling have their own individual characteristics but both have made great contributions to the development of scientific psychological explanation. Both, of course, bring with them their own retinue of problems and apparent paradoxes.

In the pages that follow, we will fully discuss these two sorts of model, which I have named intimate and formal respectively, with copious illustrative examples. We will debate the ways they have facilitated and hindered psychological research, the problems they raise and those they resolve.

The models we shall discuss are, however, not ones typically met in connection with psychological model building - flow charts, logogens and, if we are lucky, perceptrons, but bold, imaginative, mechanical fabricata, sometimes even life-size automata, real working machines in their clanking grandeur, and objects with which men of genius have made illuminating comparisons to mental functioning.

What is perhaps most surprising of all, to readers coming new to this field, is the fact that many of these psychological models were created in the 18th century. There seem to be two main reasons for the surprise. Firstly, due to what could be called the Boring myth, of which more later, many psychologists find the idea of a thriving experimental psychology of the 18th century impossible to believe. Secondly, over-whelmed by the received opinion that the 20th century is the age of the technological miracle, the ingenuity and sheer genius displayed by the engineers and simulators of the 1700s is almost literally incredible.

The motivation for this piece of research is not, however, merely the intrinsically fascinating nature of the subject matter - although, as we will see, the automata and simulacra of the 17th and 18th centuries (and even earlier) do, by their complexity, strain the limits of credibility of the 20th century mind. It is, though, a significant aim of this thesis to set straight somewhat the historical record. This is necessary in two ways.

Firstly, few, if any, of the men and simulacra discussed below, in the chapters on formal models, will be familiar figures. They seldom even figure in histories of technology - let alone histories of science in general, or psychology in particular - and when they do, it is nearly always condescendingly as 'mere' gadgeteers or toy-makers. As will be shown below, this is a gross slander. They were interested in toys - but then these men had such exuberant imaginations and apparently boundless energy that they were interested in virtually everything. Toys as such were not however their principle concern but the principles they embodied were. As we shall see, Vaucanson's musical automata were not toys but material instantiations of theories of flute-playing-man. To generalise the point, these men were trying to understand that part of the natural world which is man by simulating him.

Secondly, and more particularly, the history of psychology, post Boring, has been plagued by the erroneous conviction that psychology as an experimental science began in 1860 when Fechner got out of his armchair and went into his workshop. It is true that, in his revised second edition, Boring did pay lip service to earlier movements in philosophy and physiology as precursors of psychology but 'scientific' psychology itself was still, according to Boring, 'founded' in 1860. In this thesis will be found persuasive evidence for the existence of an experimental science of psychology long before Boring's 'foundation stone' was laid.

Lakatos has correctly remarked that philosophy of science without history is empty and it is equally true that history of science without philosophy is blind. Accordingly, I have tried, in writing this thesis, to contribute not merely to the history of psychology but also to the philosophy of psychology ... and in particular to the philosophy of current psychology.

The term 'philosophy of psychology' is, however, open to some misinterpretation. Most of the work produced in the name of philosophy of psychology recently has not gained and, I believe, has not earned, the respect of psychologists ... partly because it is a philosophy of psychology devoid of psychology!

There are three main representatives of this type of work. Firstly, what we might call the ordinary language criticisms of psychology, as given by authors such as Hamlyn, Peters and Bennett, is based primarily on the claim that the realms of discourse of 'ordinary people' and psychologists respectively are distinct, incommensurate and non-intertranslatable and thus statements of the latter can never act as an explanation of the former. By illicitly identifying the ordinary realm of discourse with reality, these writers attempt to undermine the psychologist's case. In response to this, however, the psychologist may maintain that the 'ordinary man's' language of thoughts, minds, hopes etc. is itself a theory ... and not a statement of how things are. The psychologist's and the ordinary man's theories are thus in competition as rival accounts, and the psychologist may claim, at least without absurdity, that his theory is for various reasons the better alternative. The second common form of philosophy of psychology is a sort of poor man's philosophy of science ... usually of physics... extended arbitrarily and inappropriately to psychology e.g. the Kuhnian account of psychology attempted by Palermo. Thirdly is a rather futile sort of historical exegesis apparently merely intended to show that we all had it wrong about old x, or that y had said it all before.

None of these three approaches, it seems to me, is very helpful to the working psychological scientist. The first prescriptively tells him what he is doing is impossible (and he goes ahead and does it anyway), the other two are almost totally retrospective. What is needed is a new sort of philosophy of psychology, let us call it theoretical psychology, which recognises that psychological theories must be both empirically constrained and conceptually viable, which pulls out and makes explicit the theoretical implications of empirical research and which points out the empirical consequences of theoretical positions.

There is no lack of <u>data</u> in current experimental psychology, but there is a lack of evidence. This is because evidence is evidence <u>for</u> theories, and theories are sadly missing in psychology. Data, on the other hand, are available in profusion, frequently the result of a convenient experimental procedure (maze running, nonsense syllable learning etc.), but it is not clear to what these data are relevant.

One area in which there <u>are</u> general theories, as opposed to experimental paradigms, has mushroomed in recent years, that is the field of Artificial Intelligence (A.I.). Practitioners of A.I. disagree as to whether they <u>are</u> trying to simulate human intelligence or whether they are trying merely to get machines to do, by any means whatsoever, tasks which, if a human were to do them, would require intelligence.

One well represented view, however, can be described as seeing the main aims of A.I. to be studying the structure of problem solving processes independently of its realisation (McCarthy), a view in accord with the answer given earlier to the question of what constitutes a psychological explanation. Psychologists, frequently embroiled in methodological disputes, are now more and more beginning to discuss the relevance of A.I. research to their own work. Unfortunately, apart from a few oases of interdisciplinary cognitive studies, psychologists are likely to have had dealings with computers only in the context of data-processing, an experience destined to bias them against the A.I. enterprise. Depressingly, in discussions with psychologists, one meets the standard objections to machine intelligence ... that computers are 'just' super calculators, that they cannot do anything original, can not learn,etc. ... as one hears from naive (non-cognitively concerned) laymen, as if psychologists believed themselves not to be studying a special sort of (neural) machine. Admittedly, some (dualist) psychologists <u>do</u> maintain that psychology is not confined to the study of the complex functioning of a part of the natural material world, but it is not to these, but rather to monistic psychologists with impoverished concepts of machines, to which I address this complaint. For us, the interesting question is not whether the brain is or is not a computer, but what sort of computer it is!

However, there <u>are</u> other interesting questions to be asked about the relations between A.I. and psychology, at least in so far as the former tries to model the subject matter of the latter. What, for example, <u>is</u> a model ? What relation does a model bear to what it models? What is the relation between models and further research? Between models and theories?

In the pages that follow, I shall try to give some answers to these questions. The problems I tackle are intended to be relevant to the current debate about the relation between simulation and psychological explanation. However, as mentioned above, instead of dealing with modern A.I. projects, which are usually so technically complex that they require so much lengthy exposition that there would be no time or space for discussion of the issues relevant to psychology, I deal mostly with pre-20th century models.

I maintain that the same important issues can be discussed in connection with these early simulacra as with modern ones, but with considerable extra interest owing to their rather exotic nature.

The substantiation of this claim, is however, the thesis itself ... to which we now proceed.

# CHAPTER ONE

## WHAT MODELS ARE.

"models...not products of the imagination,
but representatives of real things."
Max Born (1953/4)

What is a model? To limit the domain of discourse from the outset, we shall not here be concerned with models in the sense of ...exemplar ("Nixon was not a model President."); scaled imitation ("Gerald Ford was a model President."); type ("She owns a well preserved model T Ford."); mannequin ("Twiggy was a successful model in the 1960s."); poser ("The art students paint from live models."); etc., but rather with what it is to be a model in the sense in which a computer is a good (or bad) model of the brain (mind).

As a first approximation: a model is an entity or phenomenon which helps us understand some system's functioning. (Boden (1977): "I use the expression (Artificial Intelligence) as a generic term to cover all machine research that is somehow relevant to human knowledge and psychology").

However, this use of the term is surely over-extended...if this were the criterion, slide-rules and microscopes would count as models and tachistoscope manufacturers would automatically count as psychologists - or at least as 'artificial intelligentsia' (this ugly term is due to Dr. Louis Fein).

(Only slightly) less roughly: a model is an entity or phenomenon which stands in place of something i.e. represents it.

The problem of representation is raised in many areas of psychology, for example, in perception, psycholinguistics and modelling i.e. it seems to be a general problem. However, one might ask whether there is, in fact, a problem ... or many separate (and different) ones.

Some seem to think that there is an underlying account which applies to all cases of representation. Dennett (1977) has written: "What is needed is nothing less than a completely general theory of representation, with which we can explain how words, thoughts, thinkers, pictures, computers, animals, sentences, mechanisms, states, functions, nerve impulses and formal models can be said to represent one thing or another. It will not do to divide and conquer here - by saying that these various things do not represent in the same sense. Of course that is true, but what is important is that there is something that binds them all together, and we need a theory that can unify the variety." Note that, broad as the extension of the term 'representation' is (as used by Dennett), it does not include representation in the sense of 'deputation', or the way a lawyer represents a client ... though these are no doubt related ... I shall conform with this usage.

We cannot hope to give a general theory of representation here and now, of course. Most of what is said below concerns representation and modelling, although issues in psycholinguistics and perception are touched upon.

Perhaps we can say, then, that: Every model of something is a representation of that thing in the sense that it stands in place of that thing (for certain purposes). (c.f. Sloman (1975): "representation ... a more or less complex structure which has addressable and significant parts and which as a whole is used to <u>denote</u> or <u>refer</u> to something else.")

However, fairly obviously, not every representation of a thing, for example, a portrait of the Queen or the word "Tokyo", is a model of it, the Queen or Tokyo.

Accordingly: models are a species of representation. However, although it seems almost a truism that representations are representations - <u>of</u> something or other ... paradoxically, they may fail to be <u>of</u> anything at all (we have all seen pictures of unicorns, and what are inaccurate models models of?). It is, then, too strong a claim to say with Goodman (1969) that the essence of the representation relationship is denotation.

An approach using Frege's (1970) distinction between Sinn and Bedeutung seems to have more hope of success ... representations not only have a denotation (or reference) but also a sense. The reference of a sign is "that to which the sign refers", the sense is that "wherein the mode of presentation is contained." Two representations may have the same reference but different senses. Frege's classic examples are "The Morning Star" and "The Evening Star", both of which expressions refer to Venus but express different senses (they identify Venus by different procedures). Two portraits of the Queen, say by different artists, or two models of the brain will equally have different senses but the same reference. Some representations, for example, "the least rapidly converging series", a painting of a unicorn or a hydraulic model of motivation (literally interpreted), may have a sense but no reference at all.

Denotation, then, is not necessary for representation and neither, in the case of modelling at least, is it sufficient.

To move from the logic of the modelling relationship to the pragmatics of it for a moment, we may ask what models are (typically) used for.

Models are not merely intended to denote, they also figure in explanatory theories. As noted above, this is a point about the pragmatics of model use - rather than about the logic of modelling - but we may nevertheless use the insight it provides to help us bear in mind the question of what it is about the model-modelled relationship, which makes it suitable to be used in explanations.

Tentatively then: a model is a species of representation with a particular use ... it figures in an explanatory account.

This still requires further qualification however. As it stands, this characterises words (at least referring expressions) in theories as models. It seems that, for the case of models at least, we require a stronger notion than mere representation ... the model not only stands in place of the modelled but resembles it and functions explanatorily by resembling it.

How then does a model resemble what it models? In other words, what properties do a model and what it models have in common? A reasonable first approximation to an answer is ... a function (they 'do' the same thing!) In this sense, a mousetrap is a weak model of a cat. A stronger requirement is that, not only should the model and the modelled fulfil the same function, but insofar as function is a product of structure, they should instantiate the same structure.

The relation between structure and function is a complex one. In 19th century biology (St. Hilaire, Cuvier) a useful distinction was made between organs in different species which were homologous and those which were analogous. Organs in differing species are homologous if they correspond in position and connections relative to the whole organism and are made up of corresponding parts, whatever the difference in function. The human hand and the bat's 'wing' are thus homologous. Organs are analogous if they have the same function, irrespective of structural dissimilarities. The bat's 'wing' and the bird's wing are thus analogous. We might say bat's 'wing' weakly models a bird's wing. To strongly model it, it would also need the same structure and structural relations. Whether our pre-occupation in modelling is with structure or function depends on our concerns. If we are crudely modelling a process, we can analyse it into functional components, replace each component with a functionally similar one (a black box) and we may not mind that the structural relations are ignored or even violated. Stronger modelling (cognitive simulation) requires that the components of the model are structurally isomorphic with those of the modelled. This reflects the insight that functions are given the very possibility of their function by their structure.

Is the structure of a structure written into it or is it read into it by a modeller?

Anything which has a structure also is a structure. For example, a house _is_ a structure, but it also _has_ a structure, in fact many such. The structure which a house _is_, is quite unique and individual. My house and yours _are_ different structures. However, the structure which my house _has_ may exemplify a sort. Two houses on an estate may _have_ the same structure ("there's the same house again") but _be_ different structures. Conversely, the structure my house has qua load-bearing elements may be quite different from that which it has qua living spaces. To the engineer the important parts are girders and rafters, to the interior decorator the rooms and their connections ... to the former a hardboard dividing wall is negligible, to the latter crucial. But the structure which my house _is_ is the same to the engineer and the interior decorator ... it is what makes them talk about the same thing.

Similarly machines (and brains) both _are_ and _have_ structures. In the case of computers we are usually interested in the aspects of structure relevant to the storage and handling of information. Of course, computers are also physical objects. They must, for example, stand without collapsing. This is effected by both the structure of the machine and its material constitution. A machine made of metal and one made of tissue, whilst being identical with regard to information handling structure, would (probably) differ as regards their possible physical realisations.

Note in passing:

a) Structure does not imply construction. The human body is structure which grows and evolves, drops of snow are structures which form and a collection of pebbles thrown ashore by a stormy sea might constitute an abacus of great power.

b) Structures (and models) are not necessarily three dimensional entities. A structure is basically an ordered set of primitives (where the structure is defined over the elements) ... Ron Atkin (1972) has recently demonstrated a relationship between chess pieces, board squares and possible moves which possesses a geometrical representation in 53 dimensional space.

We are now in a better position to answer the question of whether the structure a thing has is written into the structure a thing is.

Each structure a thing has is (but is not caused by) an ordered set of primitives (order given by a syntax). This is independent of any modeller. However there are indefinitely many of these structures. The modeller reads a particular structure in, in the sense that he gives it preferential treatment as regards focus of attention.

We may take a horse chestnut leaf as a model or representation of (one aspect of) a human hand. We draw the connection. But there is still something about a horse chestnut leaf which makes it more suitable as a model of the human hand than, say, a beech leaf. It is I that make three billiard balls a model of three atoms but three billiard balls lend themselves (due to their structure) better than does a blob of jelly to be such a model.

Thus the structure a thing has, which makes it a good or bad model, is written into the structure itself (the structure of the salt crystal was discovered not invented). But many other structures are equally written into the structure; in choosing one as the root of a modelling relation, we are guided by our purposes ... the function to be fulfilled by the model.

Does any mention of similarity presuppose, as one of its conceptual criteria, an exempt agent? On a trivial level perhaps ... in the sense in which, if all living things were wiped out overnight by a mysterious virus there would be nothing which answered to the concept 'similar', because there would be no concepts at all! Equally, in this sense, there would be no stones, nor seas nor even undifferentiated matter.

But is it true at a more interesting level (in the sense in which there would be no dreams or pains)? I don't think so.

There would still be ordered sets of primitives ... there would still be disparate sets of primitives ordered by the same set of rules. Therefore there would still be 'similar' structures ... which represented each other ... instantiated in different ways.

The account so far sketched, makes use of the notion of similarity ... which is presently ill-favoured as an ingredient of the representation recipe. Why?

Evidently, similarity is not <u>sufficient</u> for representation. Two twins, or any two objects off a production line, are highly similar, but do not (necessarily) represent each other ... though a wax effigy of a person in Madame Tussaud's or a prototypical model of a car, in some sense less similar, might do so.

Is similarity <u>necessary</u> for representation? The usual answer is "no" ... anything can be taken as a model of anything else ... but is this as obviously true as is usually assumed? I do not think so.

Suppose after witnessing a traffic accident, at tea I take a salt pot to represent a car ... surely the salt pot doesn't resemble the car? (in anything but the most trivial way ... they are both physical objects).

But does it make sense to think of a single undivided entity, such as a salt pot, as representing anything?

Suppose I say, holding up a salt pot: "this represents the car" and then do not say or do anything else, my doing so was quite empty ... only when it is related to something else does it become a representation ... say the sugar bowl represents a stationary van and the pepper pot a pedestrian.

It is only systems which represent ... or, more correctly, only in systems that items can represent; but then in some sense the system does resemble what it (the system) represents. The relations of the salt pot to the sugar bowl e.g. motion of one relative to the other, are similar to those of the car relative to the stationary van ... this is just their point.

Structures are defined over primitives by a system of rules (a syntax). The primitives of two systems can only model each other via the structure. It is only as members of a system that primitives are made the primitives they are. The content of the constituents lies in their relations with the other elements of the structure. Examples could be taken from autonomous syntax, formal semantics or first order logic. In the case of the former, all that is required is that the relations between a fixed number of discrete primitives be stipulated by a grammar. What the language is (what the primitives are primitive of) is laid down (defined) by the syntax. (One of the central problems of semantics is whether there could be a formal semantics, in this sense - where semantic primitives have no intrinsic content). Two words may be spelled similarly ("bank" and "bank") but two letters cannot be. (c.f. How do you spell 'cat'? vs. How do you spell 'c'?). Sentences can be ungrammatical, words cannot. Correspondingly, systems can be representational, primitives cannot.

Thus, whilst we have found similarity to be insufficient for representation. We have not, so far, found reason to believe it not to be necessary for representation, in modelling at least; just reason to think that the unit of similarity is not the primitive but the system in which the primitive plays its part.

However, I have to admit that this is not initially quite so plausible in cases of representation other than by modelling, e.g. language use.

In the sentence "The ant is carrying a stick", the word "ant" presumably represents an ant ... but it doesn't seem to resemble one, except perhaps in their both being rather small!

Note though, that there have been attempts by psycholinguists to explain representation of words by resemblance.

Werner and Kaplan (1967) in their studies of onomatopoeia, the phenomenon of words whose sounds imitate non-verbal sounds e.g. "cuckoo", and "splash", found them to be a universal phenomenon - occurring in all known languages. Nevertheless, they are still rare and, as they never violate phonemic rules, they are at least partly conventional. It is moreover noteworthy that the sound systems of languages are chosen from a very limited sub-set of physiologically realisable vocalisations ... just the opposite of what one would expect if onomatopoeia played a significant role. The evidence for cross-modal similarity ... synaesthesia ... e.g. zigzag, is even more tenuous. On similar lines, Kohler (1947) has reported experiments showing that subjects spontaneously recognise relations between drawings of sharp objects and nonsense, words like "takete"; and Kaden (1955) has found that subjects tend to make characteristic errors of judgement of elevation of projected images of words like "climbing" or "falling". However, there seems no way to decide whether such effects are the cause or consequence of perceived similarity. Did the word "climbing" get chosen to mean what it does mean because it was in some way inherently 'similar' to some Platonic 'upness', or does it appear so similar (to those who claim it does) because of what it means? Instructively, there is no tendency for full synonyms to exhibit phonemic features in common e.g. "bachelor" and "unmarried man".

However, we should note that all of these attempts, weak as they are, are ones to show how the primitives of the system represent by resemblance ... which, from earlier considerations we would not expect to succeed.

How much more likely is it that sentences represent states of affairs by resembling them? How does the sentence "The ant is carrying a stick" resemble the state of affairs in which the ant is carrying a stick?

One sort of answer - the Davidson-Tarski Correspondance theory of truth and its semantic ramifications and Montague's Semantics can be seen as an attempt to elaborate an 'imitative semantics', albeit at a very abstract level, but a discussion of that goes beyond the scope of this chapter.

Note, however, it does at least seem plausible that the state of affairs in which the ant is carrying a stick is related to the sentence "The ant is carrying a stick" in a very similar way to that in which the state of affairs in which the ant is not carrying a stick is related to the sentence "The ant is not carrying a stick" (i.e. the state of affairs obtains whenever the sentence is true).

It is meaningless to talk of the structure of a representation; structure is relative to the processes which construct and use the representation. Pylyshyn (1975) has suggested we should speak of the structure of the representation relative to a "Semantic Interpretation Function" (S.I.F.). Then we can say representations "preserve the structure of what they represent" to the extent that the "same structure" is extracted by some appropriate S.I.F. Pylyshyn observes, interestingly that, "in that sense, the sentence, 'the book is on the table' can be said to preserve (part of) the structure of a scene containing a book on a table." Finally note, as Sloman (1975) has pointed out, the way some sentences represent is partly by analogy ... or similarity. Compare: "She shot him and kissed him" vs. "She kissed him and shot him." (The order of phrases is similar to the order of events).

On the other hand, although representations of all kinds are complex structures, linguistic representations have primitive parts which do not correspond straightforwardly to primitives of what they represent. "The man who shook hands with the Queen at t" is a complex representation which contains as a proper part "the Queen"; however, that which the complex representation denotes, say Jim Callaghan, certainly does not contain the Queen as one of his proper parts.

Summary of main points.

1. What is the relation between the model and the modelled?
2. The model 'stands in place of' the modelled ... it represents it.
3. All models are representations; not all representations are models; therefore models are a species of representation.

4. Models are used to explain/understand things.
5. Not all things which are similar to each other represent each other ... similarity not a sufficient condition for representation.
6. Is similarity a necessary condition for representation i.e. are all representations similar to what they represent? Possibly but, if so, at a very abstract level, at least for many sorts of representation e.g. linguistic.

Suggestion.

Models are a special case of representation in that they resemble what it is they represent - indeed, they can be used as they are in explanations because they resemble what they represent. (The similarity must play a role in producing the explanation, otherwise a microscope used to explain the workings of a tiny instrument, which turned out to be a minute microscope identical to it self, would count as a model).

In conclusion, whilst we are as yet unable to give an acceptable general theory of representation, we can say a few interesting things about particular cases of it e.g. modelling. The proper unit of representation is the system or structure within which primitives may represent. All representations appear to strive (but may fail) to denote ... they are representations - of; (although we may choose to deny that a representation not of anything is a representation at all). However, mere denotation seems too weak a notion to totally capture that of representation. Representations (being structures) always have structures and these correspond 'in some way' to what is represented. Some representations correspond directly with the represented. Others correspond with the structure of the procedure by which the thing is identified. The former type of representation lend themselves - because of their isomorphism - to being used in explanations as models.

Models, then, are a special case of representation which can be used in certain types of explanation because they resemble what they represent.

Some preliminary appreciation has now been gained of what, in the context of this thesis, is meant by the term "model". Further discussion will, however, be opaque without particular concrete examples. Accordingly, in the next chapter we shall look fairly closely at a case of 18th century formal model building, and in the chapter to follow, we shall discuss the impact of an intimate model (the clock) upon the development of theories of human mentality.

# CHAPTER TWO

## A FORMAL MODEL

"All things are Artificial; for Nature itself is nothing else but the Art of God ... to find the various turnings and mysterious process of this divine Art, in the management of this great Machine of the World, must needs be the proper Office of onely the Experimental and Mechanical philosopher ... he that will give a satisfactory Account of those Phaenomena, must be an Artificer indeed, and one well skill'd in the Wheelwork and Internal Contrivance of ... Automaticall Engines."

Henry Power (1664).

Jacques de Vaucanson, born in February 1709 in Grenoble, was the youngest of ten brothers and sisters. Something of a prodigy, he was fascinated by machinery from a very early age; as a small boy he managed to build a clock, which "marquait les heures assez exactement", out of nothing but wood and using only the few crude tools available to him. Only slightly after this, his mother, a strict and pious woman, allowed him to make and furnish a 'child's chapel' ... and he repaid her by equipping it with tiny angels complete with beating wings and automaton priests, who carried out certain sacerdotal duties. This experience stood him in good stead for his later career. We are told (Tence, 1836), perhaps apocryphally, that whilst a pupil at a Jesuit college, his mother of course expecting him to go into the church, he made and released some automaton angels which proceeded to fly around the chapel. The offending aviators were destroyed as heretical and the outraged Vaucanson took the opportunity as a pretext to leave the custody of the Order of Minims forthwith for Paris, where he gave himself over largely to a life of dissipation and revelling ... finding time however for in-depth studies of anatomy, music and mechanics.

Early on, Vaucanson had made plans for a pump at Lyons to raise much needed water to the town, but had been too shy and lacking in confidence to put them forward. In Paris, however, he came across a pump working on similar principles and his self-confidence appears to have been boosted. A great admirer of Coysevox's statue in the Tuileries of a rustic figure playing a flute, Vaucanson dreamed of animating it. He began plans for such an automaton, but a threatening letter from his uncle - who no doubt thought his nephew was taking leave of his senses - discouraged him. Three years later however, the urge returned irresistibly and he threw himself into the plans and calculations with such ardour that he fell ill. The hard work paid off handsomely though ... although various parts of the automaton were entrusted to several different artisans, when they were eventually assembled together the automaton played perfectly at its very first attempt. The spectacle was so awe-inspiring that Vaucanson's manservant fell on his knees in homage - thinking his master somewhat more than a mere mortal.

Vaucanson made several other automata: his famous duck and tabor player, to be discussed below, and at a fair in St. Germain in 1749 he exhibited three other automata: firstly, a Moor who swung a bell in one hand and a hammer with which to strike it in the other ... the Moor no doubt served the function of collecting a crowd; secondly, a country woman carrying a pigeon on her head and a wine glass in her hand ... on the word of command, she raised up her glass to the pigeon who filled it with red or white wine from its bill; thirdly, a grocer's stall to the counter of which an, initially seated, mechanical shopkeeper would bring ordered merchandise.

De Solla Price (1964) and Bedini (1964) have recently argued that many such early gadgets and machines were neither "trivial toys" nor "immediately useful inventions". Rather, they were <u>simulacra</u>. That is, the devices were models of a very special sort, models "... whose very existence offered tangible proof, more impressive than any theory, that the natural universe of physics and biology was susceptible to mechanistic explication". This leads Price to reverse the usual interpretation of the relationship between high technology and "pure Science" in the Hellenistic and Roman world. It is <u>not</u> the case, he suggests, that "... certain theories in astronomy and biology derived from man's familiarity with various machines and mechanical devices". On the contrary, "... some strong innate urge toward mechanistic explanation led to the making of automata, and (....) from automata has evolved much of our technology...". Another alternative might, of course, be that this urge towards mechanistic explanation led both to the construction of automata <u>and</u> to our practical technologies.

As Bruce (1977) rather inelegantly put it concerning Descartes: "He justified his mechanistic view of bodily movements after observing and constructing elaborate moving dolls. (Descartes thus anticipated not only later developments in science but also amusement park technology.)"

With Price's distinctions in mind, we shall attempt to answer the following question: Does the aim of understanding behaviour by simulating it also inform the work of Vaucanson or has the urge degenerated, as Brewster suggests, into the mere desire to amuse by displays of mechanical exuberance?

There can be no argument over the fact that Vaucanson's automata did indeed delight enthusiastic audiences. Demonstrated to courts, learned societies, and the lay (i.e., paying) public, Vaucanson's flute player, duck, and tabor-pipe player "astonished all Europe" (Brewster, 1832). So great was the demand in London that these mechanical figures could be viewed in operation at the Opera House in the Hay-market "at 1, 2, 5 and 7 o'clock in the Afternoon" (Vaucanson, 1742). When we consider both its imposing physical appearance and its virtuoso performance, it is hardly surprising that the flute player was so popular. According to Brewster "The body of the flute player was about five and a half feet high, and was placed upon a piece of rock, surrounding a square pedestal four and a half feet high by three and a half wide". The pedestal was packed with machinery, bellows and the like, and the trunk was riddled with pipes and small reservoirs. "These reservoirs were thus united into one, which, ascending into the throat, formed by its enlargement the cavity of the mouth terminated by two small lips, which rested upon the hole of the flute. These lips had the power of opening more or less, and by a particular mechanism, they could advance or recede from the hole in the flute. Within the cavity of the mouth there is a small movable tongue for opening and shutting the passage for the wind through the lips of the figure" (Vaucanson, 1742).

For the London exhibition, J. I. Desaguliers, LL.D., F.R.S., Chaplain to his Royal Highness the Prince of Wales, translated "out of the French Original" two of Vaucanson's most illuminating papers, "a Memoire, to the Gentlemen of the Royal-Academy of Sciences at Paris" and a letter to the Abbé de Fontain. The work was printed by T. Parker, and sold in the Long Room of the Opera House during the exhibition. It is thanks to this publication that we have a fairly full and accessible record of the philosophy of Jacques de Vaucanson.

What, then, did Vaucanson imagine that he was doing? In the first part of the memoire, he describes in considerable detail the structure of the German flute, its components and their inter-relationships. He also describes how such an instrument is played by a man, specifying the motions of, and positions taken up by, the lips, tongue, and fingers in obtaining various notes, sequences and timbres. Finally he discusses the properties of the "strength" and velocity of wind necessary to produce certain effects, and mentions the gross anatomy which underlies these variations in the force of the air stream.

The difficulty, however, is to know how these elements - instrument, mouth, fingers, airflow and so on - work together, simultaneously and sequentially, in a co-operative fashion to produce the desired result. There is a sense in which Vaucanson's seemingly "objective description" is really a hypothesis concerning the mode of operation of man plus flute. Vaucanson accordingly summarises the position thus:

> "These, Gentlemen, have been my Thoughts upon the sound of Wind-Instruments and the Manner of modifying it. Upon these Physical Causes I have endeavour'd to found my Enquiries; by imitating the same Mechanism in an <u>Automaton</u>, which I endeavour'd to enable to produce the same Effect in making it play on the German Flute".

Vaucanson clearly regards his automaton as a <u>test</u> of the principles he has formulated: "<u>It will then follow</u> ... <u>according to the Principle</u> settled in my First Part, the flute will give a low Sound: and <u>this is confirmed by Experience</u>". This theme is stressed in the second part of the memoire. The construction of the automaton is itemized and each part of the device is carefully linked with the proper part of the previous functional description. The four elementary operations are stated as follows:

> "By the action of the Lever, which increases the Opening of the Lips, the Action of a Living Man is imitated, who increases that Opening for the Low Sounds. By the Lever, which draws back the Lips, I imitate the Action of a Man who removes them farther from the Hole of the flute, by turning it outwards. By the Lever which gives Wind from the unloaded Bellows, I imitate the weak Wind which a Man gives when he drives it out of the Receptacle of his Lungs, by only a light Compression of the Muscles of his Breast. By the Lever which moves the Tongue, in unstopping the Hole Thro' which the lips let the Wind pass, I imitate the Motion of a Man's Tongue, when he pulls it back from the

Hole to give Passage to the Wind to articulate such a Note".

But how should these elementary motions be combined? The first note (D) is formed by a relatively simple procedure, described on page sixteen of the memoire. In order to produce E, however, various compensatory adjustments must be made:

> "If I wou'd make the Flute found the Note above, namely E, to the four first Operations for D, I add a fifth; I fix a Bar under the Lever, which raises the third Finger of the Right Hand to unstop the sixth Hole of the Flute; and I make the Lips to come a little nearer to the Hole of the Flute, by fixing or making a little lower the Bar of the Barrel which held up the Lever for the first Note, namely for D. Thus, giving an Issue to the Vibrations sooner, by unstopping the first Hole from the End, as I said above, the Flute must found a Note above; which is also confirm'd by Experience".

All now continues smoothly throughout the remaining notes of the first octave; only the appropriate programming of finger positions needs to be worked out. Further problems of co-ordination arise, however, when the simulation of the second octave is attempted:

> " ... we must change the Situation of the Mouth, that is, we must place a Bar under the Lever which serves to push the Lips beyond the Diameter of the Hole of the Flute, and thereby imitate the Action of a Living Man, who in that Case turns the Flute a little inwards. Secondly, we must fix a Bar under that Lever, which bringing the Lips towards one another diminishes their Opening; as a Man does to give a less Issue to the Wind. Thirdly, a

Bar must be fix'd under the Lever which opens the Valve of that Receptacle that contains the Wind coming from those Bellows which are loaded with two Pounds; because the Wind being then driven with more Force, acts in the same Manner as that with which a Living Man blows by a stronger action of the Pectoral Muscles ... "

The succeeding paragraphs of part two describe the further adjustments which are necessary if the higher notes of the second octave, and the whole of the third octave shall be faithfully produced. There follows a discussion of how to obtain correct tempi and phrasing, interspersed with remarks about the very subtle modifications which are needed to produce a "swelling of notes" and an echo-effect. It is clear that Vaucanson must have run into some horrendously difficult problems in debugging his device, for he closes the Memoire with the transparently heartfelt comment that:

"The fear of tiring you, GENTLEMEN, has made me pass over a great many little Circumstances, which tho' easy to suppose are not so soon executed: the Necessity of which appears by a View of the Machine as I have found it in the Practice."

There seems little doubt, then, that Vaucanson was concerned to formulate and validate - in the most precise and formal language available to him - a theory of the German flute player. Certainly this is how Vaucanson's translator interpreted him. In the preface, Desaguliers boldly claims that " ... this Memoire ... in a few Words gives a better and more intelligible Theory of Wind-Musick than can be met with in large Volumes". Explicitness is the paramount virtue. As Vaucanson himself puts it, his aim is to imitate "by Art all that is necessary for a Man to perform in such a Case."

This refrain recurs in the letter to the Abbé de Fontaine. Here, Vaucanson compares the difficulties he encountered in the construction of the flute automaton with the new problems posed by the tabor-pipe. He had expected this latter simulation to be somewhat easier; the task is merely " ... to articulate Sound by Means of a Pipe of three Holes only, where all the Tones must be performed by a greater or less Force of the Wind, and half stopping of the Holes to pinch the Notes".

Vaucanson finds, however, that the essence of pipe playing lies in its speed of operation " ... every Note, even Semi-Quavers, must be tongued", for otherwise the sound of the instrument is "not at all agreeable". It turns out that, for some melodies, his automaton actually performs better than a human player:

> "In this the Figure out-does all our Performers on the Tabor-Pipe, who cannot move their Tongue fast enough to go thro' a whole Bar of Semi-Quavers, and strike them all. On the contrary, they slur above half of them; but my Piper plays a whole Tune and tongues every Note".

In this respect, the automaton can be regarded as a competence model. Contingent limitations are lifted. Just as our finite wind and memory capacities ensure that many sentences which are strictly within our formal competence will never be uttered or comprehended, so are there many "Minuets and Rigadoons" which are formally within a (human) pipe player's competence but which are not attainable due to the tongue's lack of agility. Vaucanson's automaton is able to transcend (some) of these difficulties.

The fact that Vaucanson's theory (i.e., his automaton) must actually perform enables a variety of surprising phenomena and unforeseen relationships to be uncovered; "Discoveries of Things which could never have been so much as guessed at", as Vaucanson writes. For instance, of all wind-instruments, the pipe must be "one of the most fatiguing to the lungs".

> "For in the playing upon it, the Performer must often strain the muscles of his Breast with a Force equivalent to a Weight of 56 Pounds; For I am oblig'd to use that Force of Wind, that is, a Wind driven by that Force of Weight, to sound the upper B which is the highest Tone to which this Instrument reaches: Whereas one Ounce only is sufficient to ... produce the lowest Tone, which is an E".

The direction of information flow here is interestingly similar to the "intellectual boot strapping" talked of by present day simulation theorists. From observation of the world we derive enough information to construct a model - then from this model we infer further properties of the world. "The Performer <u>must</u> often strain etc. ... for I (qua simulator) am oblig'd to etc. ...".

Most interesting of all, however, is Vaucanson's discovery that the force required to produce a particular note is not solely dependent upon the nature of the note itself, but is rather determined, in part, by the note which precedes it:

> "That Wind, for example, which is able to produce a D following a C, will never produce it, if the same D is to be sounded next to the E just above it; and the same is to be understood of all the other Notes".

Vaucanson accordingly finds it imperative to have at least "twice as many different Winds, as there are Tones, besides the Semi-Tones, for each of which a particular Wind is absolutely necessary". It is pleasing to observe that a more recent formal model of one aspect of musical competence reaches a similar conclusion; Longuet-Higgins' (1976) computer program for the transcription of classical melodies into standard notation demonstrates very conclusively that "the tonality of any note cannot in general be established unambiguously until the following note has been heard". In both cases, then, a simple "chain-reflex" account of musical competence is falsified. This need to utilize "context-sensitive" machinery had first impressed itself upon Vaucanson during the construction of his most famous device, "the marvel of the last century" as Helmholtz (1901) wrote - an artificial duck. Vaucanson's primary aim here was "to represent the Mechanism of the Intestines". In addition, however, he made the duck capable of many other overt motions, which included stretching out its neck ("to take Corn out of your Hand"), moving the neck from left to right, flapping its wings, and raising itself up on its legs. If these co-ordinations were to be achieved without the bird falling over, the "same" piece of elementary machinery (the same structure) had to change its "function" dependent upon the overall pattern of behaviour that was being executed. Upon observing the duck, "Persons of Skill and Attention" will see that:

> " ... what sometimes is a Center of Motion for a Moveable Part, another Time becomes moveable upon that Part, which Part then becomes fix'd. In a word, they will be sensible of a prodigious Number of Mechanical Combinations".

It is the duck that affords a final illustration of Vaucanson's motives. The entire internal mechanism of the automaton is "exposed to view". This gesture is made, Vaucanson writes, because "my Design (is) rather to demonstrate the Manner of the Actions, than to show a Machine". Because the intellectual cards are more than usually on the table, Vaucanson can risk a little joke against both his public and his competitors:

> "Perhaps some Ladies, or some People, who only like the Outside of Animals, had rather have seen the whole cover'd; that is the Duck with Feathers. But besides, that I have been desir'd to make every Thing visible; I wou'd not be thought to impose upon the Spectators by any conceal'd or juggling Contrivance".

From our current vantage point, then, the most interesting fact about Vaucanson is not that he constructed automata, nor that his automata were so superbly realized. Vaucanson's main achievement lies rather in the clarity with which he perceived and articulated the character of the explanatory mode he sought to attain. Recent discussions of the philosophy of psychology have stressed a number of desiderata which are directly paralleled in Vaucanson's memoires. For example:

1. That the theories must be explicit:

It is generally conceded that psychological theories phrased in 'ordinary language' suffer from an incurable vagueness; they only work (if at all) when a liberal dose of human intelligence has been added to them, thereby opening the way to disagreements over interpretation. Chomsky's (1965) requirement that a grammar should generate without benefit of intuition all and only the 'objects' within its domain has accordingly been called "the most important conceptual demand on psychology of this century"(Suppes, 1968). The necessity of explicitness is only too obvious to Vaucanson who describes all his efforts as "raised on the solid Principles of Mechanicks", and points out (in a passage we have previously quoted) the difference between supposing that something will work and showing that it does.

2. That occult entities may not interact with machines:

A number of early mechanistic theories in psychology have always been vulnerable to attack on the grounds that they postulated (or at least left room for) a ghost within the machine. As late as the nineteenth century, one finds Johannes Müller (1842) proposing that "The fibres of all the motor, cerebral and spinal nerves may be imagined as spread out in the medulla oblongata, and exposed to the influence of the will like the keys of a pianoforte".

Needless to say, Vaucanson will have no truck with flute-players within flute-players or ducks within ducks. It is imperative for Vaucanson that the " ... Machine, when once wound up, performs all its different Operations without being touch'd any more". No hidden soul or mind pulls the levers which cause the wings to flap.

3. That stimulus-response psychology nonetheless leaves out a vital ingredient:

Whilst Vaucanson is a behaviourist in the sense that behaviourism may be contrasted with vitalism or dualism, he clearly cannot believe in an 'empty organism' approach to psychology. No manipulation of reinforcement contingencies will provoke the duck to partake of liquid refreshment ("I forgot to tell you, that the Duck drinks, plays in the Water with his Bill, and makes a gurgling Noise like a real living Duck".) - unless it has been constructed to do so. You can take the duck to water, but only the fact of Vaucanson's having got the wiring-diagram right will make it drink. The oft-repeated refrain in the memoire is " ... we must fix a Bar so that ... ". Or as Quine (1969) puts it, the behaviourist is " ... cheerfully up to his neck in innate mechanisms".

4. <u>That the behavioural repertoires of organisms must be characterised</u>:

Chomsky has emphasized that any finite corpus of sentences is only a selection from the infinitude of examples which are, in principle, known to a competent speaker. The point generalizes to other behaviour domains. As Fodor (1968) puts it: "The potential behaviour of the organism defines a space of which its actual behaviour provides only a sample". It is therefore a condition upon any theoretically illuminating simulation that, unlike a gramophone recording, it should generate the entire behavioural repertoire of the organism within a particular domain.

Vaucanson is not entirely insensible of such points, although the recursive function theory that would solve the problem was, of course, not available to him. His tabor and pipe player does not quite manage to play all and only the sequences of notes which are melodies within some formal musical system. It does nonetheless play some " ... twenty tunes, Minuets, Rigadoons, and Country-dances". (Sadly, we formulate our mathematics in such a fashion that twenty is no closer to infinity than one is.) The pipe, however, " ... employs but one Hand" of the automaton. In addition:

> "The figure holds a Stick in the other, with which he strikes on the Tabor single and double Strokes, Rollings varied for all the Tunes, and keeping Time with what is played with the Pipe in the other Hand. This motion is none of the easiest in the Machine; for sometimes we must strike harder, sometimes quicker, and the Stroke must always be clean and smart, to make the <u>Tabor</u> sound right. The Mechanism for this consists in an infinite Combination of Levers, and different Strings, all moved to exactness to keep true to

the Tune".

We suspect that "an infinite Combination of Levers" really means "many levers"; but at least the importance of modelling a reasonably large and varied behavioural repertoire is obvious to Vaucanson. In this respect, there is a parallel between the tabor-pipe and the flute automata; as Brewster remarks about the latter: "The airs which it played were probably equal to those executed by a living performer."

5. <u>That the theorist must choose an appropriate level of representation for his simulation</u>:

Any formal theory should set boundary conditions for the phenomena which fall within its scope. Vaucanson's duck " ... stretches out its Neck to take Corn out of your Hand, it swallows it, it digests it, and discharges it digested by the usual Passage". These, then, are the limits of what is being modelled. Vaucanson continues:

> "I don't pretend to give this a perfect Digestion, capable of producing Blood and nutritive Particles for the Support of the Animal. I hope no Body will be so unkind as to upbraid me with pretending to any such thing. I only pretend to imitate the Mechanisms of that Action in three things, vis. First, to swallow the Corn, <u>secondly</u> to macerate or dissolve it; <u>thirdly</u>, to make it come out sensibly changed from what it was."

Let us say, with Fodor, that a machine is <u>weakly</u> equivalent to an organism when the behavioural repertoire of the machine is identical with that of the organism within a particular domain. Let us furthermore say that "a machine is <u>strongly</u> equivalent to an organism in some respect when it is weakly equivalent in that respect and the processes upon which the behaviour of the machine are contingent are of the same type as the processes upon which the behaviour of the organism are contigent." This notion of weak and strong equivalence is obviously related to the distinction that many modern computer theorists draw between "artificial intelligence" and the "simulation of behaviour". Clearly, it is the notion of strong equivalence which is of interest to the empirical scientist:

> "The Food is digested as in the real Animals, by Dissolution, not Trituration, as some Natural Philosophers will have it."

It was likewise the strong equivalence of flute-player and flute-playing automaton which so impressed the French Academy of Science. This learned body:

> "... did not hesitate to state that the machinery employed for producing the sounds of the flute, performed in the most exact manner the very operations of the most expert flute-player and that the artist had imitated the effects produced, and the means employed by nature with an accuracy which exceeded all expectation."

Conclusions.

It should be obvious, then, that Brewster's claim that the primary objective of Vaucanson's work was "to astonish and amuse the public" is hardly fair. Vaucanson was an entertainer, but he was also deeply committed to the development of an explanatory psychology.

One might wish to praise Vaucanson both for his achievements and for his modesty:

> "I own freely, that I am surpriz'd myself to see and hear my Automaton play and perform so many and so differently varied Combinations."

And above all, for his refusal to succumb to wishy-washy metaphysics about the limits of psychology:

"And I have been more than once ready to despair of succeeding; but Courage and Patience overcame every Thing".

Yet despite Vaucanson's fame throughout the eighteenth and nineteenth centuries, his name is not to be found in any standard (twentieth century) histories of science; he is occasionally mentioned in histories of technology, albeit briefly and only in the context of his later contributions to industry. These contributions were by no means meagre. As Bedini reminds us, Vaucanson was responsible for pioneering the development of machine tools. In 1741, having been appointed an inspector in the French silk factories, Vaucanson invented and perfected an apparatus, wrongly attributed to Jacquard for the automatic weaving of brocades. In 1760, he developed an industrial metal cutting lathe with prismatic guideways, inspiring Maudslay, a generation later, to continue work on machine tools in general and metal cutting lathes in particular. Later, as examiner of new machine inventions for the Academie Royale des Sciences, Vaucanson designed countless machines, including one for producing an endless chain.

Moreover, in constructing his machines, Vaucanson was obliged to make many technological breaks through. Vaucanson was the first to make use of flexible tube of India rubber (caoutchouc), which he had employed in representing "the Mechanism of the Intestines". "While M. Vaucanson was engaged in the construction of these wonderful machines, his mind was filled with the strange idea of constructing an automaton containing the whole mechanism of the circulation of the blood. From some birds which he made he was satisfied of its practicability; but as the whole vascular system required to be made of elastic gum or caoutchouc, it was supposed that it could only be executed in the country where the caoutchouc tree was indigenous. Louis XVI took a deep interest in the execution of this machine. It was agreed that a skilful anatomist should precede to Guyana to superintend the construction of the blood vessels, and the King had not only approved of, but had given orders for, the voyage. Difficulties, however, were thrown in the way: Vaucanson became disgusted, and the scheme was abandoned." (Brewster, 1832).

Nevertheless the disparaging label of "toy-maker" seems to have stuck to him, just as it had done to Hero in earlier times. Even current day automata theorists persist in this libellous misrepresentation Raphael (1976) for example, concludes his discussion of the models of Vaucanson (and others) with the remark that "These eighteenth century gadgets were developed purely for their entertainment value."

It has, however, been pointed out by Cohen (1966) that at least one major nineteenth century scientist - Hermann von Helmholtz - did take Vaucanson seriously. After describing the automata of Vaucanson and of the elder and younger Droz, Helmholtz (1901) comments:

> "That men like those mentioned, whose talent might bear comparison with the most inventive heads of the present age, should spend so much time in the construction of these figures which we at present regard as the merest trifles, would be incomprehensible, if they had not hoped in solemn earnest to solve a great problem."

The "great problem" in question was, of course, the same problem on which Helmholtz and his colleagues were engaged, namely, to complete the "mechanization of the world picture" (Dijksterhuis, 1961) by bringing physiology and psychology within its scope.

Helmholtz is aware, however, that serious scholars other than himself had taken a very considerable interest in these masterpieces of simulation. Following a public exhibition of the writing-boy made by the elder Droz, Helmholtz notes that " ... this boy and its constructor, being suspected of the black art, lay for a time in the Spanish Inquisition, and with difficulty obtained their freedom."

The tribute is well-taken. Throughout history, those who accused the automata-makers of necromancy were at least closer to the truth than those who accused them of frivolity.

# CHAPTER THREE

## AN INTIMATE MODEL

"So if unprejudiced you scan
The goings of this clockwork man
You find a hundred movements made
By fine devices in his head;
But 'tis the stomach's solid stroke
That tells this being what's o'clock."

Matthew Prior (1721).

Clockwork has been responsible for a great many formal (purpose built) models used by simulating natural scientists ... just think for a moment of the astrolabes, planetaria, orreries, planispheres, equatoria, armillary spheres and astronomical clocks of former times. However, in this chapter we are concerned with the use of the clock as an intimate model - a pre-existing entity seen to be similar in conceptually interesting ways to something the modeller wishes to understand.

The clock has frequently figured as an intimate model, perhaps first of all in astronomy. Oresme (1370) compared God's creation of the heavens to a man's creation of a clock ... in both cases what is made is later self moving and without the necessity of further intervention on the part of the creator. Oresme took the model much more seriously than this may suggest however ... indeed, he went so far as to maintain that the actual mechanism of the heavens must be similar to that of a clock: "In the absence of any resistance similar to the balance which regulates the movement of clock hands, the speed of the spheres would not stop increasing so that to keep the stars' rate, the divine clockmaker had to calculate a complicated system of actions and reactions."

The use of the clock model was by no means confined to the astronomically naive. The supreme Kepler (1605) wrote: "I am now engaged in investigating physical causes; my goal is to show that the celestial machine is not the likeness of the divine being, but is the likeness of a clock."

Gradually, however, scientists lowered their eyes from the skies to the earth, whilst firmly retaining the same model. Thus Boyle (1686) writes that: "The world is like a rare clock such as that at Strasbourg, where all things are so skilfully contrived. that the engine once being set going ... the motions ... do not require the peculiar interposing of the artificer, or any intelligent agent employed by him", and for Powell (1661) the world was "a kind of Automaton or Engine that moves of it self much like a great Clocke with wheels and poyzes and counter-poyzes."

Once the universe and within it the world, had been conceived of as like a clock, the way was clear to think of the denizens of the world in a similar way. An intermediate position has sometimes been adopted however, where men are seen as mere components of the vast machine of nature ... for example as the bell on a clock, whose function is to chime the praises of God. Cristobal Goncalez (1609): "If we look at a clock we shall find therein a whole host of springs, wheels, chains, pins, cogs and weights, all of which move and function so that a bell, placed in the topmost part of the mechanism rings and strikes the hours. But if this bell were not to ring, nor to keep time, we should say, and rightly, that the whole of this cunning mechanism was rendered useless. After this fashion we may philosophize about the whole fabric of the world, which is like a clock, set by God on those mighty wheels (circles) of the heavens, some of which move slowly, others quickly, some turn one way, others another, and all in time with the primum mobile. And having created the elements and all the other creatures, which are the chains and weights of the clock, and having set man, the most perfect of creatures, like the bell atop all of them, so that like a well-regulated clock he should give perpetual praise and glory to God ..."

Soon, however, man was not merely a component of a machine, but a machine himself. Hobbes (1651): "For seeing life is but a motion of links the beginning whereof is in some principal part within; why may we not say, that all automata (engines that move themselves by springs and wheels as doth a watch) have an artificial life? For what is the heart but a spring; and the nerves, but so many strings; and the joints, but so many wheels, giving motion to the whole body, such as was intended by the artificer? Art goes yet further, imitating that rational and most excellent work of nature, man." (Both Hobbes' and Goncalez' models were pre-figured by Suso's 14th century conception of the soul and body as like the various parts of a single clock which, when they were in accord, rang a bell symbolising the surge of joy experienced by the heart moving towards God.)

With the comparison even of the rational part of man to a time-piece, the stage is set for a discussion of human psychological models in terms of clocks.

The influence of clocks on <u>psychological</u> models could itself be traced from the present day (Deutsch, 1960: "The mechanism of a clock is related in the same way to its behaviour as the neural mechanism of an animal to the behaviour which psychologists attempt to explain.") right back to the ancients ... were one to begin with clepsydrae and hydraulic models. But we shall begin rather later. According to some sources (Encyclopaedia Britannica 1929), the first (weight powered) clock was built by Pope Silvester the Second in A.D. 996, although main-spring clocks, called Nuremberg eggs, were not being made until the sixteenth century. However, by 1265, when he started to write the <u>Summa Theologica</u>, Aquinas had obviously given the parallel between clocks and animal functioning some thought:

"Et idem apparet in motibus horologiorum et omnium ingeniorum humanorum quae arte fiunt ... Et propter hoc etiam quaedum animalia dicuntur prudentia vel sagacia: non quod is eis sit aliqua ratio vel electio."

Nearly four hundred years later, whilst a student at La Flèche, René Descartes studied Aquinas' Summa (Gilson, 1914), and this study no doubt formed one root of his conviction that animals "have no reason at all and that it is nature which acts in them according to the disposition of their organs, just as a clock which is composed of wheels and weights is able to tell the hours and measure the time more correctly than we can do with all our wisdom." (1637). Four years before this (Treatise of Man), Descartes had fully adumbrated a completely mechanical model of man: "I wish you to consider, finally, that all the functions which I attribute to this machine, such as digestion ... nutrition ... respiration, waking and sleeping; the reception of light, sounds, odours..; the impression of ideas in the organ of the sensus communis and imagination; the retention ... of these ideas in the memory; the interior movements of the appetites and passions; and finally the movements of all the external members ..; I desire, I say, that you consider that these functions occur naturally in this machine solely by the dispositions of its organs, not less than the movements of a clock or other automaton ... Thus it is not necessary to conceive that it has a nutritive soul, or sensitive soul, or any other principle of motion and life except its blood and spirits ... which have no other nature than (that) found in inanimate bodies." Even four years before this, Descartes was already occupied with the problem of mechanical life. In his Cogitationes privatae he writes: "From the very perfection of animal actions we suspect that they do not have free will." Readers of more recent literature will recognize the above argument: when a man is more skillful than a computer, the conclusion is drawn that men are not machines; but similarly, when the computer outperforms the man, exactly the same conclusion is drawn!

(Actually, not all commentators concur with Descartes in his suspicion. As Büchner observed: "A watch, ... has, as is frequently said, a head of its own; it goes, it stops, frequently in a manner that makes it appear to have a will of its own; But how singularly crude and simple is the combination of matter and force in these machines, compared with the complex mechanical and chemical composition of the animal organism.")

In Cogitationes privatae - Descartes proposed the construction of a man-machine to be worked by magnets and, according to Père Poisson (1670), Descartes also drew blueprints for a flying pigeon and a pheasant hunted by a spaniel. And the motivation for all this? "Voulant verifier par experience ce qu'il pensoit de l'ame des bestes, il avoit inventé une petite machine qui representoit (un homme)". Descartes' observations on the regularity and perfection of animal behaviour, combined with the mechanistic leanings which had been fuelled by his dissections, led him to assert that animals are machines; he had also been greatly impressed by the hydraulically operated statues in the royal gardens at St Germain en Laye, and reasoned that if man can make such machines, then God could certainly create even more perfect ones, namely, animals.

During the period dominated by the genius of Descartes, the clock analogy was pursued at varying levels of sophistication. In his Physica Clauberg (1664) likened the living animal to a clock in running order, and a dead one to a clock that has stopped. Indeed according to Sir Kenelm Digby (1658) "The King of China upon his first seeing a watch thought it a living and judicious creature because it moved so regularly of itselfe, and believed it to be dead when run out; till the opening of it and the winding it up, discovered unto him the artifice of it."

In Exeter Cathedral, on the tomb of Lady Dodderidge, who died in

1614, there is an epitaph which also compares death to the breakdown of a clock, though in the most charming of manners; resurrection is likened to the clock's repair;

"As when a curious clock is out of frame
a workman takes in peeces small the same
and mending what amisse is to be found
the same reoiynes and makes it true and sound
so god this ladie into two parts tooke
too soone her soule her mortall corse forsooke
But by his might att length her bodie found
shall rise reoiyned unto her soule now cround
Till then they rest in earth and heaven sundred
att which conioyned all such as live then wondred."

For Malebranche (1678) who followed the main tenets of Cartesian biology but avoided mechanism by the doctrine of Providentialism, the apparent intelligence of animals (and watches) is simply a manifestation of God's divine intelligence quotient. Accordingly, Malebranche writes:

"Les mouvements des bêtes ... marquent une intelligence: mais cette intelligence n'est point de la matière, elle est distinguée des bêtes, comme celle qui arrange les roues d'une montre, est distinguée de la montre."

But the more interesting insights of the period concern the possibility that the nature of thought itself can be explained mechanistically.

Common sense suggests a continuity between the mental functioning of animals and that of man. If men are intelligent, surely animals are also - albeit to a lesser degree. Descartes, of course, denied this (in some of his writings at least) and argued that animals are purely automatic. It was now a simple step for La Mettrie (1748) to supply the missing position - animals are indeed automata, but so are men.

As Frederick the Great wrote in his Eulogy, La Mettrie believed that "thought is but a consequence of the organization of the machine". In fact, La Mettrie put it rather more strongly than that:

> "Thought is so little incompatible with organized matter, that it seems to be one of its properties on a par with electricity, the faculty of motion ... etc."

For La Mettrie, the clock analogy was an integral mainstay of this conviction: "Man is to the ape, to the most intelligent of animals, as the pendule planetaire of Huygens is to a watch of Julien Leroy." Elsewhere in L'Homme Machine he writes: "The body is but a watch, whose watchmaker is the new chyle." However, if one wants to demonstrate that a functioning system is independent of a celestial intelligence (as La Mettrie unquestionably did), there is something patently lacking in the model of a simple clock - a device, that is, which requires a winder, a great key-turner in the sky. Accordingly, La Mettrie proposes the following innovation: "Man is but an animal, or a collection of springs which wind each other up." And elsewhere he writes: "The human body is a machine which winds its own spring." It is in this fashion that La Mettrie can conceive of man as being both mechanical and self-regulating.

Was the independence issue the only stimulus to La Mettrie's developing thought, or was he heir to the whole automata tradition in Western science? Budding mechanists in Heron's time visited automatic theatres for their amusement and timed their eggs by clepsydrae; in Descartes' time weight-and-pulley clocks were sufficiently complex to make interesting models of the mind. By the time La Mettrie was writing could it be that self-winding watches were already in vogue?

It would appear so. The oldest known <u>patented</u> self-winding watch was that of Louis Recordon in 1780, but the principle had been discussed long before it was realized that, when patented, time equals money. In 1751 Kratzenstein, the inventor of one of the eighteenth century's most famous talking heads, is reputed to have designed a clock which wound itself as a consequence of variation in temperature (Chapuis & Jaquet, 1955). Before this, in 1678, Abbé John of Hanteville supposedly made "a means to provide that the weight of the pendulum should be wound up by the guidance of several pinewood boards, placed transversely in two slide ways, the said boards continually rising and falling according to the humidity or dryness of the air."

Still earlier, in 1651, Daniel Schwenter, the Professor of Mathematics at Wittenburg, entitled a chapter of his opus magnum: "Eine Uhr so man nicht aufziehen darff mit immer wärender Bewegung sonder grossen unkosten zu machen." (A watch which need not be wound because fitted with a continuous motion mechanism cheap to make). Interestingly, he starts the chapter off: "It is a well known fact that the human heart, like a going watch, is incessantly in motion and beats day and night." Breguet, on the other hand, suggests that the very first self-winding watch was made in 1600 by a Jesuit father. In 1686 an advertisement appeared in the <u>London Gazette</u>. The advertisement read: "Lost, a watch in black shagreen studded case, with a glass in it, having only one Motion and Time pointing to the Hour on the Dial Plate, the spring being wound up without a key, and it opening contrary to all other watches, 'R. Bowen, Londini, fecit' on the back plate."

Unfortunately, there is some doubt as to whether this watch, although being keyless, was indeed self-winding. Although the modern method of winding by means of a knurled knob did not come in until 1820 (and was not in general use until 1870) it now seems likely that the Bowen watch was powered by the pumping of a spring plunger on the rim of the watch body. More reliably, if we prefer not to believe all that we read in the press, Chapuis and Jaquet inform us that the first really genuine self-winding watch was invented by Abraham-Louis Perrelet around 1772.

All things considered, it seems reasonable to assume that the idea of a self-winding watch is certainly current by 1748, when L'Homme Machine was published. Furthermore, it is clear from the researches of Chapuis and Jaquet that the period which immediately surrounds 1748 is characterized by some fairly intense competition to develop small, accurate, and reliable self-winding watches. It is more than possible, then, that a technological advance in clockmaking led to the development of La Mettrie's psychological theory.

Familiarity with technology is not a totally unrecognised influence on psychological theory, as Romanes (1885) puts it: "Hobbes was perfectly right in saying that with respect to its movements the animal body resembles an engine or a watch; and if he had been acquainted with the products of higher evolution in watch-making, he might with full propriety have argued, for instance, that in the compensating balance, whereby a watch adjusts its own movements in adaptation to external changes of temperature, a watch is exhibiting the mechanical aspects of volition. And, similarly, it is perhaps possible to conceive that the principles of mechanism might be more and more extended in their effects, until, in so marvellously perfected a structure as the human brain, all the voluntary movements of the body might be originated in the same mechanical manner as are the compensating movements of a watch;".

Neither is it a coincidence that the master clock makers of the time were also, in general, responsible for the most ingenious automata; at first as jacks (automated figures to strike the hours), then as elaborate ornamentation, and simulations (that is, as theories of the natural world). Pierre Jaquet-Droz (1721-1790), an eminent horologist of Neuchâtel was also famous for his singing birds, and together with his son (Henri-Louis, 1752-1791) made astonishing mechanical puppets. There grew up around them a whole team of automata-makers: Leschot, Maillaidet and James Cox of Shoe Lane, London, who was renowned for his museum of "mechanical puppets, singing bird boxes and strange clocks" which opened in Spring (sic!) Gardens in 1773.

As can be read above, the philosophes of Europe had plenty of stimulus to keep them occupied with the problem of mechanistic explanation during these times. Only ten years before the publication of L'Homme Machine, Vaucanson had exhibited his spectacular duck, and his tabor and flute players. In 1745, Vaucanson had hoped to create "moving anatomies" which would reproduce life artificially, and he only gave up the project through lack of adequate research funds. As we have seen, Descartes had thought of making a man-machine and La Mettrie too writes that "to make a talking man, a mechanism, is no longer to be regarded as impossible."

L'Homme Machine is arguably the most important psychological text of the eighteenth century - certainly for those with materialist inclinations. La Mettrie had realized that the central problem of the conception of man as machine was neither moral or practical but cognitive: how are we to make sense of the notion that matter is endowed with the faculty of thought? The key, La Mettrie affirmed, lay not in matter itself but in its organization: "to ask whether matter can think, without considering it otherwise than in itself, is like asking whether matter can tell time."

Criticisms of the clock analogy (and, by extension, of the whole research programme of the mechanists) were, of course, as frequent in the eighteenth century as they are today. Some of the objections were fairly trivial. Bougeant (1739), for instance, pointed out that whilst Descartes had tried to represent animals as clocks, it was nevertheless the case that we do not caress our watches. But, of course, explanatory models are necessarily impoverished. Everything can't be true of a model that is true of the modelled. If it were, then the model would <u>be</u> the modelled. A brain is not a <u>model</u> of a brain, it is one. And in any case, Bougeant's remark seems peculiarly inappropriate in the light of the later Romantic Tradition in which people <u>did</u> fall in love with (and presumably caressed) automata such as Olympia. Moreover, even in 1632, Sir John Suckling was actually comparing lovers to timepieces:

> "That none beguiled be by times quick flowing
> Lovers have in their hearts a clock still going;
> For though Time be nimble, his motions
>         are quicker
>         and thicker
> where Love has his notions:
>
> Hope is the main spring on which moves desire,
> And these do the lese wheels, fear, joy inspire;
> The ballance is thought, ever more
>         clicking
>         and striking
> and ne'er giving oer.

Occasion's the hand which still's moving round,
Till by it the Critical hour may be found,
And when that falls out, it will strike
kisses
strange blisses
and what you best like."

Another of Bougeant's criticisms, that: "an inner persuasion makes the Cartesian hypothesis repugnant" had earlier been stated by La Motte (1721). He wrote that it was a "depraved way of Reasoning that could dare to make them (animals) mere Machines, or pieces of clockwork." This objection, commonly enough raised even nowadays, was also voiced by Noel Antoine Pluche, the Jansenist author of Spectacle of Nature, who criticised Locke as "a man who degrades our soul far enough to consider it a soul of clay." We can let La Mettrie answer this attack for himself: "The excellence of reason does not depend on a big word devoid of meaning (immateriality), but on the force, extent and perspicuity of reason itself. Thus 'a soul of clay' which should discover, at one glance, as it were, the relations and consequences of an infinite number of ideas hard to understand, would evidently be preferable to a foolish and stupid soul though that were composed of the most precious elements." The material in which the structure of reason is instantiated, is irrelevant to psychological theory, though it may be of interest to a neurologist, a physiologist or a physicist.

Another basis for the repugnance felt (and still felt in some quarters) was the supposed ill-treatment of animals that follows from the thesis. Thus Fontaine (1738): "They said that the animals were clocks; that the cries they emitted when struck, were only the noise of a little spring which had been touched, but that the whole body was without feeling."

A less emotional (but no less pertinent) criticism was that of Fontanelle (1764): "Mais mettez une machine de chien et une de chienne l'une aupres de l'autre, il en pourra resulter une troisième petite machine: au lieu que deux montres seront l'une aupres de l'autre toute leur vie sans faire jamais une troisième montre."

To go further than Fontanelle is beyond the scope of this chapter, but it is high time that a more thorough and systematic study of the clock analogy was undertaken. As Frederick the Great said, La Mettrie "tried to explain ... the thin texture of understanding, and he found only mechanism where others had supposed an essence superior to matter." To what extent was the rapid rate of advance in the clockmaker's art responsible for the fact that La Mettrie completed the philosophical foundations of the Cartesian revolution?

# CHAPTER FOUR

## ARE MODELS ABERRATIONS OF MINDS TOO FEEBLE TO THINK ABOUT ABSTRACTIONS?

Calculating machines and adders.

We are told (Ghiselin, 1952) (perhaps somewhat apocryphally) that, after a hard but unsuccessful period slaving over a hot desk trying to solve the chemical problem of the benzene molecule, Kékulé had a vivid daydream in which a snake swallowed its own tail. We are not told why the snake adopted such an unnatural posture; (perhaps it was like the Gryphon's whiting which "would go with the lobsters to the dance. So they got thrown out to sea. So they had to fall a long way. So they got their tails fast in their mouths. So they couldn't get them out again. That's all." (Carroll, 1865)). Luckily, we do not have to decide how the snake got its tail into its mouth ... only what the result was. Kékulé came to his senses, with the image reverberating in his mind's eye, to find that he had solved the problem - the structure of the benzene molecule was a ring rather than a chain.

The present point is that one would search in vain through chemistry textbooks for illustrations of the snake, and lecturers seem to be quite capable of explaining the structure of the benzene molecule to aspiring chemists without referring to suicidal anacondas or masochistic vipers.

Similarly, the doodles (if any) which Einstein drew in the margin as he was formulating the general theory of relativity may in some sense have helped him to solve the problems, yet neither do they figure in textbooks of theoretical physics.

According to some (e.g. Duhem, Braithwaite etc.), models are no more essential to theory than doodles or daydreams ... at the best, all three are mere psychological aids, private and personal to their individual creators and playing no part in the intersubjective language of science; at their worst (i.e. usually) models are superficial and distracting.

What these critics strive for, instead of models, are purely formal deductive systems only some of whose consequences need be interpreted into observables and empirically tested ... what they see as theories (which are abstract, logical and systematic) as opposed to models (which are concrete, visual, imaginative and incoherent).

Duhem's classic: The Aim and Structure of Physical Theory, published in 1914, was, as might be expected from the title, directed primarily at physicists, but one might as well put the "Physical" of the title in parentheses and take it as an attack on model building as a form of explanation in any science. Chapter IV in particular, which deals with "Abstract Theories and Mechanical Models", might have specifically been written about one of Vaucanson's or von Kempelen's automata ... or indeed even of a more modern computer model such as Winograd's.

In this chapter, we shall discuss the relation between theories and models, starting with a discussion of Duhem's ideas, and see that we need to split up the somewhat monolithic concept of 'model' into two separate but related parts.

Mad dogs and Englishmen.

According to Duhem, (physical) theory should be the result of the application of two processes: abstraction and generalisation (achieving economy both in substituting a few hypotheses for a vast set of laws and, previously, substituting a law for a multitude of facts). However, he realised that not all "vigorously developed minds" are endowed with high powers of abstraction but some are rather "endowed with a powerful faculty of imagination." As Pascal (1669) put it, there are: "two kinds of minds: one kind, able to penetrate quickly and profoundly the consequence of principles, we call the exact mind; the other, able to comprehend a great number of principles without confusing them, we call the geometrical mind."

Variations in terminology are common. Duhem discusses Napoleon as an example of a broad but weak intellect ... "the slightest and most fleeting detail ... did not escape Napoleon's scrutiny" but (as Taine said) "a general principle displeased him as a bad joke."

However, according to Duhem, Napoleon was something of an exception - being a Frenchman with an ample mind ... more typically an English failing: "There is one people in whom ... ampleness of mind is endemic ... the English" (who have an) "extraordinary facility for imagining very complicated collections of concrete facts and ... an extreme difficulty in conceiving abstract notions and formulating general principles." Duhem cites Dickens' novels as "nightmarish series of concrete things" and criticises Bacon's "The True Philosopher" as showing no attempt "to construct a clear and well-ordered system of truths logically deduced from warranted principles", but rather, he goes on scathingly, "its object is quite practical not to say industrial."

Duhem clearly intended the shop floor references perjoratively. Why, is not so obvious - as a theoretical physicist we might have expected him to be fascinated by the concept of mass production - but he continues in a similar withering vein in discussing a book by Oliver Lodge (1890).

First he mentions what he expects in a theory: "This whole theory (should constitute) a group of abstract ideas and general propositions, formulated in the clear and precise language of geometry and algebra, and connected with one another by the rules of logic."

Secondly he mentions what he finds: "Here is a book ... intended to expound the modern theories ... In it one finds nothing but strings which move round pulleys, which roll round drums, which go through pearl beads ... toothed wheels which are geared to one another and engage hooks. We thought we were entering the tranquil and neatly ordered abode of reason, but we find ourselves in a factory."

These strings, pulleys and gears were of course, collectively, Lodge's model. Another modelling physicist, William Thomson (1884) (later Lord Kelvin), said, as explicitly as we could wish: "It seems to me that the test of 'Do we or do we not understand a particular subject (in physics)?' is 'Can we make a mechanical model of it?' ... I never satisfy myself until I can make a mechanical model of a thing. If I can make a mechanical model, I understand it. As long as I cannot make a mechanical model all the way through I cannot understand." To this Duhem's response is predictable enough ... these models should not be taken as explanations (which would be logico-deductive systems). Rather they are ("not combinations intended to be conceived by reason, but) mechanical contrivances intended to be seen by the imagination."

That some researchers do, as Duhem said, have a "need to imagine concrete, material visible and tangible things" is undeniable. Oliver Lodge did express a desire to "form a mental representation of the phenomena which are really happening" and even Deutsch (1960) admits that expressing a theory in terms of a model by giving it an identification which is already familiar makes it easier to think about than a completely abstract system - and that, in this respect, it is a mere psychological aid, which neither adds to nor detracts from the explanatory value of the system itself. For Duhem, though, modelling doubles the problems - we not only need to grasp the operation of the model itself - often fairly complex, as in Kelvin's models of vortex atoms or Winograd's natural language understanding system - but we also have to grasp the analogy between the properties of the model and the propositions of the theory.

However, whilst Duhem did violently oppose the use of models as explanation, his opposition to models (or at least analogies) as sources of inspiration for discovery has perhaps been over-emphasized in the literature - probably because Mary Hesse's Models and Analogies in Science (1966) polarised the positions of Duhem and Campbell for the purposes of exposition.

It is true that Duhem was sceptical of claims of theoretical advance due to model use. He believed that models were often parasitic on theory - which withered away after the birth of the model, but without which the model would never have been born. Yet on the other hand, he was moved to write that "the search for analogies between two distinct categories of phenomena has perhaps been the surest and most fruitful method of all the procedures put in play in the construction of (physical) theories." Note however that for Duhem the process of analogy was an abstract one, far removed from concrete, mechanical apparatus or models: "Analogies consist in bringing together two abstract systems; either one of them already known serves to help us guess the form of the other not yet known, or both being formulated, they clarify each other."

What exactly is at issue here? Duhem objects to models but commends analogies. It is not merely that the models he was considering were mechanical. Indeed, he wrote that what he objected to in the English school was not that it had tried to reduce matter to mechanism - after all his beloved Descartes (with his "strong and restricted mind") who believed animals were machines, La Mettrie, who believed men were machines, and Huygens, whom Duhem respected, had done just that - to name but three. No, it was not the reduction itself of matter to mechanism to which he objected, but "the particular form its attempts have taken" (in the English school). The essence of the respectable mechanists was a desire for an explanation which was economical and systematic (highly abstract concerns); but that of the English was "a lapse in the faculty of abstracting, that is, to a victory of imagination" ... a yielding to the inner "need to imagine concrete material, visible and tangible things." (Though Descartes did build automata to test his mechanical theories and can Duhem have heard of Vaucanson?).

One result of the tendency to be concerned with the tangible present, which Duhem criticised in the English school, is still as relevant today ... if not more so. The English (physical) theorist "does not aim to deduce his model from a philosophical system, nor even to put it into accord with such a system." The result is that a model may be completely ad hoc - it may not be theoretically motivated at all, its only justification being that it 'works'.

Guide lines on model-adequacy may be derived from Chomsky's (1965) suggestions regarding adequacy of a theory of language. A model is observationally adequate if it is indistinguishable in the relevant respects from the modelled, i.e. it passes the Turing test, provides the same input - to - output conversion etc; it is descriptively adequate if it does the same thing in the same way, i.e. the processes upon which the output of the model are contingent are of the same type as those upon which the output of the modelled are contingent - where "same" means "receiving a common description by a theoretically relevant metalanguage"; it is explanatorily adequate if it is suggested/supported (e.g. if it is unified) by some theoretical concerns etc.

These conditions on adequacy render more precise what was said above regarding weak and strong modelling ... observational adequacy is equivalent to the weak requirement and descriptive adequacy to the strong one. These constraints on modelling have however long been appreciated. It is of considerable interest, and consistent with the general theme of this thesis, that these sorts of considerations are at once an integral part of the concrete process of building models and of the theoretical process of the philosophising about explanation. To take a couple of examples from the use of clocks as intimate models: Descartes (1647) was well aware that, although the theory of man he was offering was observationally adequate, he could not guarantee its descriptive adequacy.

"Although I may have imagined causes capable of producing effects similar to those we see, we should not conclude for that reason that those we see are produced by those causes; for just as an industrious watch maker may make two watches which keep time equally well and without any difference in their external appearance, yet without any similarity in the composition of their wheels, so it is certain that God works in an infinity of diverse ways (each of which enables Him to make every thing appear in the world as it does, without making it possible for the human mind to know what of all these ways He has decided to use.)"

Cotes (1729), too, was aware of the strong/weak distinction, but insisted that descriptive adequacy alone should be our goal:

"The business of true philosophy is to derive the natures of things from courses truly existent; and to enquire after those laws in which the Great Creator actually chose to found this most beautiful Frame of the World; not those by which he might have done the same, had he so pleased. It is reasonable enough to suppose that from several causes, somewhat differing from each other, the same effects may arise; but the true cause will be that, from which it truly and actually does arise; the others have no place in true philosophy. The same motions of the hour-hand in a clock may be occasioned either by a weight hung, or a spring shut up within. But if a certain clock should be really moved with a weight; we should laugh at a man that should suppose it moved with a spring, and from that principle suddenly taken up without further examination should go about to explain the motion of the index; for certainly the way he ought to have taken should have been, actually to look into the inward parts of the machine, that he might find the true principle of the proposed motions."

Differing enterprises have differing requirements upon them as regards satisfaction of the three levels of adequacy; observational; descriptive and explanatory.

Cognitive simulation should satisfy all three, whereas artificial intelligence need only be observationally and explanatorily adequate - it is no part oftheir brief to ensure that their programs solve problems in the same way as people, although, through the use of protocols and because the programmers are human, this may incidentally occur. To be of theoretical interest however, both C.S. and A.I. should be theoretically motivated: ad hoc patching of a program simply to get it to do what we want may be useful but yields no insight. As an example, Michael Arbib (1969) has written an interesting program simulating the self-reproducing properties of a three segment worm, which, when 'chopped' into pieces, regenerates itself. However, the simulated worm regrows itself in unnatural combinations, e.g. head-tail-head. Arbib solves the problem by adding a set of context sensitive rules to filter out unwanted combinations. Yet the only motivation for these rules is the desired end-state ... they have no natural justification in terms of the model itself

Similar allegations of arbibtrariness have been levelled against Winograd's program for understanding natural language by Drescher and Hornstein (1976).

Thus it is not the mechanical side of machines to which Duhem objects but their concreteness and the way they seduce the mind from more abstract issues such as theoretical motivation. Duhem openly endorses the use of analogies as discovery procedures. Yet mechanical models have an analogical side - it is just that in which we are interested. What we need is a way of teasing apart the concrete physical side of mechanical models from their abstract analogical side.

Before we can satisfy this need, however, we shall have to look a little more closely at the relationship between models and what they model.

Two ways to approach this are:

(i) From the point of view of the system:

(ii) From the point of view of the primitive.

(i).

Of two structured systems, A and B;

A (e.g. a chess machine) strongly models B (e.g. a brain) insofar as (a) A and B both perform some common relevant function (e.g. chess-playing);

(b) A's and B's respective functioning are a consequence of A's and B's respective structure and there is an interesting (information processing relevant) level of description such that they both satisfy a single (probably complex) referring expression;

(c) A's structure is explanatorily adequate.

(ii).

Of two primitives, a and b;

a (e.g. a transistor) strongly models b (e.g. a neurone) within a structured system A;

insofar as (a) a has a function within A (a computer);

(b) b has a function within B (a brain);

(c) A and B have some relevant function in common (e.g. chess-playing competence);

(d) the structuro-functional relationships of a and b to other elements in their respective systems satisfy some interesting common level of description;

(e) A's structure is explanatorily adequate.

Such a schematic characterisation of models has the advantage of demonstrating why such disparate phenomena as Vaucanson's duck and La Mettrie's clocks both count as models; of exhibiting what it is they have in common.

But there is a problem - the necessity of the use of the term "relevant". As has been stressed ad nauseam, everything cannot be true of the model that is true of the modelled ... or the model would be the modelled, indeed, "only by being unfaithful in some respect can a model represent its original" (Black, 1962). Model and modelled will have some commonly shared properties (relevant ones) and some properties which distinguish them (irrelevant ones). The Turing test was designed to minimise distraction by 'irrelevant' aspects of the model. The motion and behaviour-on-impact of billiard balls were relevant properties in their modelling of atoms - but not their colour or being made of pottery or ivory.

However, whilst ontologically the division of properties into those equally true of the model as of the modelled and those true of the model but not true of the modelled (and vice versa) may be exhaustive, epistemologically it is not ... there are also those properties of which we do not know whether or not they also belong to the modelled.

Let us call this latter set of properties the neutral component of the model. Let us also call that which the model and modelled have in common the positive component and that which distinguishes them the negative component.

Now let us clarify all this by means of an example. Take Craik's (1943) use of the railway system as an intimate model of the nervous system. Here the positive component is the modelling of neural pathways (neural tubes?) by the railway lines and of nervous impulses by locomotives; the negative component consists of facts like e.g. trains, but not nerve impulses, were powered by steam, made of metal, travelled on wheels etc. The neutral component (that part of the model of which we are unsure whether it is part of the negative or part of the positive component) is the growth area of the model ... that which gives rational grounds for new hypotheses and experimentation (i.e. which licenses hypotheses as opposed to mere guesses). In the Craik example, the neutral component is the way the railway system is organised ... the fact that (pre-Beeching at least) there was not only one line to each destination, but many alternative routes. Once grasped, this neutral component can be transferred to the modelled from the model and it generates predictions e.g. that just as bombing one of many lines to/from a destination (Craik was writing during World War II) would only minimally disrupt communication to/from that place, interrupting a neural pathway - and thus derailing the nervous impulses - would only minimally disrupt the function of the area to/from which the impulse was travelling (a sort of naturally occurring response buffer?) ... other impulses would reroute and eventually arrive, though possibly late (unless they put on steam in the meantime). The neutral component yields almost indefinitely many testable predictions - we can block lines and time trains over set routes, try and isolate an area completely by blocking all its lines etc., etc.

Note though, that this sort of analysis does not seem to apply in an equally natural manner to all of the models with which we have concerned ourselves. Take, for example, Vaucanson's flute player - the positive and negative components are easy to identify but where is the neutral one? It is true that expression of a theory in rubber, leather and metal forces you to be totally explicit about your intuitions and shows the difference in degree of difficulty between glossing over theoretical inadequacies on paper and making a machine work by wishful thinking. The working out of concepts in steel (or transistors) reveals connections and interactions you had made, or failed to make, inadvertantly. It may even be that there are unexpected bonuses (implications for the modelled), which stem from unlooked-for aspects of the assembled model, but, compared to the Craik model, the role of the vestigial neutral component is minor. The major concern of Vaucanson was to present a theory of the flute player in terms of leather and rubber i.e. he was interested in exhibiting in as explicit a mode as possible just that which the model and the modelled <u>shared</u> - the positive component. His aim was precisely to exclude from the model any properties not shared by the modelled. He purposely created a machine to possess precisely those properties which we have expressed as the positive component. Craik, on the other hand, seized on a ready made artefact - the railway system - knowing full well that, in some respects, it was quite unclear as to whether or not properties of the railway system were properties of the nervous system.

We now have the concepts at our disposal to meet an apparent criticism of the account we gave in Chapter I as to the nature of a model. Recall that we suggested there that models are a species of representation which resemble what they represent.

It may be objected that models tend to resemble other models more than they resemble that of which they are models (for example, two IBM computers, modelling completely different psychological processes, may, in some sense, seem to be more similar to each other, than either is to what it respectively represents.)

Now, though, we can see that, as A can only model B by leaving something out, A will have both a negative and a positive component. However, what we take as the negative and positive components depends on our interests. The model may be like other models in some respect(s) e.g. in its colour, or that of which it is made, and it may be like that of which it is a model in other respects e.g. in its information handling capabilities. To assert that any one respect is the positive component (wherein lies the similarity) is but a shortsighted and unjustified dogmatism.

Two sorts of model.

> "In everyday practical thought, physical analogy metaphors play a large role, presumably because one gets a large payoff for a model of apparently small complexity ... It would be hard to give up such metaphors, even though they probably interfere with our further development, just because of this apparent high value-to-cost ratio. We cannot expect to get much more by extending the mechanical analogies, because they are so inflexible in character. Mental processes resemble more the kinds of processes found in computer programs: arbitrary symbol associations ... etc. In short, we can expect the simpler useful mechanical analogies to survive, but it seems doubtful that they can grow to bring us usable ideas for the parallel unification of the internal modelling mechanism."
>
> (Minsky, 1968).

Let us now distinguish, and name, the two types of model we have illustrated in the cases of Vaucanson and La Mettrie (and Craik) respectively:

A FORMAL model is one which has been intentionally devised or constructed to match, in some relevant respect, the structure and/or function of some system the modeller wishes to understand.

e.g. those of Descartes (in his automata), von Kempelen, Longuet-Higgins, Vaucanson, Winograd.

An INTIMATE model is a pre-existing entity or phenomenon which is seen to be similar, in a conceptually interesting way, to something the modeller wishes to understand.

e.g. those of Craik, Descartes (in his use of the clock), Pearson, Plato, Romanes.

These two categories are not mutually exclusive; for example, a computer program (potentially) constitutes one of the most precise and explicit formal models of cognitive psychology available, yet, merely to say that man is a computer, tout court, is to indulge in the use of an intimate model.

Both sorts of model contribute to research - but with differing emphases. Formal models concentrate on rendering explicit the positive component, the neutral component purposely being kept to a minimum. Intimate models seldom fit so well or so extensively (not being specifically designed for the job), but exploit a fertile neutral component as a source of testable predictions beyond the positive component.

"A fact m, known to be true of A, is more likely to be true of B, if B agrees with A in some of its properties (even though no connection is known to exist between m and those properties) than if no resemblance at all could be traced between B and any other thing known to possess the attribute m." (Mill, 1891). In the limiting case, where B resembles A in all its properties, its possession of m would of course be certain, and it does (intuitively) seem that the more B can be shown to resemble A, the more likely is B to possess m too. It is however, far from trivial to show that this intuition is valid, although there is limited inductive support for the notion. Take the model ABD, the modelled BC, the evidence that all ABs are D and the problem of whether BC is D. The evidence is compatible with each of four hypotheses: (i) everything which is A is D and everything which is B is D; (ii) everything which is A and B is D; (iii) everything which is A is D; (iv) everything which is B is D. Of these (i) and (iv) support BCs being D; (iii) is irrelevant and only (ii) is contrary. On balance this appears to support, if only weakly, the belief that BC is D. (Accounts in terms of probability, falsify ability and simplicity are no more successful.)

Certainly the neutral component is not a stopping but a starting place. Mill: "the competent inquirer into nature (will) consider the analogy as a mere guide post, pointing out the direction in which more rigorous investigations should be prosecuted."

We can now reap the benefit from the ambiguity of the term "intimate" - these models are intimate (homely) but they are also intimate (point a direction) by virtue of their neutral component.

As I have written elsewhere, there is no entity or phenomenon too low or ludicrous to function usefully as a psychological model. Mill supports this notion (and gives a reason why): "any suspicion, however slight, that sets an ingenious person at work to contrive an experiment, or affords a reason for trying one experiment rather than another, may be of the greatest benefit to science." Indeed, Mill is quite adamant in his insistence that 'inquirers' should not "restrict themselves arbitrarily to the particular hypothesis which is most accredited at the time, instead of looking out for every class of phenomena between the laws of which and those of the given phenomenon any analogy exists, and trying all such experiments as may tend to the discovery of ulterior analogies pointing in the same direction."

By this stage, the reader should have a fairly good idea of what I am trying to do in this thesis, of what I mean herein by the term "model", both <u>intimate</u> and <u>formal</u>, of how important I think models are and, finally, some idea of how modelling is actually achieved.

We now move on to consider more models in more detail. In the next four chapters (i.e. 5 - 8) we discuss some formal models, and in the following three chapters (i.e. 9 - 11), we consider some intimate ones.

# CHAPTER FIVE

## ANCIENT AUTOMATA ... AND ALL THAT JAZZ!

"the Greek civilisation ... favoured intuition, insight and the intellectual processes, but not the extraction of secrets from nature by mechanical contrivance and experimental technique. This was not to come until almost 2,000 years later."

E. Boring (1950)

We are going to look presently, in some detail, at some 18th and 17th century automata and androids (completely mechanical figures which simulated a living human or animal, operating with apparently responsive action: Bedini, 1964.). and the motives of their builders. Of course, this work did not spring into being fully-fledged and uniquely original, but was part of a long and venerable tradition. But in virtue of which aspects did it belong to a tradition and (indeed) just which tradition?

This section is short, partly because information about automata before the 17th century is not always as totally reliable as one might wish and partly because it is not as obviously of crucial import to the central arguments of this thesis, which could anyway be maintained purely on the basis of the later simulacra. However, some brief survey of more early automata and androids is in order for several reasons: firstly, there is good reason to think that these early efforts, and the records kept of them, did actually materially affect the work of the later European simulators (e.g. the Alexandrians via Islam to France,); secondly, we gain further insight into the (multiple) attractions of simulation; thirdly, there is some little evidence that these early precursors were, whatever else they were doing, indeed trying to understand by modelling; and fourthly, the material is intrinsically fascinating.

It has been claimed that it is a "deep rooted urge of man to simulate the world about him through the graphic and plastic arts" (de Solla Price, 1964); that there is a dualism deep in man's very nature, which leads him both to try "to uncover the mystery of man's making" (and perhaps even try to usurp the powers of the gods) <u>and</u> to "exercise to the full all that his ingenuity has to offer" and that, moreover, this dualism "is already apparent in his early twofold role as cave artist and toolmaker" (Cohen, 1970). The reason that, although there seems to have been a continuous and strong tradition leading man to simulate living animals and even man himself, there is little (and especially little concrete) evidence of early simulacra, was the limitation due to lack of technical skill (de Solla Price 1964). However, limited though that skill might seem as compared to that of the engineering genius Vaucanson and his contemporaries, the skills we are about to witness are surely remarkable.

Let us now briefly trace the early history of the automaton, which is, incidentally, a particularly interesting example of cultural transmission (Battisti, 1960).

There are two proven routes for the passage of interest in automata from Alexandria to Europe:-

(1) Alexandria ⟶ Islamic World ⟶ Sicily ⟶ court of Frederick II (1194-1250) ⟶ France.

(2) Alexandria ⟶ Arabic table games ⟶ 16th century Germany.

and two possible (but unproven) ones:-

(3) Alexandria ⟶ India.

(4) Alexandria ⟶ China.

We shall take a (slightly) closer look at these major centres of the automaton industry, but first let us take a glimpse at some (alleged) automata from Ancient Greece, contemporaneous with, or even earlier than, the Alexandrians.

It is almost impossible to disentangle legend from (purported) fact in early (and for that matter modern) Greece, but one thing is certain ... there was a profound interest in the notion of androids; after all the very word 'automaton' is Greek in origin (αυτόματος).

For Homer, an automaton was anything which operated independently (Iliad). In the Iliad Homer mentions banquetting hall tripods, built by Vulcan, which ferried to and from the tables with wine and delicacies, whilst in the Odyssey (vii, 91, ff) we hear of Talos, the Cretan hero with the bronze body and of the golden and silver dogs "endowed with a certain intelligence" in the palace of Alcinous. Aristotle spoke of a wooden Venus, who moved about as a consequence of quicksilver being poured into her interior (Brewster, 1832) and Daedalus, also according to Brewster, "enjoys the reputation of having constructed machines that imitated the motions of the human body." Plato and Aristotle said of Daedalus' "spontaneously moving statues" that it was necessary to tie them to prevent them from running away and Callistratus, Demosthenes' tutor, said the statues were mechanical. One of the most bizarre androids of all was perhaps the mechanical snail which, according to Polybius (XII, 13, II) appeared in the triumphal procession of Demetrius Phalerus in 307 B.C.

In discussing the Alexandrians, who were by no means averse to automata, we are on a much surer footing. The Alexandria School is centred around the work of Ctesibius (c250 B.C.) who constructed water-clocks and automata; Philo (200 B.C.), who wrote a treatise on hydraulic and pneumatic mechanisms and, above all, Hero (285-222 B.C.), who wrote many treatises on hydraulic and pneumatic automata, such as libation vessels, singing birds, automatic wine and water mixers, temple doors which opened automatically on the lighting of the altar fire and a slot machine holy water dispenser (a "Sacrificial Vessel which flows only when Money is introduced") which accepted a 5 drachmae piece.

Hero was also a builder of mechanical theatres. In his Treatise, he describes one and explains that he purposely built it in miniature to preclude the suggestion that a man was hidden inside! (The interest of this remark becomes even more apparent when we consider von Kempelen's chess automaton).

Hero's work is of the highest importance and Battisti is almost certainly right when he says: "Alexandrian treatises, especially the writing of Hero, were determinant in influencing the development of automata towards technical and decorative virtuosity." Naturally, the Greek word for automaton appears in the title of one of Hero's treatises: "περι αὐτοματοποιητικης ", which appeared in French in the 16th and other western languages in the 17th centuries. The Pneumatica was translated into French in 1598 and German 1688. I give here three examples, all involving androids, from the latter work to give an inkling of the complexity achieved. (Woodcroft, 1851).

Number 40.

> "On a pedestal is placed a small tree round which a serpent or dragon is coiled; a figure of Hercules stands near shooting from a bow, and an apple lies upon the pedestal; if anyone raises, with the hand, the apple a little from the pedestal, the Hercules shall discharge his arrow at the serpent and the serpent hiss."

Number 49.

> "A trumpet, in the Hands of an Automaton, sounded by compressed air."

Number 78.

> "An Automaton, the head of which continues attached to the body, after a knife has entered the neck at one side, passed completely through it, and out at the other; the animal will drink immediately after the operation."

All of these are, of course, supplemented in the treatise itself with detailed construction information.

As we have seen, historians of ideas believe that the tradition in automata, with their complicated gearing mechanisms, shifted from

Alexandria to Islam. In fact, Hero's Mechanics were translated into Arabic by Qusta b Luqua about 864 A.D. and Philo's Pneumatics were also translated early into Arabic, although the exact date is not known. Al Kindi, an Islamic philosopher of 970 A.D., refers both to Hero (as knowledgable of geography, pneumatics and time measurement) and Philo (as skilled in the construction of Ingenious Mechanical Devices). Moreover, examination of the devices of the Banu Musa shows they were directly derived from Hero's and Philo's work, and al-Jazzari, the most important automaton maker of Islam in this period, actually refers to the Banu Musa as one of his sources.

'The Banu Musa' (813-833 A.D.) has been given as a name to a work by three brothers, only one complete copy of which exists for certain, and that in the Vatican. It describes 100 devices: fountains, self-trimming lamps, an automatic musical instrument, a mechanical grab for excavating stream beds, a vast number of drink dispensing trick vessels, and, interestingly, a gas mask for approaching polluted wells. These devices do employ an advance in valve technology (cone instead of crude clack or plate valves) but, as Hill (1974) says, this is the main difference between the Banu Musa devices and those of Philo and Hero apart from "the greater complexity of the former."

We pass over Mafatih al-Ulum (975-991 A.D.) and a work of Ridwan, to discuss al-Jazzari's Book of Knowledge of Ingenious Mechanical Devices, a treatise which is "the most elaborate of its kind and may be considered the climax of this line of Muslim achievement." (Sarton, 1931).

Al-Jazzari, originally from Mesopotamia, had settled in Diyar Bakr (now northern Iraq) and constructed automata and mechanical curiosities for the local prince. In 1206 he wrote down for posterity descriptions of 50 of the devices. Many of these were hydraulic and among them there was an astonishing range of living beings represented: birds, cows, dancers, ducks, elephants, falcons, fish, horses, lions, kings, monks, monkeys, peacocks, sailors, scribes, slaves and swordsmen; made from materials as various as beaten copper and papier mache, wood and brass. There were clepsydras in which as many as 8 or 9 different jacks were activated by a single water flow, automaton whistles and flutes, robots which dispensed water and towels for washing, automatic wine and water dispensers and so on almost indefinitely.

It is worth describing two such automata in more detail:

The Water Clock of the Drummers.

> "It is a frieze like a ledge projecting from the face of the wall about 4ft in a straight horizontal line. Along it's edge there are 12 battlements and at the end of the ledge stands a man. His right hand is outstretched and his index finger points towards the battlements; when he moves behind the battlements his finger almost touches their points. Above the frieze and parallel to it are 12 glass roundels in a straight line (set) in holes (cut) through to the inside of the house. Below the centre of the frieze is a mihrab with falcon in it ... with a vase in front of it on a projecting bracket, with a cavity (in the wall behind it).
>
> In the floor of the chamber is a platform occupying all the foreground, raised about the height of one man above the ground. On this platform are seven men: on the right, two blowing trumpets, on the left two playing cymbals - the rest are drummers ... at daybreak the man is at the end of the frieze and moves smoothly until he is behind the first of the 12 battlements, whereupon the falcon leans forward and casts a ball from its beak onto the cymbal in the vase, and the musicians play. This happens every hour."

A Boon Companion - a man who drinks the king's leavings.

This little fellow, who looked like a five year old boy, was employed at drinking parties. The steward having poured the leavings into the boon companion's goblet, the automaton lifted the goblet to his lips, lowered the empty cup, nodding and raised a water lily in his other hand. However, as its party piece, the boon companion evacuated over the lap of an unfortunate guest, whom the king had persuaded to hold. In fact, ingeniously the wine went from the goblet not between the android's lips, but through the stem of the goblet, down the arm, into a reservoir. The raising lily signalled to the initiated that the automaton was near capacity and the evacuation was accomplished by siphon action.

Of course, al-Jazzari made technical advances in constructing these machines. In addition to "elaborate systems of trip-levers, pullies, tipping buckets, floats,runners, trapdoors and ballraces", we witness the first appearance of conical valves (previously thought to appear in Leonardo's drawings) and segmental gears (previously thought to appear in de Dondi's astronomical clock), and we have the first unequivocal description of metal casting in closed mould boxes with green sand, not met in the west until the 15th century. He was also the first to describe (and perhaps invented) the suction pump and he introduced the notion of a leaking float, which sank in a given period.

These automata raise interesting questions in many directions. One would have expected them never to have existed at all for two reasons: firstly, there was a traditional Moslem predilection for two dimensional treatment (Ettinghausen, 1960) and secondly, there was a more general Moslem horror of images, which commonly precluded statues (Keller, 1975). Still, interestingly from our point of view, historians do suggest, I use no stronger term, that simulation was an underlying urge. Thus Keller: "It is surprising to find figures that move and really do try to imitate human actions" and Ettinghausen: "in these automata, Islamic art attained a greater realism than in any other medium."

Two other possible directions of influence from Alexandria were mentioned earlier ... to India and China, respectively.

Philostratus informs us that Apollonius saw self-moving banquetting tripods amongst the sages of India (Brewster), but the first concrete evidence comes in the 11th century with the treatise by Bhoja (Samaranganasutradhara), in which is described a wooden flying machine in the form of a bird and powered by steam from a shell of boiling mercury (Vallauri, 1960). Treatises and stories agree in drawing a parallel between the machine of the universe, composed of the five elements ether, air, fire, water and earth, and the yantra, manufactured by men, in which the same elements reappear in a new order. Knowledge of the science of automata was yantrasastra.

In China, there are many records of automata. There is mention of a mechanical orchestra being in the treasury of Ch'in Shih Huang Ti in 206 B.C., whilst (more psychologically?), in the 2nd century B.C., Ch'u Chi built a rat's market in which automata closed doors as soon as the rats tried to escape (the first Skinner box?). Later on, in the 4th century A.D. Wang Chia speaks of a mechanical man made of jade who could walk and jump by means of a well concealed mechanism and in a 6th century manuscript (Hsi ching tsa chi) are mentioned "12 men in bronze 3 feet high ... seemed like living men ... the figures performed like real musicians." (Needham, 1960).

In Europe itself, automata have figured in legend and record, and that hazy ground that lies between, for nearly 1,000 years. Pope Sylvester II (999-1003) is reputed to have built a speaking figure and Albertus Magnus (1193-1280) an iron man. Villard de Honnecourt (active

1230-35) tells of a moving angel which pointed to the sun with its finger and Vergil the magician built a mechanical fly to frighten other flies out of the butchers' shops of Naples. Two drawings by the Tuscan architect Giovanni Fontana of 1420 survive, showing devils with horns, eyes, tongue, arms, fingers and wings all moved by strings; and, when the Emperor Maximilian arrived at Nuremberg on 7th June, 1470, an artificial eagle, constructed by Johannes Muller, or Regiomontanus, allegedly flew out to meet him, then returned to perch on the town gate. Muller was also responsible for "an iron fly, which was put in motion by wheelwork, and which flew about and leapt upon the table" (Brewster.)

Gianello Torriano de Cremona (c1500-85), who was entertainer to Emperor Charles V, is reputed to have built fighting armies, flying birds, a ballerina a palm high, who danced and played the tambourine, and a working cornmill which could be concealed in a glove. In 1509, Leonardo da Vinci made, in honour of Louis XII, an automaton lion which crossed the room to the throne and opened its breast with a paw to reveal the lilies of the royal house of France. In 1588, Agostino Ramelli's Diverse et Artificiose Machine was published, followed a year later by Hans Schottheim's automaton lobsters and ten years later Reidel's mechanical spider, now both in Dresden. Perhaps the most complicated set of androids was that of Friedrich Hentsch (1660) who constructed two armies of more than 100 men: horsemen, foot-soldiers and musketeers who fought and performed complicated manoeuvres to the sound of firearms.

Hentsch's prototypical 'action men' brings this brief survey up to the heyday of Vaucanson and his contemporaries. But to what interesting theoretical questions is this amassed data relevant?

Were these fabricata indeed simulacra? This is a difficult question to answer. Some historians think so. As Battisti puts it: "their essential aesthetic quality lies in their imitation and elaboration of movement." However there are at least three tendencies evident in the tradition of android building. One is indeed "man's dauntless ambition to exercise to the full all that his ingenuity has to offer" (Cohen) and we can think of "old technical experiments" (Battisti) in this way too. Separate from (but parallel to) this tendency was the desire to understand the mystery of life itself, with perhaps a view to usurping the power of the gods; this tendency is evidenced in the stories of Prometheus (remember that Voltaire called Vaucanson a Prometheus), Paracelcus and Faust. The third (related) tendency was towards magic and the black arts - an accusation often levelled against 'respectable' scientists. In ancient China the making of masks and puppets was believed to be related to the cult of the dead (A. Bulling); a chronicle of Nuremburg of 1398 states that "Turning mechanisms which perform strange gestures come directly from the devil." An interesting later example of this sort of interpretation is provided by Thomas Nashe in "The Unfortunate Traveller":

> "These birdes by the mathematicall experimentes of long silver pipes secretlye inwinded in the intrailes of the bough whereon they sate, and undiscernablie convaid under their bellies into their small throats sloaping, they whistled and freely carold their natural fieldnotes ... But so closely were all these organising implements obscured ... that everye man there present renounst conjectures of art, and sayd it was done by inchantment."

Modern model-building psychologists and practitioners of A.I. are seldom accused of black magic or prometheanism, although critics often seem to suggest that there is something rather profane about the enterprise. The perjorative use of the term 'hubris' as applied by critics like Dreyfus or Weizenbaum has intriguing classical connotations - tacit recognition perhaps of the long tradition to which I have made, in this chapter, a mere gesture.

# CHAPTER SIX

## OPINIONS ON PINIONS - AN EIGHTEENTH CENTURY TURING GAME.

> "that an Automaton can be made to move the chessmen properly, as a sagacious Player, in consequence of the preceding move of a stranger ... is ... utterly impossible".
>
> Philip Thicknesse (1784).

"How much of the business of thinking could a machine possibly be made to perform?" (Peirce, 1870). Thus stated, the question displays two depressingly common properties - it is apparently central to many important issues in cognitive psychology yet it is hopelessly vague.

In 1950, in an influential paper, Turing sought to tighten up such questions by introducing the concept of an "imitation game". The principle of such a game is simple; an interrogator is alone in a room with teleprinter connections to two other rooms; in one room is a person, in the other is a machine. By means of questions and general conversation, the interrogator has to decide which responder is the machine and which the person. Both machine and person may, of course lie, although it is probably in the latter's interest to tell the truth. A machine which (in this situation) is indistinguishable from a human being is said to pass the Turing test, as it is now known. More precisely: "A machine successfully simulates the behaviour of an organism when trained judges are unable to discriminate the behaviour of the machine from the behaviour of the organism in relevant test situations." (Fodor, 1968). Programs need not be staggeringly complex to pass this test; Weizenbaum's "Eliza-Doctor" program, which worked by reflecting the interrogator's statements back at him, passed the test with such flying colours that, to Weizenbaum's horror, suggestions were made, e.g., by Colby et al (1966) that the program would soon be ready for clinical use. Weizenbaum (1976) admits that Eliza had a "relatively simple computer program". Nevertheless, the Turing test is an extremely powerful lower constraint on programs, at least.

The principles of the Turing game are however by no means new. From 1769 to 1834, the "whole of Europe" (and much of America) was "astonished and delighted" (Brewster, 1832) by what amounted to a grand (commercial) Turing game - von Kempelen's (later Maelzel's) Automaton Chess-Player.

In 1769, Wolfgang von Kempelen was summoned to the court of the Empress Maria Theresa of Austria to explain various tricks of magnetism being performed by the itinerant Frenchman Pelletier. Von Kempelen, a court counsellor, was already famous for his mechanical ingenuity and would later be responsible for a talking head and the hydraulic fountains of Schönbrunn.

After the demonstration, von Kempelen vowed to produce something still more amazing and, six months later, he reappeared at court with "an object of intense curiosity to all persons who think" (Poe, 1836): the Automaton Chess-Player.

This Automaton Chess-Player resembled a man, approximately life size (only Poe disagrees about this), sitting at a chest. The whole construction was mobile. The upper part of the body was dressed in Turkish style and had a roughly carved wooden face; both the head and eyes moved. The right arm reached forward towards the spectators and the left hand held a long pipe, which was removed just before commencement of play - the Turk, as he was soon called, playing left-handedly. The chess board was permanently fixed to the top of the chest. The lower figure was not body-like, but consisted of a tall box, apparently full of machinery, although a pair of drooping slippered legs and a long cloak disguised the fact during a game.

The chest itself consisted of two cupboards and a drawer at the front, and three cupboards at the back. These cupboards were opened with a considerable element of ritual before a game and audiences were generally satisfied that it was impossible for anyone to have remained inside. An eye witness allegedly reported: "I searched into its darkest corners but found no possibility of its concealing any object of even the size of my hat." (Levy, 1976). A third of the cabinet was to all appearances densely packed with machinery and the largest cupboard was almost empty; the drawer was very shallow and extended the full depth of the chest. A pair of candelabra, one on each side of the chest, ensured that all was well lit and the whole Automaton was wheeled about on casters to any point in the room requested by a member of the audience.

A challenger from the audience having volunteered, the machine was (seen and heard to be) wound up by the exhibitor or "governor", as he was known. It is interesting to note that the term "governor" was also used to refer to the person employed to look after, repair and collect clocks - he was usually either the inventor (maker) or, at least, a talented mechanician (Cipolla, 1967). The challenger sat at a nearby table, in order not to obscure the view of the audience. The pipe was removed and, the governor having manipulated machinery inside the lower torso, the game began. The Automaton played white (i.e. first) and moved his pieces with jerky angular movements to the accompanying sound of clanking machinery. He would occasionally drop a piece if it had not been exactly centred in the squares (the governor moved the opponent's pieces on the Turk's board) - but carried on as if it hadn't "noticed". The Automaton was occasionally rewound.

Games lasted for about thirty minutes, in order not to weary the audience, who were presumably usually more interested in the mechanical man than the game itself. In addition to merely playing, the Turk would roll his eyes and move his head as if surveying the pieces. When he threatened the opponent's Queen he would nod twice and three times when putting the opponent's King in check. He was also somewhat imperious. If an opponent made an illegal move, he would rap the table top and shake his head, correct the move and then himself move without giving his opponent a second chance. On occasion, in an apparent fit of mechanical pique, he would sweep the pieces from the board. According to some, this happened in 1806 when the Turk was playing against Napoleon. Three years later, in a second contest at Schönbrunn, Napoleon was soundly beaten. The game is reproduced in Levy (1976). The Automaton generally, but not invariably, won his games, and at times would undertake virtuoso demonstrations and end games.

The willingness of the governor to move the Automaton to any point requested by the audience, and its being raised from the ground on castors, precluded the use of trapdoors, thin wires or whispering. The inside was openly shown to be either empty or else packed with machinery. What could one conclude except that one was witnessing the "most amazing automaton which has ever existed." (Windisch, 1784)? If Augustus, Duke of Brunswick was right in 1676 when he remarked "Chess is the Art of Human Reason" then the Automaton seemed to offer, at last, tangible proof that machines _could_ think. As Edgar Allen Poe reports: "we find everywhere men of mechanical genius, of great general acuteness, and discriminative understanding, who make no scruple in pronouncing the Automaton _a pure_ machine ... The most general opinion ... an opinion too not unfrequently adopted by men who should have known better, was ... that no immediate human agency was employed - in other words, that the machine was purely a machine and nothing else".

This was certainly true in _some_ cases, at least. In his worthy _A Dictionary of Mechanical Science, Arts, Manufactures and miscellaneous Knowledge_, published in 1827, i.e. fifty to sixty years after the creation of the Automaton, Alexander Jamieson described and discussed it without revealing the slightest suspicion whatsoever of fraud. Francis

Bowen still took the Automaton Chess-Player seriously enough in 1877 to take the trouble to draw (for him) worriesome conclusions ... he mistakenly pays Maelzel, the Turk's second owner, the compliment of attributing the authorship to him: "Mr. Huxley pithily expresses the necessitarian doctrine, when he protests, 'that if some great Power would agree to make me always think what is true and do what is right, on condition of being turned into a sort of clock, and wound up every morning before I got out of bed, I should instantly close with the offer.' The ingenious Mr. Maelzel, who, nearly a century ago, constructed a wooden man, about three feet high, that played a good game of chess, also fashioned a smaller puppet, which pronounced quite distinctly a number of words. Now, it matters not at all, most persons will think, whether a sentence uttered by this puppet be true or false, since there would be just as much merit, or demerit, in the one case as in the other. And if all mankind were wooden images so constructed, I think that the difference between truth and falsehood, or between a right action and a wrong one, would not concern them in the least, and in fact would have no meaning for them. Mr. Huxley's remark, if intended to be taken seriously, merely shows the lamentable cynicism, which is the only state of mind that can logically result from belief in a materialistic and fatalistic theory of the universe."

However, if this was the "most general opinion," many records of it have not survived. Claude Shannon (1950) is probably nearer the truth when he remarks "Most analysts concluded ... that the automaton was operated by a human chess master concealed inside." The interesting thing for us though is not what these conclusions were, so much as the way in which they were reached.

The acid test, of course, would have been to interrupt the Automaton mid-game, to have dismantled it and tried to find the hidden "master". Happily this never came about, although it was at times a possibility from at least two directions. Early automata makers, with their supposed Promethean aspirations and Paracelsian leanings, were frequently suspected of necromancy or black magic. Windisch remarks on an old lady in the audience who, during a game, crossed herself and moved to a far corner of the room "in order to be as far as possible from the evil spirit which she probably thought animated the machine". Fortunately, von Kempelen, a court counsellor and highly respected mechanician and later Maelzel, a consummate showman, never allowed any suspicion of hysteria to develop.

The other more "empirical" line was voiced by Thicknesse (1784), a highly incensed Englishman, outraged at the thought of a "Foreigner ... collecting an immense sum of money in this Kingdom, to carry into some other, by mere tricks" and indignant at the "folly of my own Countrymen, and the arrogance of the imposing stranger". He recalled, gloatingly, his exposure of a previous "automaton" ... "a coach which went without horses". Then "a small paper of snuff, put into the wheel, soon convinced every person present, that it could not only move, but sneeze too, perfectly like a Christian". It doesn't take much perspicuity to see that Thicknesse would dearly have liked to extend this "empirical" line of enquiry to the Automaton Chess-Player. It is, presumably, a muddling of stories that makes Raphael (1976) wrongly maintain that "the secret finally came out one day when, in the midst of a match and in front of a large audience, a loud sneeze was emitted by the midget, an expert chess player who was hidden in the cupboard".

Given that commentators <u>couldn't</u> dismantle the machine, what then was their approach? Interestingly, not to say why von Kempelen's creation wasn't an automaton - but why it couldn't be. Poe again (typically) gives the least cautious line: "It is quite certain that the operations of the Automaton are regulated by mind and by nothing else. Indeed this matter is susceptible of a mathematical demonstration a priori". Unfortunately (but predictably) he doesn't give that demonstration. David Brewster too, with the benefit of hindsight, thought: "Upon considering the operations of this automaton, it must have been obvious that the game of chess was performed either by a person enclosed in the chest or by the exhibitor himself".

On the contrary though, it was not obvious. Why should it have been? The crux of the matter (sic) was the near total antonymy in many minds of "rational" and "inanimate". Thicknesse remarks in a slightly different context (of a speaking doll): "a rational answer conveyed through the head of an inanimate being ... consequently made by a rational being". Either the Automaton Chess Player was rational and animate and therefore not an automaton, i.e. a fake, or inanimate and arational and therefore not a chess player, i.e. a fake. The very expression "Automaton Chess Player" seemed to make analytic nonsense. Moreover since it manifestly did play chess, it simply couldn't be an automaton.

Commentators then, decided that there must be a human agent involved: the "movements (were not) really performed by mechanical powers ... (but) supported ... by invisible confederates" (Thicknesse, 1784). But where was the confederate?

There were two ready resolutions of the dilemma - either someone(s) in some fashion influenced the game from outside, or was/were concealed inside. In these early days of rapid technological innovation, the mental ground was fertile for suggestions of the employment of new physical forces. The willingness of the exhibitors to move the Automaton about at request, did away with mundane direct contact methods of communication, e.g. wires, strings, etc., but some commentators did suggest the use of that strange new phenomenon - magnetism. Hindenburg (1784), for example, thought the effect to be caused by a combination of machinery and magnetism. Certainly the exhibitor at times encouraged the credibility of the use of magnetism - for example by keeping in a pocket a hand which was obviously holding something. The exhibitor "is supposed to be the person with whom the stranger actually plays, by causing the arm and head of the Automaton to move the Chess-men by some incomprehensible and invisible powers ... a Magnet concealed in his pocket" (Thicknesse). At the same time, however, both von Kempelen and Maelzel allowed a member of the audience to stand a powerful magnet on the cabinet, thus mystifying the speculators. Note that von Kempelen could quite reasonably have objected - if his Automaton were a sensitive machine, it would have been quite possible that a strong magnet could have interfered with its delicate workings. And, after all, magnetism is a physical force - what an achievement to have built an Automaton built on magnets! But, of course, the public was set on the notion that although given a rational director, magnetism might be the way of communicating via the machine, the thinking would of course have come from an intelligent (i.e. animate) source.

The other immediate reaction was the "midget concept". This crops up again and again: the "real mover is concealed in the counter" (Thicknesse) ... "the Automaton Chess-Player is a man within a man; for whatever his outward form be composed of, he bears a living soul within". The midget concept has several obvious sources. Firstly, the exhibitors had apparently demonstrated that there simply wasn't room for a full-sized man. Secondly, there were precedents. Brewster mentions the case of M. Raisin, the organist of Troyes, who built an automaton harpsichord player - later discovered to contain a five-year-old. Moreover there seems to be something intrinsically satisfying about the "little man" type of explanation. Who is able to resist the temptation of an assembled set of Russian nested dolls? More formally (but only slightly) Osgood has inveighed against theories which suppose "a new homunculus in our head", whilst Jerry Fodor (1968) has openly advocated their adoption as explanatory tools. To return to the Automaton. Legend embellished the tale by suggesting that von Kempelen, whilst visiting Russia to study the mechanics of speech, built the Automaton to smuggle out of Russia a Polish insurgent named Worousky, whose legs had been blown off in battle. (An heroic tale which is, alas, quite definitely false.) In any event, the mistaken view that the Automaton contained a midget or child persists in some quarters till this day. A recent children's introduction to computers (Scharff, 1975), states that the Automaton contained "a midget who was an excellent chess player and who manipulated the automaton from his hiding place" and Raphael makes the same mistake.

If intelligent systems are necessarily animate, what characterises machines? According to Poe "regularity" is "important in all kinds of mechanical contrivance". Once given data and started "it should continue its movements regularly, progressively, and undeviatingly towards the required solution", "these movements, however complex, are never imagined to be otherwise than finite and determinate". Thicknesse corroborates this view: "that an Automaton can be made to move ... in certain and regular motions, is past all doubt; but that an Automaton can be made to move the chessmen properly, as a sagacious Player, in consequence of the preceding move of a stranger ... is ... utterly impossible".

Robert Willis (1821): "however great and surprising the powers of mechanism may be, the movements which spring from it, are necessarily limited and uniform".

Commentators expected the Automaton, if truly mechanical, to make moves almost to the beat of a metronome. However, not only did it vary in the time it took to move (this was surely to be expected even on the mechanical hypothesis), but it flexibly adjusted itself to its opponent's game. This was unthinkable: "it cannot be made to vary its operations so as to meet the ever varying circumstances of a game of chess. This is the province of intellect alone" (Willis); even if the "movements of the Automaton Chess Player were in themselves determinate, they would be necessarily interrupted and disarranged by the indeterminate will of his antagonist". It is interesting to compare this attitude with that of Spielmann: "In the opening a master should play like a book, in the middle game like a musician, in the ending like a machine" (Levy, 1976). Ironically then, the Chess Automaton would have been more convincing if it had been less successful. Alas the mechanical simulation of intelligent processes stands to lose either way; if it is too good then it cannot be mechanical, if it is not good enough, it is not intelligent.

Connected via appropriateness, to the problem of its mental competence (in the twentieth century style) was that of its behavioural performance. Two points were made in this respect - firstly it was noticed that whilst the Turk did indeed shake his head and roll his eyes during a game - he did it most of all during easy periods of the game! It did occur to the critics that this was perhaps because only then was there sufficient "computational space" to co-ordinate such actions - that during difficult periods, the directing intelligence was too fully engaged in working out its moves; but they took this as proof that the directing intelligence must be human not mechanical.

Secondly, critics were suspicious of the exaggerated primitiveness of the figure itself and its movements. The face was roughly hewn, but worse, the manoeuvres were "awkward and rectangular". Critics were suspicious of the motives of the builder, for it was widely believed that he was a very skilled mechanician indeed and could have made the movements immeasurably more natural. The only intention he could have had, to thus underperform, was to suggest the idea of pure unaided mechanism.

Several other features of the performance were considered to be red herrings ("merely to puzzle" (Thicknesse)). Robert Willis considered this to be the function of all the machinery which he noted was never actually seen in action. Attributing to the creator a modicum of pride in his own achievement, Willis reasoned that he should have been only too willing for the innards to be seen in their full glory - especially as the audience could contemplate it as long as they wished whilst it was inoperative. There was a "glaring contradiction between eager display on the one hand, and studied concealment on the other". Even more perceptively, he paid attention to the actual winding up of the (alleged) clockwork and concluded that not only did the key meet too little resistance to power such a lot of machinery, but, more importantly, that there was gross inconsistency in the frequency of windings. The machine was variously rewound by the exhibitor after sixty three, seven, three and on one occasion - presumably inadvertantly - after no moves at all; each winding consisting of approximately the same number of rotations. Willis took this as "positive proof that the axis turned by the key is quite free and unconnected ... with ... any system of machinery".

The extravagant candelabra too, in the well lit exhibition rooms, raised suspicions. As Poe observed, they can't have been for the benefit of the audience - who could see well enough already, and "if we suppose the machine a pure machine, there can be no necessity ... for any light at all". There must, then, have been some ulterior motive. For theatrical effect? This didn't seem to occur to anyone. It was suggested instead that the strong light was required so that the "inner man" could peer through the gauze of the Turk's breast, whilst being safe from detection by the resulting dazzling reflection from the gauze. The actual reason - to cover up the smell of a candle burning elsewhere - never, apparently, occurred to anyone.

Other arguments against the possibility of the Chess Playing Automaton being a "pure machine" were, of course, put forward. These range from the misinformed via the plain silly to the genuinely original.

Poe, for example, wrongly claimed the Automaton to be a lot <u>larger</u> than it was, while Bowen, on the other hand, maintained that it was only three feet high. Poe also made the astonishing deduction that because the Turk played with its <u>left</u> arm, this "absurdity" must be so that the inner man could more easily manipulate it with his <u>right</u> hand. However he also commented on a character called Schlumberger, who was often seen around Maelzel's (the then owner and exhibitor) suite, whilst the Turk was not playing, but who was never to be seen when it was; once when Schlumberger became ill the performance was cancelled at short notice. Worse though, Schlumberger professed total ignorance of chess, although all the others in the party could play; an incongruity Poe found difficult to accept. In true detective style, Poe finally left the discussion of the mysterious Schlumberger - with the observation that he "has a remarkable stoop". Carroll (1975) informs us however that Schlumberger who was about six feet tall and well proportioned had made his name as a player and teacher of chess (he taught Pierre St Amant, one of the great masters of the nineteenth century) at the Café de la Regence, the chess centre of Europe. Moreover, Maelzel introduced Schlumberger to the Bostonian chess circles and frequently played with him himself.

Another piece of Poe's deduction was that the opponent sat apart from the Turk, not in order to give the audience a clear view, but so that he would not hear the sound of breathing from inside the cabinet. Less comprehensible is Poe's reaction to the fact that the machine did not always win. "Were the machine a pure machine," he says, "it would always win. The principle being discovered by which a machine can be made to <u>play</u> a game of chess, an extension of the same principle would enable it to <u>win all</u> games". This deduction is, no doubt, based on the "self evident principle" that the "difficulty of making a machine beat all games, is not in the least degree greater, as regards the principle of the operations necessary, than that of making it beat a single game". The A.I. community at Stanford and M.I.T. would surely be glad to know this.

As a refreshing contrast to Poe's sophism, consider Thicknesse's realistic psychological argument: the Automaton couldn't be a pure machine because, if it were, the exhibitors would not exhibit it only such relatively short and infrequent periods, but would have it almost permanently on show: "the invisible player could not bear a longer confinement; for if he could, it cannot be supposed that they would refuse to receive crowns for admittance from 12 o'clock to 4 instead of only from 1 to 2". There must be a confederate inside who simply could not stand it any longer than he did. The serious side of this point is, of course, the ethical one, that we don't feel a moral responsibility towards machines; nor of course do they (in an important sense) have a choice about co-operating with us. Remember the term "Robot" itself comes from the Czech for "work".

The last major line of scepticism was that stemming from the undeviating way the innards of the cabinet and Turk were displayed. Most commentators mentioned this point. Robert Willis speaks for them all in his conclusion that the "regular and undeviating mode of disclosing the interior" must mean that "more (is) intended in the disclosure than meets the eye". What was intended, it was assumed, was that something should remain hidden, whilst the impression was given that all was displayed. And that something? - a man, of course.

An interesting feature of commentaries given by the critics is the way they analysed the Automaton Chess Player into psychological components. Hindenburg (1784) was perhaps the first to suggest that the Automaton must have two mechanisms: A Gehewerk (motive force) to move the arm, head, etc., and a Schlagewerk (directing force) to play the game. Thicknesse is really onto the same point when he remarks, in a footnote, that the "pretended Automaton ... points both a directed, and adds to it the human faculties, by playing with judgement". Robert Willis (1821) characterises the problem as explaining the mechanical functioning, the directing intelligence and the communication between the two. Surprisingly enough, the perceptual problem - how the Turk perceived the chess configurations - seldom seems to have concerned the commentators. Only Brewster really seems to see there is a problem and resolves the issue by deeming as bogus all three problematic functions (perception, movement and intelligence) ... a man sees through the waistcoat, manipulates the arm and does the thinking ! Certainly the essential insight that a machine is both a transformer of power <u>and</u> of information went unreceived.

This is the place to make clear - as you will already have guessed -

that von Kempelen's Automaton Chess Player was not in fact a "pure machine". It contained not a midget, nor a child, but a fully grown man (at one stage in its career Schlumberger). The cabinet was so skilfully contrived, by means of a telescopic drawer, foreshortened cupboard and hollow machinery, that a man could, and did, remain undiscovered inside. But for more than sixty years, no-one except the exhibitors and players themselves knew this for certain. For our purposes, the fact that the machine was in fact not totally a machine is irrelevant - for of main interest was the argument, or rather conviction that it could not be.

To be fair, neither von Kempelen nor Maelzel ever actually claimed that the Turk was a pure machine. Indeed, Maelzel's refusal to openly state that it was, was one of the reasons which convinced Poe that it was not. Thus, although Thicknesse entitled his essay: "The speaking figure and the automaton chess player exposed and detected" and admitted later that it was a "good deception" (as a "good deception" rather than a bad piece of sincerity, it merited a price of entry but "the price at least may be reduced"), in fact, according to Brewster, "its ingenious inventor, who was a gentleman and a man of education, never pretended that the automaton itself really played the game". In fact, von Kempelen had openly stated that "the machine was a bagatelle which was not without merit in point of mechanism but that the effects of it appeared so marvellous only from the boldness of the conception, and the fortunate choice of the methods adopted for promoting the illusion". (Brewster). It is unfortunate then that the slander still continues; even Claude Shannon, after discussing the Automaton Chess Player, goes on to talk of L. Torres y Quevedo's chess machine as a "more honest attempt".

We now have the strongest reason to believe that there are no a priori reasons against the possibility of chess playing automata - the existence proof. As Yorick Wilks (1976) has remarked in connection with Dreyfus' scepticism over whether machines can play chess: "A.I. workers have contented themselves, with ... Johnsonian refutations - remembering his treatment of Berkeley ... Dreyfus played a mechanical chess-player in 1969, was beaten ... and for many A.I. people that settled the question". As mentioned, Torres y Quevedo had, as early as 1890, constructed a machine capable of an endgame. As for his motives we are told: "The machine was created as a scientific toy in order to attract attention to the feasibility of Torres' theory on automation". (Levy). It is an interesting question how far von Kempelen himself shared this motivation. More recently, in August 1974, Kaissa became the world computer chess champion in Stockholm. Even so, although machines can play a "good" game, they do not (yet) come up to grand masterclass. It remains true that von Kempelen's Automaton was "almost certainly better than any computer program that has yet been developed" (Raphael, 1976).

Modern qualms over the possibility of chess machines, are mostly concerned with the problem of formalisability. Is chess playing the sort of process "which can be formalised so that it can be represented as a series of instructions for the manipulation of discrete elements" (Dreyfus, 1972)? Bronowski (1965) generalises the point: "Is man a machine ...? (the) answer hinges on possible modes of knowledge ... If all knowledge can be formalised, then the human self can be matched, in principle by a machine". What is to count as formalisable is however a function of the theory, but in A.I. theory, model and program are very closely linked - if not identical. Indeed as Weizenbaum comments, many programmers think of the language they use as <u>being</u> the computer. In a real sense, programming creates a new machine; or transforms an old machine into a new one. A "malfunctioning" machine is miscalled - it is in fact a well functioning one - although not the one we want. In this way of thinking "formalisable" (which is often synonymous for many with "strictly explicable") comes perilously close to meaning "programmable", rendering Bronowski's point trivially analytic. Anyway, even Dreyfus doesn't maintain that no computer could play chess (well) - only that digital ones could not.

Even Joseph Weizenbaum, doing his best to shoot down the man-machine analogy and with the advantages of the close familiarity of a professor of A.I., doesn't doubt that computers can be regarded as organisms or that man is (at least) a machine. His argument is not that there are human functions which a computer couldn't take over - but that there are some which, for moral reasons, they shouldn't be permitted to. McCarthy (1974) comes closest to our own position: "Processes of problem solving depend on the class of problems being solved more than on the solver, e.g., playing chess requires lookahead, whether the apparatus is made of neurons or transistors".

It is these problems with which the (theoretical) cognitive psychologist, who wishes to avoid an ontology of occult entities, is concerned. For the brain is not only an intelligent information handling system - but also a model of one. A brain is as much a model of a computer as a computer is a model of a brain, in so far as they can both solve the same classes of problems. Which we take as a model of which is a pragmatic issue depending upon which we are more familiar with in a relevant respect. Both are physical entities; both are machines. One consists of transistors, etc., the other of cells. Even cells are sorts of machines: "the thought that the cell is a machine (wherein occurs) ... conversion of inputs to outputs via a device admitting of states invokes a familiar class of objects as cellular models, while apparently vindicating a very old philosophical perspective" (Berlinski, 1972).

By the end of the eighteenth century, genuine automata were by no means infrequent visitors to the intellectual milieu. There were mechanical theatres, conjurors, musicians and animals galore, amazingly complex clocks and jackwork of astonishing calibre. What was the particular incredulity attached to the idea of a mechanical chess player due to? The critics, we have discovered, baulked at the notion that the difference between this and earlier automata was merely a matter of sophistication of mechanism, but considered that what was required was the introduction of a new and qualitatively incompatible element, that of intelligence. Today we have still not solved the problem of how organised matter can be endowed with the faculty of thought. This remains the great challenge to cognitive psychology.

Let us return to the Turing test. We can now see that our mental attitude, in playing the Turing game, has undergone a great change. For the audience of the Automaton Chess Player were in fact playing the game, where the teleprinter was replaced by a wooden doll and the other room by a wooden cabinet. What was in the box - man or machine? For the eighteenth century man of letters there could be no doubt - it had to be a man, a priori. The Fodorian canon would have been senseless to him. How could a machine successfully simulate the behaviour of a man, when machinehood entails inanimacy, a man's cardinal feature is his rationality and rationality entails animacy.

The influence of the Automaton Chess Player spread however beyond its immediate application. It became a pervasive psychological model ... an example of the (apparent) blatant fraudulence of claims that machines could (can) think. In 1879, long after the controversy over the actual Turk had died down, we find Bowen writing: "Let the physiologist or chemist contrive what mechanism he may, if an indivisible Ego of consciousness is not allowed to come in, the machine will not work. The automaton won't play chess if an Ego be not smuggled into the cupboard". References to the ghost in the machine persist to this day.

# CHAPTER SEVEN

## SPEAKING MACHINES:

## utter nonsense or the psychological history of the chatter-box?

"Hasce igitur penitus voces cum corpore nostro
Exprimimus, rectoque foras emittimus ore,
Mobilis articulat verborum daedala lingua
Formaturaque labrorum pro parte figurat."

Lucretius.

(When we force out these utterances from the depths of our body and launch them through the direct outlet of the mouth, they are cut up into lengths by the flexible tongue, the craftsman of words, and moulded in turn by the configuration of the lips.)

In 1692, the patron of Jonathan Swift, Sir William Temple, wrote: "We pretend to give a clear account of how thunder and lightning (that great Artillery of God Almighty) is produced, and we cannot comprehend how the voice of man is framed - that poor little noise we make every time we speak." That "poor little noise", and its method of production had, however, puzzled and intrigued men from ancient times, through the middle ages and was to occasion an explosion of experimentation in the late 18th and early 19th centuries, which would lay the foundations for a modern psychology of speech perception, production and speech synthesis.

The subject of speaking machines is of considerable interest in its own right and, during this chapter, I shall mention and/or discuss some twenty (alleged) speaking machines. However I shall concentrate for the most part on those produced by Wolfgang von Kempelen. This is not only because we have already met him in connection with the Turk, his Automaton Chess-Player, but also because his intentions behind the construction of his speaking machines are, to my knowledge, in his case the best documented and most interestingly stated. Indeed, although I hope to have mentioned in this paper most speaking machines of any note, and thus provide a useful source of references, my prime aim is not one of historical description (with its implication of strictly applied minimal standards of evidence) but an illustrative piece of philosophy of psychology.

The case of speaking machines is particularly interesting for four further reasons. Firstly, the readiness with which it lends itself to support the thesis that many automaton and machine builders, as simulators and thus explainers of that part of the natural world which is man, should be considered as a rich and fertilising part of both the history and actuality of psychology.

Modern psycholinguists still consider speaking machines to be important elements in research. Liberman et al (1959), after stating their aims with regard to speech synthesis ... "a set of rules, written down, perfectly explicit in all particulars, so that a person with no knowledge of speech or spectograms could, by reference to the rules, synthesise speech as well as anyone else", suggest: "It may help in setting the problem to think in terms of a machine that will process a discrete phonemic input in such a way as to produce a speech output."

Secondly, we there witness a major maturing of the notion and practice of simulation. In the steady development of speaking machines we see the tacit requirement that the simulation look like the simulated dropped. As speaking machines stopped looking like human heads and became more like blacksmith's appurtenances, we see a move from imitation to truer (functional) simulation; yet, because it was a putative theory of the human speaker (a bit of cognitive simulation rather than artificial intelligence, in modern parlance), the requirement that the machine and person produce equivalent utterances in the same way, was retained.

Thirdly, the case of speaking machines serves to link together several figures we have already encountered in previous chapters and, interestingly to show both the interconnections between and cross-fertilisation among researchers in the field and the intellectual links both back into antiquity and forwards into modern times. Thus we see the problem as appreciated, by research workers of the time, in the terms of debate set by Descartes and La Mettrie, who were responsible for a focusing of attention on the clock analogy and Jaques de Vaucanson, who gave the abstract notion of 'automaton' some concrete substance.

Von Kempelen was creator of both the Chess-Player and some speaking heads; Maelzel, who exhibited and repaired von Kempelen's Turk, also made a speaking head, as did Rechsteiner who, after spending years repairing Vaucanson's duck, had also made a duck himself. Kratzenstein, another maker of a speaking head, was also the designer in 1751 of an early self-winding clock.

The fourth reason for special interest in this area is the crucial nature, to the 17th and 18th century mind, of the question of the possibility or impossibility of getting a machine (or animal) to speak.

First though, what do I mean by a speaking machine? It is most important to distinguish the subject of this chapter from phonographs, graphophones and gramophones, which, in their early days, (around 1880) were often collectively referred to as 'talking machines' n.b. the Edison Centenary Exhibition programme (Royal Scottish Museum, Edinburgh) was entitled: "It talks, it whispers, it sings." These later machines <u>can</u> serve as (intimate) models of psychological abilities ... witness the record or tape in theories of memory. Indeed, even a musical box may serve this function. As Huxley wrote: "A song which has been learned has its molecular equivalent which potentially represents it in the brain just as a musical box wound up potentially represents an overture." But phonographs etc. are not elucidatory models of speech <u>production</u>. Partly this is because they are so inflexible (they require so much preprogramming) but much more because they are so less obviously structurally isomorphic with the human vocal production apparatus. A speaking machine actually produces and synthesises words rather than merely reproducing prerecorded ones. As Chapuis and Droz (1952) put it: "The gramophone is in no sense a machine which talks by forming for itself the words and phrases which it produces. It is confined to the repetition, now almost perfect, of what it has recorded and kept." The converse is true of speaking machines, which must try to produce the entire behavioural repertoire of the organism.

In 1774, Abbé Mical exhibited two speaking heads to the public, thus confirming the old adage that two heads are better than one. Compardon, in "Les Spectacles de la Foire", quoting from Journal de Paris of 1st May, 1778, affirms that, in so doing, Mical "has solved the problem which has been believed insoluble from Archimedes to Vaucanson."

What was this problem? In 1748, in L'Homme Machine, Julien de la Mettrie wrote that: "Vaucanson, who needed more skill for making his flute player than for making his duck would have needed still more to make a talking man, a mechanism no longer to be regarded as impossible, especially in the hands of another Prometheus." It is important to realise that the issue of a talking mechanism was a crucial one in La Mettrie's interpretation of Descartes.

To put it briefly, Descartes proposed two principles for distinguishing men from machines (given that to all outward appearances they were identical). The main one was that a machine could never use speech or other signs. As Descartes wrote to Henry More: "For language is the one certain indication of latent cognition in a body, and all men use it."

La Mettrie accepted that Descartes had proved that animals are machines, whilst disagreeing with his (alleged) view that men were not machines as they were bestowed with an immaterial soul. On the contrary La Mettrie (1748) believed that there was no essential difference between animals and ourselves: "The transition from animal to man is not violent." Indeed, "men are at bottom only animals and machines which though upright go on all fours." La Mettrie thought it possible to teach animals to speak. Talking of the lack of speech in the ape he wrote: "What is the reason for this, except some defect in the organs of speech? But is this defect so essential to the structure that it could never be remedied? In a word would it be absolutely impossible to teach the ape a language? I do not think so."

But anyway, as he is to "conclude boldly that man is a machine" and manifestly does speak: as animals are machines and manifestly 'could' speak; it must follow for La Mettrie that machines could and can speak. However la Mettrie, who distrusted a priori arguments, avowing that only "experience and observation should be our only guides", rejected "idle theories" and appealed for the "stuff of experience" from those who have "laid bare to us springs of life." Remembering that man is but "a collection of springs", who better to lay them bare than the automaton-simulationists? In short, prior to la Mettrie the crucial difference between men on the one hand and animals and machines on the other, was the former's ability to use language or produce speech. (True language use, of course, involves not only speech production but also symbolic manipulation. The mechanicist's reaction to this was discussed in the chess-player paper. There is, of course, all the difference in the world between speaking (which parrots do) and saying (which, as far as we know, they do not.) If the mechanistic philosophy was to achieve true generality, it was necessary to demonstrate that animals and/or machines could speak (this is still true today).

Speaking of the Past.

References to speaking heads and figures reach far back into antiquity, when there was a dual motive for automata making: that of the exhilaration of developing a technical skill in simulating human performance and the more dubious task of usurping the power of the gods

- often associated with interest in necromancy and magic. This is a suggestion supported both by Bulling's study of ancient Chinese mask and puppet making, which was related to the cult of the dead, and by Cohen's (1966) researches. Certainly the first speaking machines I shall mention are connected with religion and the magical art of fortelling the future.

The statue of Memnon at Thebes, west of the Nile, was referred to by Strabo, Pliny and Pausanius. It is accredited with different abilities by different commentators. According to some, it produced lyre-like music but for Tacitus it was a vocal sound; still others reported oracular verses. Philostratus reports that: "when touched by the first ray of the sun it uttered a sound as soon as the sunbeam reached its lips." This suggests that, if it really existed, it may well have operated according to mechanical principles - powered by air or steam pressure, much as Hero's automata in Alexandria. It may, of course, have been a natural rather than manufactured machine. Von Humboldt, in South America, came across a rock which produced musical sounds at sunrise ... caused by the escape of air, under pressure due to warmth, from crevices. Whether or not, in the case of Memnon, an enterprising artist discovered such a rock and carved it into a statue rather than quarrying out the ducts etc., is a matter of conjecture.

There were many ancient oracular heads - one of Orpheus at Lesbos foretold the death of Cyrus. The Egyptians possessed prophetic statues which replied to questions by nods and associated cries. Some were operated by steam or fire. An interesting large limestone bust of Re-Harmakis was discovered in the 1930's. This proved, upon examination, to contain a cavity in the neck with a narrow canal leading to an aperture under the right ear ... presumably a priest was hidden behind the statue and made pronouncements on its behalf. This faking technique was inevitably used again and again. In the 18th century we read: "the brains of this wonderful doll is nothing more than the continuation of a tin tube, which is fixed to its mouth, so as to convey the Question and Answer to and from an invisible confederate."

Pope Silvester II (alias Gerbert) was an amazing polymath by any standards. A master of astronomy, astrology, necromancy, music, bird song and flight, geometry, mathematics; he also introduced Arabic figures into Western Europe, took over and developed the abacus from the Saracens and constructed both a clock and an organ. More germane to this discussion, he also built a speaking head. According to William of Malmesbury (12th c.) this head "spake not unless spoken to, but then pronounced the truth, either in the affirmative or the negative. For instance, when Gerbert would say, "Shall I be pope?" the statue would reply "Yes""! Robert Grossteste (1175-1253) is also reputed to have built a speaking head but little is written of it. We know more about that of Albertus Magnus. Becherus, in 1680, wrote that Albertus Magnus spent thirty years making a walking talking automaton which saluted and spoke to his (A.M.'s) friend Thomas Aquinas, who thereupon immediately smashed it. We may suspect that Aquinas saw in the automaton evidence of the intervention of the powers of darkness, but, according to Brewster in his entry on Androids in the Edinburgh Encyclopaedia, Aquinas had a less superstitious and more human motive, he "came to see it, purposely that he might boast how in one minute he had rendered fruitless the labour of many years." Roger Bacon (1214-1292) too, is reputed to have built "a brazen figure capable of speaking", a head commemorated by Samuel Butler in Hudibras where he writes: "My head's not made of brass as Friar Bacon's noddle was."

We now start approaching our main focus of interest. In 1650, in Musurgia Universalis, Athanasius Kircher, a Jesuit scholar of Rome describes an "automatic organ machine which utters the voices of animals and birds," and thirteen years later, in Phonurgia Nova, maintained that it was possible "to produce a head which moved the eyes, lips and tongue, and, by means of the sounds which it emitted, appeared to be alive." It appears that he tried to build such a head for Queen Christina, but failed. Serious work in phonology was now beginning to get under way and in 1668, John Wilkins D.D., a founder Fellow of the Royal Society, published "An Essay Towards a Real Character and a Philosophy of Language", the tenth chapter of which was devoted to "A rational account of all the simple sounds that are or can be framed by the mouths of men."

There is some evidence that Valentin Murbiz, a rector from Dresden, devoted energy to the problem of a speaking machine around 1705, but the next concrete landmark was not until 1770, when Friedrich von Knauss of Vienna made four speaking heads. Friedrich Knauss (b. 1724) and his brother Johann Philip Ludwig (b. 1715) were both astonishingly skilled automaton builders. Apprenticed as boys to a clockmaker, Friedrich was only 21 when they built the Ritterspieluhr, now in the Hofburg Museum, Vienna, a clock in which two armies of horsemen joust whenever the hour strikes. By 1781, however, when they completed the Maria-Theresian Uhr, a horological tableau representing the coronation of the Empress and Francis I, Ludwig had specialised in the clock side of the business and Friedrich in automata. Friedrich went on to become the head mechanician of the Physicalischen Kabinett, building four writing machines (three disembodied hands and a Grecian lady) and four speaking heads, one of which was sent as a gift to the Grand Duke of Austria. In a very rare book of 1789, <u>Selbstschreibende Wundermaschine auch mehr andre Kunst und</u>

Meisterstücke, Knauss apparently (Hillier, 1976) expressed the aim of linking a writing hand to a speaking head, so that one could say out loud what the other had written, an aim he, hardly surprisingly, never realised. Although by now it really should not need re-emphasising, it is perhaps worth noting Mary Hillier's comment that "He (Knauss) had the sort of mentality now employed by those who invent computer systems."

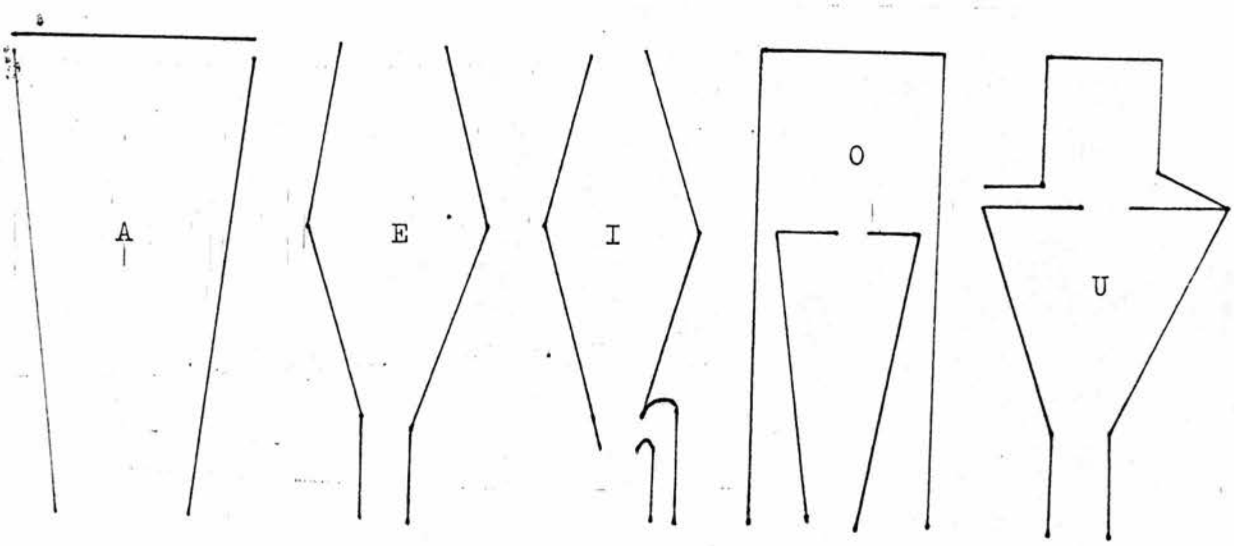
Evidently, however, Knauss's heads, though stimulating interest, were not very successful, for sever years later occurred the most important event in the history of talking machines. In 1777 the Imperial Academy of Sciences, St. Petersburgh, offered a prize for the best enquiry into "the nature of the five vowel sounds A, E, I, O and U; and for the construction of an instrument for artificially imitating them."

There were three entries ... by Mical, Kratzenstein, who received the prize, and Wolfgang von Kempelen.

Abbé Mical, we have already heard of as the solver of Vaucanson's problem. Compardon attributed Mical's lack of acclaim less to his ability than to his temperament. "Since", Compardon writes, (Mical) "is not a schemer, since he is working on his own without organised support or without accomplices, since he has not bribed anyone to puff him and has not secured the goodwill of the newspaper writers, little has been heard of his machines, which have secured the widespread admiration of natural philosophers". Indeed Mical did lack a showman's flair ... when he exhibited his two heads to the public in 1774, his advertising display read rather modestly: "Problem in mechanics solved." Mical seems also to have been rather temperamental. Apparently he himself considered the praise lavished on his heads to be excessive, perhaps even impious, and himself smashed them.

Kratzenstein, Professor of Physiology at Halle and later Copenhagen,

FIGURE A.



also the designer in 1751 of a clock which wound itself up by harnessing variations in temperature, actually won the prize with a set of five tubes with specially shaped ends, each of which "distinctly produced" a vowel, when their lower end was blown through. The tubes are illustrated in Figure A.

There are two points of special interest in Kratzenstein's winning entry. Firstly that these tubes were "suggested by observation of the form and dimensions of the human mouth when sounding different vowels" (Paget, 1930) i.e. there was a rudimentary attempt to simulate; but secondly, although they imitated a man's voice "with tolerable accuracy", they gave no indication of the underlying acoustic nor psychological principles.

The third member of the trio was Baron Wolfgang von Kempelen, Aulic Counselor of the Chamber of the Domain of Empress and Queen Maria Theresa, who also made the plans for the fabulous fountains at Schönbrunn, designed the Royal Castle at Buda and organised wool manufacturing in southern Hungary.

Owing to his book, "Le Mecanisme de la Parole suivi de la description d'une Machine Parlante" (1791), we have a fairly full record both of the development of von Kempelen's machine and of how he conceived of what he was doing.

As von Kempelen admits, "at the time when I was working on my Chess Player, in the year 1769, I started to examine diverse musical instruments, intending to find the one which came closest to the human voice"* Initially he thought only of "imitating a few vowels with a few instruments", it was only "little by little and much later " that

---

* All translations from von Kempelen by the present author.

he got the idea of the possibility of constructing a speaking machine which pronounced anything. At this time he was attracted most by the woodwind instruments, but already simulation requirements were influencing him; as he said, the woodwinds resembled the human voice most because the action of the reed used in them "resembles a little the functions of the voice box." Indeed, "for a long time," he writes, "especially in France, organs adapted with so-called human voices, had been composed of large and small clarinet reeds, but they imitated only very imperfectly the human voice."

On a country excursion, von Kempelen came across an instrument in the hands of a rustic. Rushing home, he set up an impromptu laboratory in the kitchen. His experiments continued with the aid of a pair of bellows from the stove and various handy pieces of kitchen apparatus. When he found an air leak, due to the orifice of the instrument being too large for the bellows pipe, he availed himself of a "moist bull's bladder" as muffler. Later he added the lower funnel shaped end of a clarinet, which he saw as "representing in some manner the open mouth." He soon had a rudimentary speaking machine in operation. By putting his hand in or across the 'mouth', thus effectively opening or shutting it and acting as a tongue, he obtained "first of all diverse vowels, as I opened by left hand more or less." If, however, he kept his hand still, he seemed to hear an 'A'. From such observations he "soon drew the conclusion that the sounds of a word only become distinct through the proportional relations which exist between them, and that they only obtain their perfect clarity in liaison with each other in whole words and phrases." This astonishingly modern sounding realisation is, even now, not appreciated by all psycholinguists. As Fodor, Bever and Garrett (1974) have had to spell out, the perceived phonetic value of an acoustic signal is heavily dependent on the environment in which the signal appears.

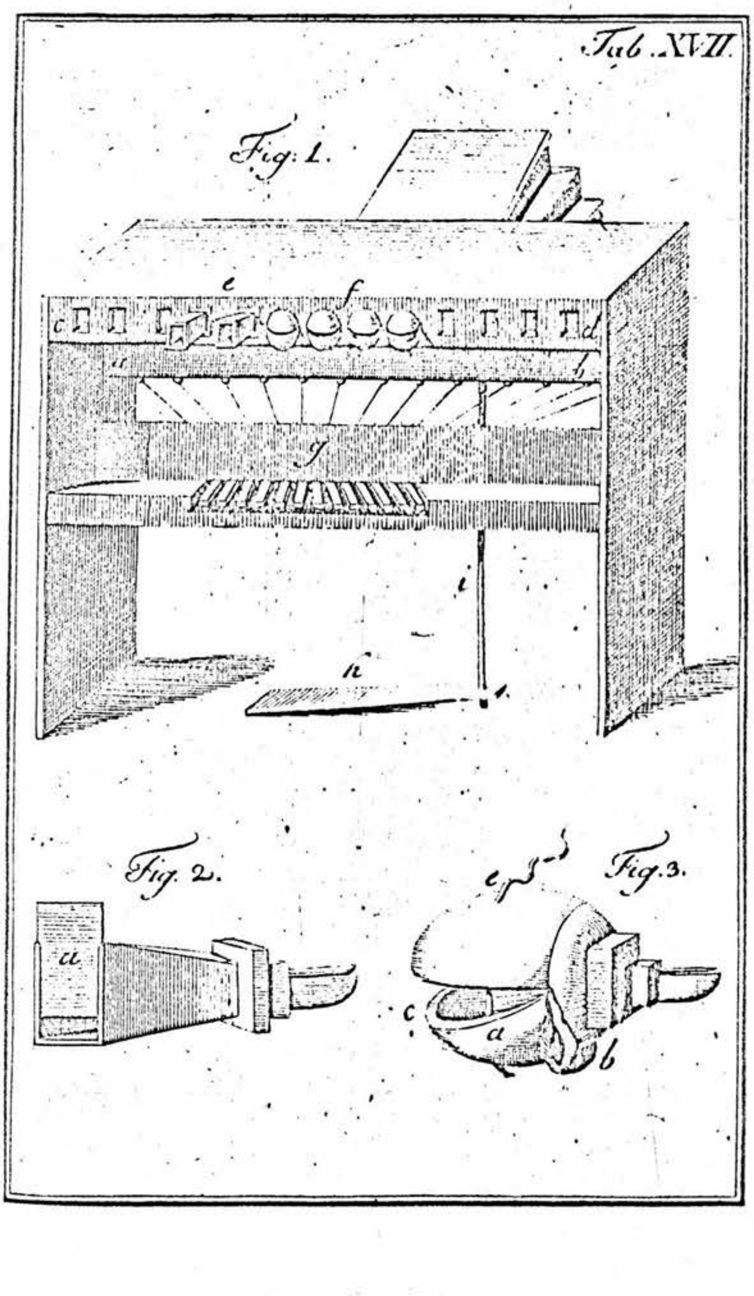
By coincidence, these kitchen dabblings also attained the status of controlled scientific experiment. As an unwitting participant in a Turing test, Frau von Kempelen, who with her children had been banned from the kitchen, and who no doubt was having second (or even third after the Turk!) thoughts about her husband's sanity, thought her husband had a visitor in the kitchen - for she heard a voice "high pitched and zealous", without being able to distinguish which it was.

Modest as these beginnings were, they were "the first formulation on which I later constructed my whole edifice."

We now move on to von Kempelen's second speaking machine. That speech was just a matter of voiced air passing through differing openings, he was now convinced with a "mathematical certainty". He had established that the three essential elements were lungs, glottis and mouth. For months and months he made no progress, kept going only by the certainty that "speech was imitable." Eventually he went to an organ builder to get some organ bellows instead of kitchen ones, and found there a small 'voice organ', which he bought at once, took home and began to adapt.

The most interesting modification from our point of view are the four spheres projecting from the front (see Figure B). Von Kempelen had a lathe operator make the hollow spheres as simulated mouths. They were each bisected into two hemispheres to "represent the two sets of teeth" and the bottom jaw, attached by a leather hinge-bag, was opened and shut by a cord operated by an organ key. Note how closely the machine is simulating a person, but how it is getting to look less and less like one.

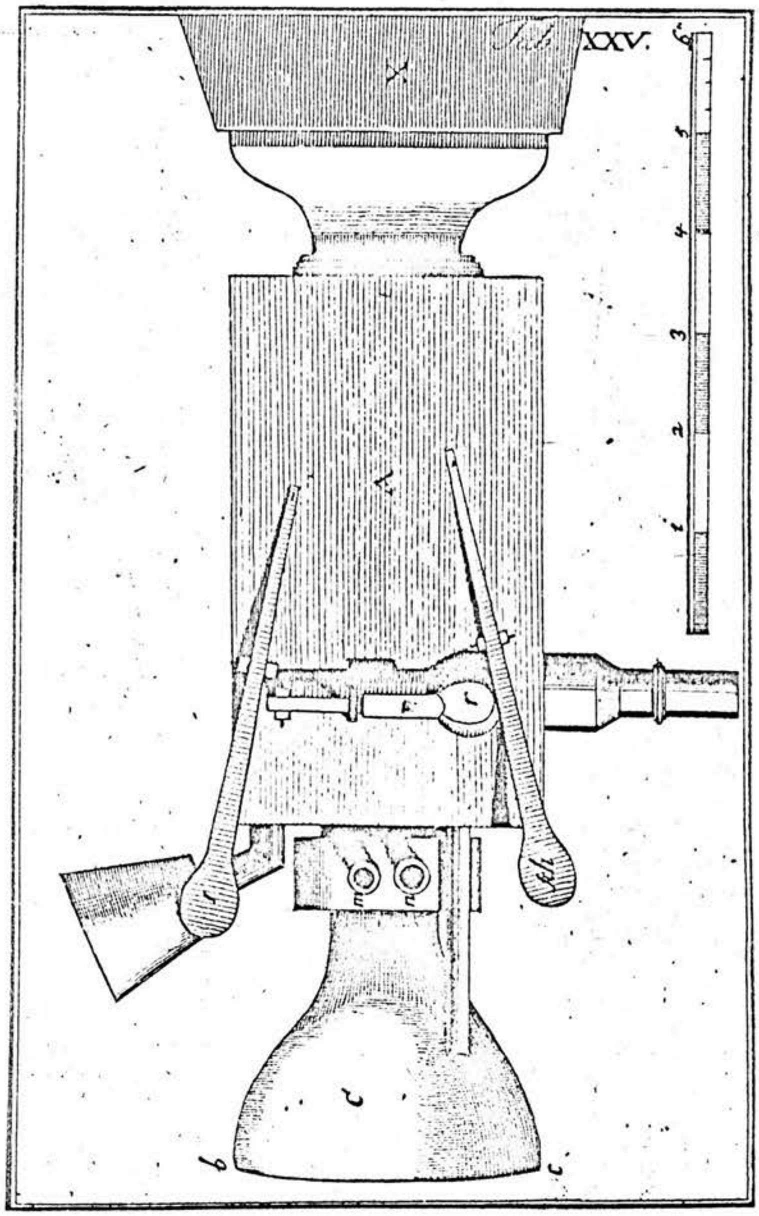
FIGURE B.



However, the machine was not very successful. Von Kempelen obtained the vowels A, O, OU and an imperfect E, but could find no trace of I or U. He fared even worse with consonants; after two years he managed to get P, M and L and this enabled him to pronounce words like 'MAMA' and 'PAPA' by merely pressing keys and pumping with his foot. However, (i) he was obliged to make slight pause between letters otherwise they merged; (ii) he couldn't eliminate a 'k' sound before each vowel, and (iii) he always got an annoying aspiration after each P. Eventually, reluctantly, he decided he was on the wrong track and abandoned altogether his work on the speech organ (sic.). In a heartfelt passage, von Kempelen, who earlier had said "to invent a speaking machine and to implement it in a considered way (was) one of the most difficult projects which could enter a man's spirit", admitted that a vigorous horse would have had difficulty in dragging away a cart loaded with all the bits of machinery he had discarded whilst trying to improve the voice organ.

The third and final von Kempelen speaking machine is illustrated in Figure C. This was a lot smaller than the organ, but much more successful. Its creator "found it necessary to imitate the human organs of speech by having only one mouth (which was, however, flexible) and one glottis." The mouth was a "bell-shaped piece of elastic gum, which approximated, by its physical properties, to the softness and flexibility of the human organs." Connecting to the mouth piece was "a nose made of two tin tubes." When both tubes were open and the mouth piece closed a perfect M was uttered, whilst when one was open and one shut there resulted an N. To make the sound of the machine mellower, von Kempelen lined the canal and inside with dog skin but it was still a fairly shrill and childlike voice. When Goethe (1797) heard it, he

FIGURE C.



commended it as "not very loquacious but it pronounces certain childish words very nicely." Nevertheless its output was, according to its maker, fairly respectable: "I can make it pronounce all Latin, French and Italian words which are proposed to me, some, its true, better than others, but at least several hundreds of words clearly and distinctly, e.g. Marianna, Roma, Maladie, Santé, Astronomie, Chapeau, Racine, Constantinopolis, Missisipi, Vous êtes mon ami, Je vous aime de tout mon coeur" and much more.

The machine was played somewhat like an Hawaiian bagpipe. It was laid flat on a table, the right elbow pumping the bellows, whilst the right hand produced consonants by flaps controlling the stops. The left hand produced the vowels by distorting the space of the mouth piece.

Von Kempelen regretted that "the part of my machine which represented the mouth" - the most important part - was in fact the most imperfect, lacking teeth, tongue and soft palate ... parts whose importance he had stressed earlier in the theoretical section of the book. This shortcoming resulted in an inability in the machine to produce D, G, K or T. However, interestingly, von Kempelen found that by putting a P in their place and rather suddenly withdrawing his hand a little from the mouth piece, he could deceive an audience, who would hear a K or T i.e. ASPRAPAN would be heard as ASTRAKAN! This is again relevant to the point made earlier by Fodor, Bever and Garrett - the perceived phonetic value is context dependent. Von Kempelen seems to have realised this point (through his simulation studies) in 1791.

He was explicit about the isomorphism, and its limits, between the machine and man: "the structure of the machine differs from the man's organ, in this - in the former the nose is blocked from the outside, and in the latter it closes from inside that is to say the palate veils it. But in execution this difference is not remarkable." ... the particular machinery is slightly different but the function is the same.

Von Kempelen is also quite explicit about his aims: "The mechanism in which all the different intonations form is the principle object of this work. We will consider each intonation or letter individually, we will examine the structure, the situation and the movements of each organ which contributes to its formation", and about the means: this machine "is a question of imitating animal organs." He aspired both to "establish a complete system of human speech" and to build a simulating speaking machine, but the two became an integrated concern. He has proceeded by bootstrapping - as Fodor would say - "my speaking machine and my theory of speech have made equal progress, the one serving as a guide for the other."

Around the period von Kempelen was building his third simulation, but unknown to him, Erasmus Darwin, Charles' grandfather, was busily occupied in a similar enterprise. An account of this work was eventually published in 1806, but, as Erasmus Darwin himself tells us there, the actual work itself was accomplished "many years" before.

Darwin's speaking machine was built specifically for the purpose of enquiring in detail into, and improving, shorthand. However, as with so many of the other simulations discussed here, it was intimately connected with its builder's theory (in this case of language). Darwin believed that to develop an adequate shorthand it was necessary to isolate and eliminate all redundancy in spoken English. As a result, his theory of language was designed to analyse the information expressed in utterances, and his speaking machine was to generate a sort of comprehensible acoustic Dalton's weekleese.

As eccentric in his mode of expression as in his interest, Darwin's researches are reported in the form of poetry (The Temple of Nature) and appended explanatory notes. In his view, language was acquired by imitation and words got their meaning by association:

> "Thus the first LANGUAGE, when we frown'd or smiled
> Rose from the cradle, Imitation's child;
> Next to each thought associate sound accords,
> And forms the dulcet symphony of words;"

For Darwin, all words are nouns ("names of things") because they are all "names or symbols of ideas"; ideas in turn, "consist of synchronous motions or configurations of the extremities of the organs of sense;".

In line with his main interests in redundancy, Darwin adopted some ideas of Mr. Horne Tooke, maintaining that language has essentially only two functions: (1) "to communicate our thoughts" and (2) "to do so with despatch". This leads Tooke, and Darwin, to divide all words into two kinds ... those strictly necessary to express our thoughts, and abbreviations of these. For example, they argued that the conjunction "if" was just an abbreviated version of the imperative of the verb "to give".

With this theory in mind, Darwin undertook (sic) an "attempt to investigate the number of the articulate sounds, which constitute those names of ideas by their successions and combinations; and to show by what parts of the organs of speech they are modulated and articulated;". Again he repeats his findings in the form of poetry:

> "The tongue, the lips articulate; the throat
> With soft vibration modulates the note;"

Darwin gives a prosaic expansion in note XV. Vocal sounds are produced by streams of air passing from the lungs through the larynx, wherein an aperture is opened and closed by means of a multitude of tiny muscles. Availing himself of the use of an intimate model, Darwin explains that this process is "something like the trumpet stop of an organ, as may be observed by blowing through the wind pipe of a dead goose."

Darwin's researches on the formation of vowel sounds did not satisfy him. He was concerned to find out in which part of the mouth each vowel was articulated. To do this he "rolled up some tinfoil into cylinders about the size of my fingers; and speaking the vowels separately through them, found by the impressions made on them, in what part of the mouth each ... was formed." The results were, however, inconclusive.

On the articulation of consonants Darwin was more confident, because he had actually simulated the process on a machine. His account is worth quoting in detail. He writes that he had: "contrived a wooden mouth with lips of soft leather, and, with a valve over the back part of it for the nostrils, both which could be quickly opened or closed by the pressure of the fingers, the vocality was given by a silk ribbon about an inch long and a quarter of an inch wide stretched between two bits of smooth wood a little hollowed; so that when a gentle current of air from bellows was blown on the edge of the ribbon, it gave an agreeable tone, as it vibrated between the wooden sides, much like a human voice."

Both Darwin's theory and his machine had, almost literally, much to say about the question of redundancy. Many letters are, they say, redundant, whilst others are wanting. Some simple articulate sounds have two letters to represent them, in other cases two articulate sounds are represented by one letter. Darwin maintained that only thirteen characters (for the letters P, T, K, F, Th, H, plus marks for antesonance, narisonance, orisonance, sibilance, sonisibilance, less and open vocality) are required to sound <u>all</u> European languages.

Within certain limits then, von Kempelen and Darwin, with their speaking machines, extended the mechanisation of the world picture a little further into the human domain. As Darwin put it:

> "Love, pity, war, the shout, the song, the prayer
> Form quick concussions of elastic air."

The St. Petersburg competition really opened the floodgates. In 1815, a Mr. Robertson was reported to have a "wax working figure of a child who could pronounce all the letters of the alphabet ... and say several words." In 1823, Maelzel, who had bought, toured with and repaired von Kempelen's chess automaton, made his own talking doll, which could say "Mama" and Papa", as did Rechsteiner, the repairer of Vaucanson's, and builder of another, mechanical duck. These, though, were all regressions to mere imitation. The next real advance fell to Robert Willis M.A., F.R.S., F.C.S., Fellow of Caius College, Cambridge, who in 1829 published 'the first systematic investigation of the nature of vowel sounds - verified by their synthetic production by models' in the Trans. Camb. Phil. Soc. Willis came to the conclusion that it was the pitch of the resonant note of the mouth cavity which determined which vowel sound was produced and that each vowel sound was characterised by a definite resonant note.

The signs by this time were so encouraging that Brewster, in 1832, felt confident in quoting, in his much read "Letters ...", M. Savant, who wrote that "no doubt that, before another century is completed, a talking and a singing machine will be numbered among the conquests of science."

Indeed about eight years later, a machine which could not only talk and sing but even whisper was going the rounds. Professor Joseph Faber, who had retired from the Vienna observatory because of failing eyesight, had built the Euphonis, worked by bellows and keys. Faber used a "caoutchouc imitation of the larynx, tongue, nostrils, and was able to operate it very skillfully even simulating song" (Chambers Magazine, 1845).

Faber's speaking machine does not actually seem to have made much of a material advance on von Kempelen's in terms of its principles of operation, but it is of interest here, for there is in this case some documentary evidence of the impact of the simulation work on more conventional physiology. Thus, in a supplement to the second volume of Müller's Elements of Physiology (1848), dealing with the mechanism of speech, entitled Recent Advance in the Physiology of Motion and sense, generation and development, Baly and Kirkes admit that "very little new information has been contributed to this department of physiology." They go on, however, to discuss as relevant the Faber artifact, which they describe as "by far the most perfect speaking machine yet invented" and say that by it, the human voice is "very closely imitated."

A fairly full account of the machine, from a physiological point of view, is given in the Entire Medical Science Weekly (1842) by Dr. Eduard Schmalz, a Dresden speech and hearing specialist. According to Schmalz, the machine was misnamed, for, although it spoke German fairly clearly, it could by no stretch of the imagination be described as pleasing. The Euphonia, which looked like a small organ with a doll's torso and head sitting on top of it, had a glottis, mouth, tongue and nose which were "der Natur nachgebildet" (fashioned after nature) in rubber. The machine could sing so well that, at least as regards its German pronunciation, it put some human singers to shame! However, it is the physiological implications in which we are here interested. According to Schmalz, the setting up of the machine is "eine der wichtigsten Fortschritte der neuern Zeit. Nicht nur die Physiologie der Stimme und Sprache, ins besondere die Theorie der Sprachbildung, (wozu auch der Unterricht der Taubstummen in Sprechen, und dieser und der Schwerhörigen im Absehen der Wörte vom Munde gehört,) sondern auch die Pathologie und Therapie der so haüfig bei dem Sprechen vorkommenden Fehler mechanischer Art (namentlich des Lallens, Schnarrens, Lispelns u.s.w. konnen hieraus mäncherlei Nützen ziehen." (Freely translated: the setting up of the machine is one of the most important advances of modern times. Useful applications can be drawn from it, for both the physiology of the voice and speech (regarding the instruction of deaf mutes and lip reading for the hard of hearing) and for the pathology and therapy of more mechanical disabilities like stammers, lisps, nasal accents etc.).

However, it was only two years after Brewster's forecast that a real advance was made. Professor Wheatstone of King's, London, using a very slightly improved von Kempelen machine, discovered multiple resonance: the fact that you can get two or more resonant notes simultaneously from one resonator because the air inside can resonate not only as a whole but in sections at the same time.

Both Willis and Wheatstone were later (1875) supported by Helmholtz, who in his "Sensation of Tone", argued that some vowels like a (calm), ɔ (more), u (who) are characterised by a single resonant note, but others like ae (hat) and e (men) are produced by double resonances, i.e. two separate notes, one produced behind the tongue and one by a constriction of the mid tongue and hard palate. Later (1890-1896) R. J. Lloyd was to complete the picture by showing that <u>each</u> vowel had (at least) two resonances and that the identity of the vowel was not due to absolute pitch of resonance ... but to their mutual interval. Even Lloyd inherited the simulation methodology, however. He writes: "the assignment is fairly certain because it can be confirmed by direct observations in whisper, by the behaviour of imitative cavities, and by careful measurement and calculation."

Some of the careful measurement Lloyd did, was by phonographic technique and it is around this time (actually in 1877) that Edison began developing the phonograph itself. In fact in 1890, as Lloyd began his serious work, Edison marketed some talking dolls - but which of course, talked by phonographic means. At this stage the phonograph, graphophone and later gramophone dominated the speaking market and the commercial days, and thus research funds, were virtually over.

The early attempts at speaking machines were bound to fail, because apart from horrendous mechanical and technological problems, they were based on what Fodor, Bever and Garrett have called "the most primitive model of the relation between acoustic events and the phones that represent them." This model says that each phone has a list of, speaker and context independent, criterial acoustic properties. But, as we have seen, von Kempelen had already realised that this primitive model was inadequate and was making use of context to dupe his audience. Nevertheless, his attempt to generate speech was fundamentally based on a stockpile technique. If this were to have been even moderately successful, it would have entailed staggering storage problems for early simulators. Peterson et al (1958) in a relatively (i.e. not very) successful speech synthesis, used fragments containing two phones mutually influencing each other, chopped from the middle of words, ... but they needed 8,000 segments to synthesise a single idiolect of American English!

Still, from what has been presented, I hope it is at least plausible that:

a) In a welter of activity around the late 18th to early 19th centuries, an empirical start was made on the psychology of speech production/perception.

b) The main method used was that of mechanical simulation, which was

c) breaking free of the immature requirement that the simulation superficially imitates as well as isomorphically models the modelled.

Apart from the desire to explain by simulating, were there any practical applications of speaking machines?

I know of three, if not useful at least interesting cases:

i) Sir Richard Paget, who filed British Patent no. 214281 ... the general method of producing speechlike sounds, in the heady high-speed motoring days of 1923, noticed that pedestrians, who apparently did not notice a rattling engine and blasting horn, would at once perceive a shout ... a noise actually lower in volume. He correspondingly designed a motor car horn which articulated the word "Away!". The first experimental hand-manipulated signal horn, with electro-magnetic larynx, was demonstrated at the British Association meeting at Liverpool on 17th September, 1923. For some reason it did not catch on.

ii) The second brainwave was the cheirophone. This was a variable cavity formed by the hands, the middle three fingers of one hand forming a tongue, the palm of the other hand as the palate and an artificial larynx to be held between the thumb and first finger. An exterior wind supply was required. Apparently, with practice, a degree of expertise can be developed, although N, G and K are impossible to produce. In a British Association meeting in Toronto in 1924, Paget is supposed to have produced by cheirograph the utterances: "Hullo London, are you there?" and "Oh, Leila I love you." One practical application Paget suggested was, when gagged in the dentist's chair, one could, providing one had one's cannister of compressed air with one I suppose, plead "Easy there, you're on the nerve."

iii) The third and final application was that of Casimir Swan, who tried (with little success) to make a <u>talking clock</u>.

Earlier, I mentioned how the topics of these chapters are connecting up ..... but as yet I have no reference to a self-winding talking clock ..... this seems like a good place to wind up this discussion.

# CHAPTER EIGHT.

## A MODERN MECHANICAL SIMULATION.

"What Winograd has done - indeed, what all of artificial intelligence has so far done - is to build a machine that performs certain specific tasks, just as say, 17th century artisans built machines that kept time, fired iron balls over considerable distances, and so forth."

Joseph Weizenbaum (1976).

"There are men who are capable of loving a machine more deeply than they can love a woman. They are among the happiest men on earth ... Men who worry themselves to distraction over the perfecting of a machine are indupitably blessed beyond their kind." (Bennett). As long as there have been such things, machines have extorted a tithe of respect and even love from men. This 'love' is, of course, on a distinctly lofty plane; as Raphael (1977) has emphasised: "I have no intention of ever kissing a computer, or of programming one computer to caress another." It is rather that machinery has an "intellectual, almost spiritual appeal" whose "curious fascination, more than the wish to build something useful or the hope for material rewards ... makes men devote their lives to machinery" (Mayr, 1976).

The popularity of the opinion that machines are instructively analogous to human beings has, of course, had its ups and downs, depending on the relative states of advancement of the sciences of physiology and machine theory. For example, from the vantage point of fifty years ago, one pair of psychologists (Sturt and Ogden, 1926) felt justified in writing: "In the past it was agreed that our body was a beautiful machine so perfect that it not only ran, but repaired itself, looked out for jolts on the road. Now scientists are abandoning that hypothesis, and the more eminent the scientist the more completely he seems to have abandoned it. The theory was, indeed, the product of partial ignorance. When physiology knew less of the phenomena it studied, a mechanical interpretation was fairly easy; now that knowledge is greater, and wonder after wonder, fine adaptation after fine adaptation is revealed, it becomes ever more necessary to postulate some power beyond the machine." Today, however, some psychologists are hailing the "emergence of a new (and not yet well understood) notion of mechanism" (Pylyshyn, 1976) in the form of the electronic computer, which in its power and flexibility, not only makes the earlier machines seem impoverished but makes many of the traditional problems of philosophy and psychology take on a "renewed interest and vigour" and provides an alternative language for psychological theories - mechanistic but irreducible to physics and chemistry. One sympathiser (Boden, 1977) with this approach is even moved to write, perhaps rather optimistically: "The new concept of 'machine' provided by A.I. is so much more powerful than familiar mechanism that the old metaphysical puzzle of how mind and body can possibly be related is largely resolved."

Earlier, we have discussed some 17th and 18th century machines and discovered that they were indeed, contra the received impression, very sophisticated attempts to explain by simulation and, as Diamond (1974) has noted, "a careful reading of history will show that 20th century theory is often following 18th century models." What, however, of the intellectual descendants of these early automaton-simulators? Weizenbaum (1970), a practitioner-turned-fierce-critic of the A.I. enterprise, leaves no doubt about his opinion: "Newell, Simon, Schank, and Winograd simply mistake the nature of the problems they believe themselves to be 'solving' ... as if they were benighted artisans of the 17th century." As it has been a major burden of the present study to argue that the 'artisans' concerned were anything but benighted, we cannot but take exception to Weizenbaum's slander, however, his main assertion - that Winograd and the 17th century artisans are engaged in essentially the same enterprise - can be accepted provisionally at least.

I wrote a little while back that modern electric computers make the earlier machines seem impoverished. The qualification ("seem") was necessary because strictly speaking any machine (be it electronic, mechanical or hydraulic) which moves through definite discrete states is a Turing machine ... and in principle any Turing Machine is as powerful as any other (given no limitation of time or space). The advantage of digital computers is that, via the convenience of programming, they can mimic any discrete state machine. As Turing (1963) has put it, digital computers are "universal machines ... considerations of speed apart, it is unnecessary to design various new machines to do various computing processes. They can all be done with one digital computer, suitably programmed for each case."

It surely hardly needs restating that digital computers (despite the name) are not mere arithmetic wizards. They can crunch numbers - where the syntax their processes are programmed to instantiate are the rules of arithmetic and the abstract symbols they manipulate stand for whatever arithmetic primitives (numbers) stand for (say classes of objects) - but that is a special case. Digital computers operate with abstract symbols which can stand for anything and with processes which can mimic any process. "A computer is not merely a number manipulating device; it is a symbol manipulating device." (Newell and Simon.) The extension of this idea to the functioning of the human being leads directly to the project of simulation by models which preserve symbol manipulating capacity: "We can postulate that the processes going on inside the subject's skin - involving sensory organs, neural tissues and muscular movements controlled by the neural signals - are also symbol manipulating processes, that is, patterns in various encodings can be detected, recorded, transmitted, stored, copied, and so on, by the mechanisms of this system." One A.I. worker even defines A.I. as "the study of complex information processing problems that often have their roots in some aspect of biological information processing." (Marr, 1977).

One symbol manipulation task in particular, natural language use, has concerned psychologists, philosophers, logicians, A.I. workers and linguists and it has, not infrequently, been posited as that which distinguishes people from both 'lower' animals and machines. When the typical 'lower' animals was taken to be a bee the former part of this claim seemed fairly strong (von Frisch and J. Bennett), but recent studies with apes (Premack) are less supportive. However, this was never seriously a logical claim but an empirical one. There was no question that, a priori, animals could not talk, only that there were in fact no animals which could talk. As regards machines, philosophers in particular have been keen to ridicule the idea of a machine which could understand natural language.

There have been several attempts by A.I. workers to program computers to understand natural language e.g. Colby's PARRY and Weizenbaum's ELIZA, but these are mere linguistic imbeciles compared to the relatively Joycean stature of Winograd's (1972) SHRDLU. In the remainder of this chapter we shall look at SHRDLU's good points and failings, as simulating psychologists rather than technologists, and see to what extent SHRDLU is, or is not, an adequate simulation.

To begin with we should, if not answer at least rule out of court, one objection - that what such a program does can by no stretch of the imagination be called language understanding but, at best, language receipt and response. The point of such objections is not entirely clear. The objection cannot be that what SHRDLU does, does not count as understanding because it does not do what people do <u>in the same way</u>, because what it is that people do is precisely what the model was built to find out i.e. how people do understand is a mystery, so it cannot be through comparing this with the way SHRDLU works that objectors reach their conclusion. In any case, most such objectors are prepared to deny that what SHRDLU does is understanding before, or even without, knowing at all how it works.

It has been said that up to the advent of locomotives, people tended to speak of men as machines (machinomorphism), whereas, since that advent they have tended to speak of machines as men (anthropomorphism). Some claim that the attribution to a machine of understanding is a bad case of anthropomorphism. Gauld and Shotter (1977), for example, claim that "the issue of whether or not machines ... could possess and exhibit ... understanding ... resolves itself into the issue of whether or not it is in principle possible to design a machine with concepts as opposed to blindly following some algorithmic routine, on which someone else could put an interpretation." However, apart from some vague phenomenological intuitions and introspections, it is not clear what having a concept is, even for people - certainly not so clear that we could categorically deny the attribution of concepts to machines. One useful way to think of concepts is as functional components of a conceptual processing system, and this clearly makes the question of concepts secondary to the question of the attribution of understanding. Moreover, a particular advantage of Winograd's approach is that by using procedural rather than assertional representation of knowledge, it offers an alternative conception of what is going on to the wholesale reification typical of mainstream psychology, which insists on postulating entities (concept, idea, mind, thought, intention etc.). Just as many apparent 'entities' in the physical sciences have turned out to be processes e.g. a 'desk' is a continually changing fluctuation of molecules, many concepts of psychology may turn out to refer to processes or procedures, rather than things.

It is frequently extremely useful to attribute propositional attitudes to programs. As Dennett (1976) has remarked, it may be very "useful and predictive" to say of a particular program: "It thinks it should get its Queen out early", where we are describing "features of the program that are, in an entirely innocent way, emergent properties of the computational processes that have engineering reality." Moreover, as Boden points out, "it is not clear that what programs do could be expressed as well - or even at all - without borrowing from everyday psychological language. And the more complex and flexible the program, the more this is so." (providing "the term chosen is carefully selected with the precise functional details of the program in mind"), nor, one might add, is it clear what would be achieved by refusing to so borrow.

Instead of continuing this wrangle, let us pass it by to consider how SHRDLU does what it does. But first, I will give the briefest of outlines of just what it is it does do.

Basically, SHRDLU is a simulated robot who lives, with a single hand, in a world of toy blocks (cubes, pyramids, boxes etc.). The system accepts information and executes commands which are typed into it in natural English. It manipulates the toy bricks according to the commands (this is simultaneously simulated visually on an accompanying display package) and answers questions in natural English, again of course via the teletyper, taking between five and twenty seconds to analyse and answer questions or obey commands. SHRDLU also has a model of its own 'mentality' and, in addition to merely carrying out its plans, can discuss what it has done and intends to do.

Example of dialogue:

Find a block which is taller than the one you are holding and put it into the box.

The system has to figure out what is meant by "one" and "it", as well as using deduction to decide which block it is holding.

BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.

Printed out to keep us informed.

OK.

The hand has to set down the red block to pick up the taller blue one.

What does the box contain?

THE BLUE PYRAMID AND THE BLUE BLOCK.

What is the pyramid supported by?

Earlier the system responded to "the pyramid" by saying "I don't understand which pyramid you mean." Now, it assumes that we mean the one just mentioned.

THE BOX.

SHRDLU can fruitfully be seen as a collection of separate but interacting programs, the main ones of which may briefly be described as follows: INPUT which takes an English sentence and by using a dictionary and carrying out morphemic analyses feeds an acceptable string to the GRAMMAR and SEMANTICS which work together with PLANNER (a deductive system which both directs the parser and deduces facts about the world of blocks) to interpret sentences; PROGRAMMER is a parsing system which builds and explores parsing trees; BLOCKS a collection of theorems containing knowledge of the properties of the blocks world. In addition, there is a DICTIONARY containing syntactic and semantic definitions (in the form of mini-programs), a network of SEMANTIC FEATURES and a store of DATA. The whole system is called by a MONITOR but then functions autonomously until the understanding process is complete, when the MONITOR is called again. In between time, communication and control passes between the various subprograms as necessary to achieve understanding.

To give a very approximate flavour of how SHRDLU would procede, consider the command: "put the blue pyramid on the block in the box."

When this is typed in, the MONITOR calls INPUT which, by using the DICTIONARY and SEMANTIC FEATURES starts to send into GRAMMAR an acceptable string ('put' is in the infinitive, so does not need morphemic change, whereas e.g. 'putting' would be sent in as put + ing (infinitive + present participle)). The GRAMMAR and SEMANTICS directed by PLANNER would interpret this as an imperative and start looking for a noun group and a preposition group. But how is SHRDLU to procede? There are two possible parsings by PROGRAMMER: Put (the blue pyramid on the block) in the box + Put the blue pyramid on (the block in the box) ... both yielding noun group and prepositional group. SHRDLU continues by bringing in BLOCKS and PLANNER, which say that in the current state of the world there is no blue pyramid on a block ... consequently this interpretation is ruled out.

It is important to realise that the knowledge of the world, of grammar and of semantics interact and co-operate continuously in order to bring about an understanding ... the results of semantic interpretation guiding parsing and vice versa. As Winograd points out, linguists have suggested (and this seems intuitively plausible) that people make use of both general and specific knowledge in understanding language. It is surely our knowledge of the world which prevents us from interpreting "he gave the boy plants to water" in the same way as we would "he gave the house plants to charity" i.e. our knowledge that there are no such things as 'boy plants'.

There are three facts to note in particular. Firstly, the fact that SHRDLU has a world at all to talk about and against which to check possible interpretations, is in fact a major advance beyond other natural language programs. Weizenbaum (1966) for example, chose to simulate non-directive therapy discourse precisely because it is "one of the few examples of natural language conversation in which one of the participating pair is free to assume the pose of knowing almost nothing of the real world."

Secondly, the control of SHRDLU is distributed and passes throughout the system (it is heterarchical) rather than being rigidly imposed from above (hierarchical). It is not the case that all the syntactic analysis is done, then the semantic, then the deductive rather the process of interpretation is heuristic ... only as much computation is done as is necessary and in the direction necessary. As Winograd puts it: "the way of treating ambiguity is not through listing all possible interpretations of a sentence, but in being intelligent in looking for the first one, and being even more intelligent in looking for the next one if that fails."

Thirdly, the form of representation is procedural rather than assertional. Theorems and definitions are actually written as mini programs. The assertion that all dogs bark ((x)(dog x $\rightarrow$ bark x)) would be written instead of putting two procedural items into the program's knowledge base:

1. If trying to show something barks, search data base for information that it is a dog.
2. If you add to data base an item that something is a dog, also add

an item that it barks.

This innovation, taken from Hewitt, has important implications for implementation - if a theorem is written in the form of a procedure, steps of the procedure can actually be internal actions of the robot (or person). Incidentally, this may be an important way of characterising Fodor's (1975) 'language of thought'; procedural representation, which he doesn't discuss, is a natural way to "devise some mapping which pairs physical states of the device (computer, brain) with formulae in a computing language in such a way as to preserve desired semantic relations among the formulae." In Winograd's SHRDLU "knowledge is expressed as procedures written in special languages ... These languages have the control structure of a programming language, with the statements of the language explicitly controlling the process. The steps the system takes in understanding a sentence can be determined directly by special knowledge."

Winograd obviously feels that the type of representation used in SHRDLU is relevant to the question of how people understand. As he admits: "in some theoretical sense predicate calculus sentences could express all our knowledge, but in a practical sense there is something missing. A person would also have knowledge about how to go about doing the deduction." Some psychologists too, are coming round to the view that what we need are "the appropriate concepts to summarise blocks of similar operations conducted by (the brain) and an appropriate language." (Sutherland, 1974). As Mackay (1974) puts it there is "much better hope of spotting correlates (of subjective experience) if we focus on terms for events and activities not static states." Certainly, as mentioned earlier, the tendency for psychologists to reify (thoughts, minds, intentions, meanings) has done little but petrify the field of enquiry.

This said in eulogy of SHRDLU, it has, of course, many failings. Firstly, its capacity to understand English is far greater than its capacity to 'talk' - but then it does only set out to be a natural language understand<u>er</u>. However, even within its range of application, due to its intrinsic (reasoning and linguistic) limitations, it is unable to deal with hypotheticals, paraphrase or handle garbled grammar or logic (where a person would extract some sense). In view of SHRDLU's massive achievements, it perhaps seems a little ungenerous to carp at these limitations, but they do raise the real problem of whether these limitations are ones of the particular program, which could be extended without any difficulty of principle, or whether they are intrinsic to the very way of going about solving the problems.

This is well brought out by consideration of the implications of the limited domain (world) in which SHRDLU lives. Whilst it is true, as Winograd says, that "the language programs do not depend on any special subject matter" - any 'world' could be substituted for the blocks world - they <u>do</u> depend on some limited domain of subject matter in each case. In particular, SHRDLU must live in a closed world, in which everything which is the case is entailed by the known data.

e.g. can a pyramid by supported by a block?

    The deductive system finds an actual example, so it knows this is possible.

YES.

(Indeed, all SHRDLU's processes are rigidly logical - it even has to prove its own hand is not already holding anything before it can pick something up).

This consideration lends some plausibility to the claims of critics that "Winograd's work is not directed to finding out how human languages are organised by theories of explanatory power - but rather how language might be organised for the purposes of a machine performing particular tasks in a limited domain." (Drescher and Hornstein, 1976).

These critics argue that, contrary to Winograd's claim that what he aims for is a "better understanding of what language is", "the true goals for this work are primarily technological rather than scientific." To be more brutally precise: "his interest remains to design a model which will work reasonably efficiently for the domain he has chosen without going into any of the theoretical or empirical considerations which are of primary interest for a scientific theory of language or intelligence."

What are the conditions on scientific adequacy which SHRDLU allegedly does not meet? There are three levels of adequacy: (i) a model is observationally adequate if it can "reproduce observed facts about human processing" e.g. can process the set of sentences processable by people (many different processes may satisfy this level); (ii) a model is descriptively adequate if it "incorporates just those principles of processing that speakers use"; and (iii) a model is theoretically adequate if it providesexplanatory principles from which such a model will necessarily follow, i.e. it must possess some theoretical structure which prohibits purely ad hoc adjustments to bring about observational adequacy.

Obviously if you are motivated purely by a "practical desire to have a usable language-understanding system", descriptive and explanatory adequacy are not your concern ... though presumably observational adequacy is still required (Pylyshyn (1976) claims that "any A.I. system is at some level a psychological theory simply because the description of the intelligent task to which it is addressed already is essentially a description of some psychological process.") Equally obviously, the satisfaction of observational adequacy does not guarantee the satisfaction of either of the other levels. Consequently in a weak sense of 'simulate': "the existence of a machine that can simulate certain aspects of human intelligence will not necessarily contribute to the development of a theory of human thinking." (Drescher and Hornstein). Some A.I. researchers do indeed disregard the higher levels of adequacy. Thus Raphael has written that "being able to solve a problem, even though no one may understand precisely how it works" is a more important aspect of a program than its being "explainable in terms of some deep underlying theory" (Boden 1977b); and Wilks denies A.I. is theoretical or scientific work at all, but rather an engineering activity. But then Raphael and Wilks do not see A.I. as any sort of simulation! Researchers who do, including Winograd, are bound to strive for higher levels of adequacy.

There are two particularly strong lines of argument against SHRDLU being theoretically adequate: the apparent lack of general principles and allegedly arbitrary 'patching'. In order to do more than "mimic in an unenlightening way some small aspect of human performance" (Marr, 1977) a theorist must formulate principles which underlie each of the components and their interaction. Drescher and Hornstein again are the least compromising: most purported computer simulations of natural language performance are "of virtually no psychological - as opposed to technological - interest because it is totally devoid of any principles which could serve as even a basis for a serious scientific theory of human linguistic behaviour" (Drescher and Hornstein). Weizenbaum (1976) reiterates this criticism of Winograd and (of course) extends it to our earlier automaton builders: (Winograd's) heuristics express no interesting general principles ..."such principles cannot be discovered merely by expanding the range of a system ... Even the most clever clock builder of the 17th century would never have discovered Newton's laws simply by building ever fancier and more intricate clocks."

This criticism is partly deserved. I am not convinced that SHRDLU does express general principles relevant to all natural language understanding; partly because of its being designed for operating in a limited domain and, as we shall see, certain aspects are open to the charge of arbitrariness. However, Weizenbaum's wider point is not well taken, as he chooses to ignore the analogical aspect of simulation. Of course, due to referential opacity, it might be true that a 17th century clock builder in constructing a timepiece is actually constructing a model of the planetary system, whilst he might not know he is building one. He might understand the way the clock works but not understand the way the planetary system works ... though they work in the same way. This is because 'understanding' does, but 'constructing' does not, set up a referentially opaque context. (Similarly, the police may be looking for the murderer, but not looking for Joe Bloggs, although Joe Bloggs is the murderer). This brings out that which is undeniable in Weizenbaum's assertion that we couldn't discover the laws of planetary motion just by building clocks, although clocks may instantiate those very laws. However, as has been argued elsewhere, truth of a system as a model of another system is not a "local" property of the model, but a global property of the pair of systems. It is by using the clock as a model of the planetary system (or the brain) that we discover laws of the modelled. Of course, just building "fancier and more intricate clocks" in itself will not help us understand the world, but using the clocks to develop "fancier and more intricate" hypotheses about the world might well! Of course, not every 17th century clock builder was a simulator of the natural world, but it is merely committing the sin of affirmation of the consequent to infer that every simulator was not a clock builder, and it is just false to say that one could not discover laws (which govern the movement of the planets) by simulating the planetary system by a clock.

Another major problem for Winograd is the accusation of patching. The basic notion behind this is easy to grasp. Suppose one orders a suit from the tailor, but instead of a well-fitting carefully-made garment with neat joints, a roughly cut approximation arrives, which the tailor takes in and patches in an arbitrary way to accommodate one's shape. By application of enough patches, he may actually keep the wind out - but you would hardly be satisfied with the end result. It may even be that the structure of the finished suit is governed by the size and shape of the patches available; certainly it will be responsive to a desired endstate (a practical result) rather than guided by 'principles of tailoring'.

It has been said of Winograd that his approach is to "arbitrarily stipulate what are in reality matters than can only be decided by empirical research and what can only be explained on the basis of theoretical work" (Drescher and Hornstein). In his semantic component, Winograd uses Katz-Fodor type semantic markers, but rather than deciding in a principled way which features are basic, and because SHRDLU lives in a closed world, he simply stipulates all the features necessary to converse about the world of blocks. As another example, Winograd writes that his parser is one which operates top-down-left-to-right, but that it "modifies these properties when it is advantageous to do so" (Winograd); this latter qualification smacks of mere ad hocery and anyway, Winograd gives no reasons why his parser should be top-down-left-to-right, again this is merely stipulated without any explanation or justification - one is tempted to believe that its only justification is that it works!

What was Winograd's aim in building SHRDLU? Was it merely to get a 'natural language understanding system' which works, or to simulate the human psychological process?

Winograd states three goals of his research: a practical desire for a natural language understanding system; a better understanding of language and a better understanding of intelligence. The first two of these aims may obviously be pursued with virtually no regard for the psychological reality of the processes postulated, but, expanding on the last aim, Winograd expresses the conviction that understanding language "may lead to a better theory of how our mind works." But, is the way he chooses to gain this better theory of our mind, that of simulating psychologically real processes? In the text, Winograd seems to support this interpretation. Thus he writes: "When a person sees or hears a sentence he ... To model this language understanding in a computer, we need a program which ... ", where the blank in each case is filled in by the same requirements. Elsewhere, he writes: "our notation ... is intentionally general, so that our system can deal with concepts as people do." However the editorial, in the edition of Cognitive Psychology in which U.N.L. first appeared, is quite explicit that "Winograd's system is not a 'simulation' but it incorporates important ideas about human syntactic, semantic and problem solving abilities, and, in particular, about their interactions in understanding natural language". When we look closely at U.N.L., we can see that, whilst Winograd seems to be strongly simulating psychological processes, often he retreats from that task. Thus on page 26 he writes: "language is a process of communication between people ... (and it) ... is enmeshed in knowledge of the world." This knowlédge is, we are told, "a collection of concepts designed to manipulate ideas." Thus far, then, it seems that insofar as people communicate linguistically, they make use of concepts. This seems to be confirmed by the claim further down the page that the "meaning of words is represented by ... meanings of 'concepts' which exist in speaker's and hearer's minds." However, even further down the page, we are told "the 'concept' representation of meaning is not intended as a direct picture of something which exists in a person's mind. It is a function that gives us a way to make sense of data, and to predict actual behaviour." Thus we have been led gently from process as psychologically real simulation to process as explanatory construct to systematise data. Winograd's claim that "The justification for our particular use of concepts in this system is that it is thereby enabled to engage in dialogs that simulate in many ways the behaviour of a human language user", then, can only be taken as seriously claiming observational adequacy for his simulation, despite an earlier apparently stronger claim.

It is my opinion that Winograd does not really hope for more than this. Elsewhere he writes: "It is not yet clear what connections (our models) have with the processes going on in the human mind. Yet they give us a clear framework for thinking about what it is we do when we understand and respond to natural language." This 'clear framework' is the "formal metaphor", given by computers and computer language, "within which we can model the processes and test implications of our theories;" "to write a program we need to make all of our knowledge explicit ... this provides a rigorous test for linguistic theories, and leads us into making new theories to fill the places where the old ones are lacking." Winograd, then does aim to make SHRDLU observationally adequate and he has certain views about the language understanding process, which he attempts to make more formal through the discipline of computer modelling. It is, as Wilks has described other A.I. work, a particularly brilliant "expression in some more or less agreeable semi-formalism of intuitive common sense knowledge revealed by introspection", supplemented in Winograd's case by extensive linguistic research.

Weizenbaum criticised both the '17th century artisans' and Winograd for letting "their successes lead them to the conclusion that they had begun to approach a general theoretical understanding of the universe, or even to the conclusion that, because their machines worked, they had validated the idea that the laws of the universe are formalisable in mathematical terms." However, failings and limitations admitted, it is my conviction that these automata do indeed point a way to a new kind of explanation, mechanistic in principle but irreducible to physics and chemistry. The demonstration that processes usually considered essentially and inalienably human in nature can be imitated by a machine, must cast doubt on the conviction that the human brain is a totally unique and even magical entity.

# CHAPTER NINE

## SHALL I COMPARE THEE TO A SUMMER'S DAY?

"The general design of the following sheets is to inlist Imagination under the banner of Science; and to lead her votaries from the looser analogies, which dress out the imagery of poetry, to the stricter ones, which form the ratiocination of philosophy."

Erásmus Darwin (1806).

In the four chapters immediately preceding this one, we have discussed various formal models (ones intentionally constructed to match, in some relevant respect, the structure and/or function of some system the modeller wants to understand). In this chapter, and the following two, we will discuss some intimate models which, the reader will remember, are pre-existing entities which are seen by the modeller to be similar, in some relevant respect, to something he or she wishes to understand.

As a rough approximation, it is true to say that no entity or phenomenon is so lowly or ludicrous that it cannot be used - or indeed that has not been used - as a psychological model.

For example, consider the following short story:
"Bill woke up one Sunday to the sound of the church bell and decided to spend the day in the country. He telephoned the station to find out the time of the next train and soon he was in the countryside admiring the scenery, flowers and animals. He returned home that evening, intending to go to the cinema, but found a telegram inviting him to a concert. He cooked and ate, whilst listening to records, and went out. When he returned it was late. He lit a lamp, covering the budgie's cage with a cloth, and read for an hour before seeing to his toilet and retiring to bed."

Most psychologists, if asked to pick out from the story phenomena which have actually been used by psychologists to help understand/explain psychological processes, would have little difficulty in picking out three or four, but it may come as a surprise to hear that there are in fact (at least) fourteen such models directly mentioned above.

In this chapter, I intend to draw out and discuss some of these models at some length (and I will mention the others) in partial support of the claim that:

> "The mind of a Newton, a Bach or a Michelangelo, it was said, differed only in complexity from a printing press, a whistle or a steam engine." (Jeans, 1937).

Church Bell.

"The bells of St. Mary's over the way always ring for a quarter of an hour before the University sermon; yet the ringing of the bells is not the cause of the sermon." (Romanes, 1885). Unlikely as it may seem at first, Romanes is indulging here in (philosophico-)psychological theorising. That it is theory in which he is interested is shown by his admitting that "although, as long as the association remains constant, there would be no harm in assuming, for any practical purposes, that it is so", we might in fact be wrong. What parallel is Romanes drawing here between the vibration of the bells in St. Mary's church and the sermon, though, in the psychological case? Well, he is leaving the sermon constant, but reading "vibration of a number of little nerve-cells in the brain" for that of the bells. Inference? "We may be similarly wrong if we were definitely to conclude that the sermon is produced by the vibration of a number of little nerve cells in the brain of the preacher."

To what interesting psychological question is this model relevant? ... That of materialism. Romanes admits: "materialism ... is ... at once the simplest physiological explanation of facts already known and the best working hypothesis to guide us in our further researches. But it does not follow from this that the theory of materialism is true."

This, then, is one typical scenario (one leading to a negative conclusion) for a piece of intrinsic model use. There is an interesting theoretical problem; some analogy of structure or function is drawn between a phenomenon one wishes to understand and one one understands relatively better (we might have believed that the bells caused the sermon "if we did not happen to know so much about the matter (sic) as we do"); then a line of argument rehearsed with the understood system (the model) is extended to the previously not-understood one and conclusions drawn.

Telephone.

The model of the brain as a telephone exchange or switchboard is, of course, an old favourite. Craik (No. 27, 1966) wrote that "... a system of neural pathways in which the number of fibres, rather than the particular identity of a single fibre is important ... has its analogy in a ... telephone system" and, slightly more recently, Miller (1956) pursues the same analogy. The classic statement of the metaphor, though, is that of Pearson (1892):

We must compare "the brain to the central office of a telephone exchange, from which wires radiate to the subscribers A, B, C, D, E, F, etc. who are senders, and to W, X, Y, Z, etc. who are receivers of messages." Which psychological function is Pearson modelling?...The connection between sensory and motor nerves and the measure of consciousness of impulses. Lines from A etc. model sensory nerves, lines from Z etc. model motor nerves and consciousness is modelled by the operator or 'clerk'. Certain analogies are manifest at once, wired into the model, so to speak. Take, for example, "instinctive exertion following unconsciously on a sense impression." Well, this can be understood by comparing it to the situation where A has previously notified the company that he will only ever want to speak to W, so that A's wire is joined permanently to that of W and "the clerk remains unconscious of the arrival of the message from A and its despatch to W, although it passes through his office." Only slightly more complex is "habitual exertion following unconsciously on a sense impression." Here the operator "finds by experience that B invariably desires to correspond with X, and consequently, whenever he hears B's call-bell he links him mechanically to X." "An exertion following consciously on the receipt of a sense impression" is modelled thus: C, D, etc. "set their bells ringing for a variety of purposes," the operator has to listen to their "special communications" and perhaps employ additional resources (directories etc.) before he "shunts their wires so as to bring them into circuit with those of Y and Z, ... to best suit the nature of the demands."

There is no denying the attractiveness of this model. There is even a (gross) physical resemblance between nerves and telephone wires and both function in systems of communication. (In a more technologically naive world, Descartes (1733) had pictured the afferent nerves as hollow tubes, like the bell ropes used to summon servants). More particularly, both motor and sensory nerves and caller and receiver are (intuitively) distinguishable by the direction of travel of the relevant impulse. Moreover, "in all cases the activity of the exchange arises from the receipt of a message from one of a possibly great but still finite number of senders" and "the originality of the clerk is confined to immediately following their behests or to satisfying their demands to the best of his ability by the information stored in his office."

Once the model is stated, it is possible to extend it; for example, in the case of instinctive unconscious exertion after a sense impression, we supposed the message to pass through the clerk's office. However, as was well known even in Pearson's day, certain instinctive exertion following sensory stimulation still occurs even though the central telephone exchange has been demolished (as in the case of a frog whose brain has been removed, but which will still remove an irritant from its body). Extension of the model ... "If these wires were connected outside the office, we should have an analogy to certain possibilities of reflex action, which arises from sensory and motor nerves being linked before reaching the brain" (for example in the spinal cord).

Nevertheless, of course, the brain is not a central telephone exchange; as Pearson saw ... "the analogy must not be pushed too far." The reality is far more complex than the model suggests. Pearson was only too aware of (some of) the failings, for example, "senders and receivers must be considered distinct, for sensory and motor nerves do not appear to interchange functions."

The eclipse of the telephone exchange as a model of the brain provides another interesting object lesson. (Intrinsic) models fall from favour when they have been superseded by other models, usually suggested by some technological advance, rather than by their own intrinsic limitations, which are anyway often patched by ad hoc augmentations. Technological advance is as least as important as intrinsic developments of psychological theory in the determination of which models are adopted. Thus Beloff (1973) was correct both in stating that "man has a brain and ... this brain is more than a switchboard ... What exactly that something is, forms the real point of departure for the cognitive psychologist.", and in pointing to the (approximation towards an) answer cognitive modellers are giving: "the brain should not be conceived either as a vast telephone exchange of reflex arcs or as a vaguely defined field of interacting forces ... it must be thought of as a computer receiving inputs from many sources and combining them to produce an output which is unique to each particular occasion though lawful"

(Welford on Craik in Beloff). In time, of course, the computer itself will be superseded by some other model suggested by the developing technology.

Trains.

Railway systems and trains themselves have also been popular models with psychologists virtually as long as there have been such things (trains that is!), but, of course, the analogy can be read in many ways ... and often is, even by the same author in the same book. William James (1891) at the end of chapter 5 of the Principles, uses the train model to illustrate the deterministic action of a machine which, unlike a person (for James), "acts fatally in one way", "knows nothing of wrong or right" and "has not ideals to pursue." The style of the argument goes as follows ... concede a point to the opposition for the sake of argument, ... suppose we were machines, then we would be, in relevant respects, just like a train, but "a locomotive will carry its train through an open drawbridge as cheerfully as to any other destination", but we would not do that, therefore we are not trains, and therefore a fort. not machines.

James' other use of the model is in connection with the identity theory of mind: "however numerous and delicately differentiated the train of ideas may be, the train of brain events that runs alongside of it must in both respects be exactly its match." The definitive statement of the use of this model must be that of Clifford (in James), whose presentation would seem like a parody were it not so earnestly put: "The train of physical facts between the stimulus sent into the eye ... and the exertion which follows it, and the train of physical facts which goes on in the brain ... these are perfectly complete physical trains ... the two things are on utterly different platforms -

the physical facts go along by themselves, and the mental facts go along by themselves. There is a parallelism between them, but there is no interference of one with the other." "If anybody says that the will influences matter, the statement is not untrue, but it is nonsense" says Clifford positivistically. In support of this he again uses the locomotive analogy and, as a thought experiment, asks us to "imagine a train, the forepart of which is an engine and three carriages linked with iron couplings, and the hind part three other carriages linked with iron couplings: the bond between the two parts being made up out of the sentiments of amity subsisting between the stoker and the guard." In the case of both organic bodies and iron carriages "the only thing which influences matter is the position (or motion) of surrounding matter." Huxley (1898), in a famous passage, was really illustrating the same point (non-action of mind on matter), though from an epiphenomenalist standpoint: "The consciousness of brutes would appear to be related to the mechanism of their body simply as a collateral product of its working, and to be completely without any power of modifying that working as the steam-whistle which accompanies the work of a locomotive engine is without influence on its machinery."

Craik, on the other hand, uses the analogy in two quite different ways. Firstly, he is concerned to understand (in the context of the localisation of function controversy) how it can be that injury to a particular region of the brain does not necessarily destroy a particular function. To Craik "this suggests a system of neural pathways in which the number of fibres, rather than the particular identity of a single fibre, is important." This, plus the fact that his theorising was done during the second World War, led Craik to compare the effect of damaging the neural system of the brain with that of bombing a railway system;

where, that is, railway lines model neural pathways and trains nervous impulses. Now the problem ... how would a railway system have to be organised, so that minimal disruption to the system as a whole occurred upon specific damage? ..."If there is only one railway ... to a particular place, a bomb on it will have a definite, localised effect - it will completely hold up communication with a particular place; but it will be otherwise if there are many alternative routes ...(then there will be) only a slight decrease in general efficiency and rapidity of communication."

Craik (1943) also uses the railway system as a model to help understand how it could be that we could conceive of the brain as a "neural calculating machine", which consists of "interconnecting parts and the continuous transmission of motion from one part to another", and yet maintain that there are some objects or events with "a greater degree of 'thinghood'" than others, and which are, for example, the physical counterpart of the (psychological) image.

Again railway lines model neurones and trains impulses, but now the focus is not on the <u>pathways</u> themselves but on their <u>confluences</u>: "Just as a railway system consists of communicating lines on which trains are in continuous motion and stations where they stop and where definite events - such as the growth of factories - tend to occur, so there may be patterns of excitation or moments in a series of impulses at which there is some kind of demarcation which issues into consciousness as an image. Some parts of the process of neural transmission from sensory receptor to motor organ may have a physiological definiteness which is correlated with their psychological definiteness or emergence into consciousness as images."

So far, our poor psychologising hero (or perhaps victim?) has hardly started the day - having merely been rudely awakened and phoned the station. Let us let him get out into the countryside ... surely there he is safe from models. I'm afraid not...

Landscape.

"In the brain we find mountains and valleys, bridges and water courses, beams and arches, pins and hooks, claws and ramshorns, trees and sheaves, harps and trumpets ... No one knows the significance of these wonderful shapes" wrote Husche (Büchner, 1870), betraying the bewilderment felt by early neuro-psychologists at the complex, yet highly structured nature of that which they were trying to understand. ("The brain is not a simple organ, but is in the highest degree composite, rich in structure and delicately formed": Büchner.) In a less 'scientific' context, Gerard Manley Hopkins harped on the notion of a mental landscape, stressing its dramatic nature:

> "Oh the mind, mind has mountains; cliffs of fall
> Frightful, sheer, no-man fathomed. Hold them cheap
> May who ne'er hung there."

Mercier (1885) in an attempt to classify feelings, again used the landscape as a model, but for him the important notion was that of one thing underlying another. He drew an analogy between "physiological craving" and soil or earth, and between the 'higher' emotion love and the growth of verdure. Then, by discussing the relationship between verdure and earth, he was able to suggest a similar one between love and craving: "The emotion of love ... Underlying all its varieties there is the fundamental substructure of physiological craving, just as underlying every landscape there is the bare earth; and just as in some landscapes there is nought but bare or lichen-crusted rock, so in some natures there is nothing or little besides this craving. In other natures this substructure is covered and hidden by a luxuriant growth of higher forms of feeling ... (love) must in any classification be included under the same title with the simple physiological craving - a feeling from which it differs as the tropical luxuriance of a Brazilian forest differs from the lichen-covered rocks of Spitzbergen."

One major factor contributing to tropical luxuriance is, of course, water and we may expect our 'hero' to have encountered at least one stream. The flow of water has been associated with cognitive theory at least since the hydraulic automata of Hero of Alexandria and achieved notoriety in Descartes' thinking automata, which were to a great extent inspired by the hydraulic androids built by Thomas de Francine in the gardens at St. Germain en Laye. Actual brooks and rivulets are, however, also sometimes mentioned in psychological theories, as for example here by Francis Bowen (1877): "... we change the whole current of thought at will. We arrest the flow when we please, and thus force the river into a different channel. No one allows his thoughts always to drift at random, as they often do in aimless reverie or a dream. But the action of the Unconscious, which is the fountain that keeps the river always full and generally determines whether its waters shall be bitter or sweet, and which way they shall run, is often checked and controlled by the conscious Ego ..." A real stream of consciousness model !

Variations on the landscape analogy are no respecter of decade or century. La Mettrie wrote in 1748 that "if the brain is at the same time well organised and well educated, it is a fertile soil, well sown, that brings forth a hundredfold what it has received" - perhaps he should have added that a wealthy patron and a liberal atmosphere make good growing conditions. Nearer our own times, in 1974, Gardner (sic.) writes of brain injury as an "Avenue to the Mind", talks of the phrenologist's diagram of mental faculties as "road maps" and suggests that "locations and juxtapositions of parts of the nervous system sometimes offer ... additional guideposts to the relations among (or distances between) various psychological functions."

As you will recall, our victim admired the flora ("the brain that highest and fairest blossom of all terrestrial organisation": Büchner) and stalked the fauna ("the course of ideas in a given train of thought will admit of having its footsteps tracked in the corresponding pathways of the brain": Romanes; "the brain has its muscles for thinking as the legs have muscles for walking": La Mettrie).

If he was lucky, the tracks our hero followed may even have led him to a ferret. Incredible as it may seem, even the humble ferret has served as a model of the mind. In a brilliantly extended analogy, Charron (1801) exploits every aspect of his model. First he identifies the mind with the ferret, though stressing from the beginning the latter's (alleged) bad points: "the mind ... is a dangerous instrument both to itself and others, a little troublemaker, a ferret which is to be feared, an annoying and importunate parasite, which, like a juggler playing at sleight of hand, under the guise of some gentle motion, subtle and smiling, forges, invents, and causes all the mischief in the world; for without it there would be none." Then Charron consolidates the analogy: "... the mind is perpetually active; it cannot be without action."; "if it is not occupied with some action, it will run riot in imagination". "The action of the mind is to search, ferret, and endlessly twist about, like one that is famished for want of knowledge, to seek and enquire ... There is no end to our enquiries; the pursuits of the human mind are without form or end; its food is doubt and ambiguity; it is perpetually in movement, without rest or bound;

the world is a school of investigation; the chase and its excitement are our proper meat, but whether we catch the prey or miss it is another matter ..." The model established, Charron starts to draw out the (for him) unpleasant and worriesome aspects of it: "It is also universal meddling in everything, ... subjects vain and of no account as well as those noble and weighty, those we can understand as well as those we cannot."; "how rash and dangerous the mind of man may be, especially when it is lively and vigorous; for being so eager, so free and universal, so unrestrained in its movements, using its liberty so boldly in all things, without bowing to any, it may easily shake off common opinions, and all the rules by which one tries to bind and restrain it, as being an unjust tyranny." There are of course certain 'dangers' in this freedom, for example, "it is to be feared that it will wander and lose its way, and in fact we see that those who have extra-ordinary vivacity and rare excellence (of mind) are usually disordered in their opinions and in their conduct." ! After having followed the model through, Charron uses it to justify a reactionary and restrictive politics: "That is why there is good reason to give the mind strict limits, to bridle and bind it with religions, laws, customs, sciences, percepts, threats, promises mortal and immortal, yet still we see that in its unruly manner it escapes and makes itself free of all restrictions, for it is by nature stubborn, fierce and proud, and if it is to be led it must be by deception, and not by force."

It is at this stage our hero returns home intending to go to the... Cinema.

As long ago as 1748, La Mettrie had written: "all the faculties of the soul can be correctly reduced to pure imagination in which they all consist. Thus judgement, reason, and memory are not absolute parts of the soul, but merely modifications of this kind of medullary screen upon which images of the objects painted in the eye are projected as by a magic lantern."

The 'cinema model' has proved to be one of the most popular, and perhaps ultimately one of the most problematic, models in cognitive psychology, neatly exemplifying a notion of representation but simultaneously raising the reduction threat of the homunculus who is apparently required as audience. The most charming example of the use of this model is, perhaps, that of Erasmus Darwin (1806):

> "But as each mass the solar ray reflects,
> The eye's clear glass the transient beam collects;
> Bends to their focal point the rays that swerve,
> And paints the living image on the nerve.
> So in some village-barn, or festive hall
> The spheric lens illumes the whiten'd wall;
> O'er the bright field successive figures fleet,
> And motley shadows dance along the sheet -
> Symbol of solid forms is colour'd light,
> And the mute language of the touch is sight."

We must not imagine, however, that the cinema model is one entirely of the past. As recently as 1976, Arbib et al. have used a "metaphor drawn from the making of movie cartoons." Arbib is really concerned with a model within a model. He accepts that an organism models the salient features of its environment, but the problem then is how we shall model the organism's model. Arbib's particular problem has two roots: firstly it seems over-whelmingly likely that the brain is a precise computational network which, while functionally unitary, is spatially distributed. The hypothesis, then, is that computation in the brain "may involve the co-operation of many subroutines that work simultaneously and in parallel (and wherein) a computation may be effected by a subset of the routines, (and where) for some tasks some subroutines may be irrelevant" (and thus be ablated without relevant function loss).

Now, apparently in the making of cartoons, it is too inefficient (and therefore costly) to draw each frame individually in its entirety before they are photographed and strung together as the animation. Instead, a "layering technique" is employed. In this technique, the final photograph is taken looking down a "slide box" which contains many different layers; each layer will have depicted upon it a different aspect of the total picture e.g. background, foreground, midground, figures, limbs, etc. Now, for any one frame, only some minor change may be required, for example, the background and a tree in the foreground may remain the same, but some aspect of the figure ... the position of the legs ... may change. Thus only one layer need be altered and the work of duplicating all the unaltered information in the slide is rendered unnecessary. All slides not at present in use would be kept in a slide file. This is enough by way of background to understand Arbib's use of the model: "a similar strategy for obtaining a very economical description of what happens over a long period of time may be used in the brain with a long term memory (L.T.M.) corresponding to the slide file and a short term memory (S.T.M.) to the slide box. The act of perception might then be compared to using sensory information to retrieve appropriate slides from the file to replace or augment those already in the slide box, experimenting to decide whether a newly retrieved slide fits sensory input 'better' than one currently in the slide box. Also, part of the action of the organism in changing its relationship with the environment might be viewed as designed to obtain input that will help to update the S.T.M. by deciding between 'competing' slides, as well as helping update the L.T.M., by 'redrawing' or 'editing' the slides."

Arbib's model, then, helps us make sense of the notion of distributed computation. But isn't it unlikely (physiologically) that we have a neural cartoon-making-mechanism between our ears? (This objection is made to all models, but especially the more outlandish). The fact is that psychological theories constrain physiological theories even more than vice versa ... the only physiological constraint on a psychological theory is that the latter must somehow be physiologically realisable; but the psychological theory tells the physiologist what must be realised. If the psychological theory dictates a distributed computational system, the fact that we have "little feeling for how to 'wire up' a neural 'slide box system' is a measure of how much further we have to go if we are to understand the neural mechanisms."

Telegraph.

Whilst Samuel Morse was travelling from Le Havre to New York on S.S. Sully, he is reputed to have seen, in October 1832, Dr. Jackson demonstrate an electro-magnet and allegedly said: "If the presence of electricity can be made visible in any part of an electrical circuit closed by an electro-magnet, I see no reason why intelligence may not be transmitted instantaneously by electricity" (Larsen, 1960).

Like the telephone system (but of course earlier) the telegraph system forms a natural model of neural pathways, where nerve fibres are modelled by wires and ganglions by the electrical apparatus (Buchner). The most thorough explanation of the analogy is found in Shelly and Stenhouse (1911) and it is worth quoting fairly fully.

After telling us that "in the spinal cord and the brain are certain very remarkable cells, called nerve cells which are able to communicate with each other, as well as with the sense organs muscles, glands, and other tissues, by means of fine threads called nerve fibres"; they go on to say: "It may simplify our ideas of these things to think of them as the telegraph system of a country. The country is the human body, the telegraph wires are the nerve fibres, and the electric batteries are the nerve cells. We shall also picture to ourselves the batteries as being nearly all collected in certain definite parts of the country - the spinal cord and the brain, which are together referred to as the central nervous system.

"Further, we can suppose the organs of touch, sight, hearing, smell and taste to be so many sets of telegraph instruments at certain offices along the frontier of the country, receiving and transmitting information concerning foreign affairs."

Once the model is set up, it can be used to make predictions which can be empirically verified. Thus just as "many telegraph wires may be bound up together to form one cable ... on what may be called main lines, many nerve fibres are bound up together to form one nerve." Moreover, just as telegraph wires are protected from interference from each other and from losing their currents "by the sheaths of india rubber or silk ... nerve fibres in use are kept from touching each other - and their messages therefore kept from going astray - by sheaths of fatty substance."

Perhaps you will remember that the telegram our victim received was to invite him to a ...

Concert.

Music has provided many analogies for understanding man. Many of these are based upon a supposed analogy between the vibration of musical (e.g. violin) and nerve strings. Thus La Mettrie wrote: "As a violin string or a harpsichord key vibrates and gives forth sound, so the cerebral fibres, struck by waves of sound, are stimulated to render or repeat the words that strike them.", apparently as a straightforwardly physical account of what happens. Romanes seizes on the same phenomenon (vibration) to explain a way of achieving a 'higher synthesis' of "spiritualism and materialism." Romanes wants to demonstrate that the antithesis between mind and motion is only phenomenal not real, (or that this is at least a possibility). Accordingly, he needs to find a parallel system, in which another phenomenological antithesis is not actual, to use as a model. Thus he writes: "When a violin is played upon we hear a musical sound, and at the same time we see a vibration of the strings. Relatively to our consciousness, therefore, we have two sets of changes, which appear to be very different in kind; yet we know that in an absolute sense they are one and the same: we know that the diversity in consciousness is created only by the difference in our modes of perceiving the same event - whether we see or whether we hear the vibrations of the strings. Similarly, we may suppose that a vibration of nerve-strings and a process of thought are really one and the same event, which is dual or diverse only in relation to our modes of perceiving it." The same argument, based on vibration, is also rehearsed by Huschke (see Büchner), substituting, or rather augmenting with, colour for music: "As colour is to the vibrations of light, as sound is to the vibrations of elastic fluids, so is thought related to the neuro electrical vibrations of the brain fibres." William James' (1891) famous dictum: "the melody floats from the harp string, but neither checks nor quickens its vibrations" though meant to illustrate an epiphenomenalist rather than identity theory, makes the same point.

An interesting development of this metaphor is employed elsewhere by Romanes. Nerve strings and violin strings have this in common ... they both vibrate; they are both accessible through different senses. But note that a vibrating violin string may be either in or out of tune and hence harmonious or discordant with the rest of the orchestra. Accordingly, Romanes extends the model: "What we know on the side of mind as logical sequence, is on the side of the nervous system nothing more than the passage of nervous energy through one series of cells and fibres rather than through another: what we recognise as the truth is merely the fact of the brain vibrating in tune with nature."

Cookery.

It may be surprising to be given by a psychologist literally food for thought, but it not uncommonly happens. William James talks of "the brain ... (as) ... a sort of vat in which feelings and emotions somehow go on stewing together" and La Mettrie, in stressing the shared heritage of man and beast, likens them to types of bread: "Man is not moulded from a costlier clay; nature has used but one dough and has merely varied the leaven ... (man and animal) must necessarily be in the same condition." Later, he develops the notion of difference in leaven to say: "animals, composed of the same matter, lacking perhaps only one degree of fermentation to make it exactly like man's, must share the same prerogatives of animal nature." Jung (1955), too, favoured carbohydrate models. In an interesting interview with Stephen Black he said: "In those days one talked of psychiatric illness as a sort of by-product of the brain. Joking with my pupils, I told them of an old text book for the Medical Corps in the Swiss Army which gave a description of the brain, saying it looked like a dish of macaroni, and the steam from the macaroni was the psyche." Here Jung uses a crude model of epiphenomenalism, to ridicule the concept modelled purposely and effectively (but of course to ridicule is not to show false).

(Slightly) more serious use of food and cookery as models has however been made, e.g. by Craik.

Craik was intrigued by how we could make sense in a mechanical way of our apparent ability to generalise on abstract properties of objects. It struck him that we were perhaps not taking something extra in (a difficult abstract property) but leaving much out ... everything which was irrelevant in that respect ... in virtue of the way we are constructed. Thus it might be "that in genuine generalisations apparently abstract properties of objects are really recognised as the same because, acting on the brain mechanisms of the animal, they produce the same effect, just as a pound of butter and a pound of bacon both produce the same deflection on a balance ... Thus, a balance 'recognises' a weight of one grm., whether it be the weight of a piece of brass or of lead. In the same way, we can recognise colours apart from the nature of the coloured objects."

This kitchen-scales model underlies much theorising on cross-sensory comparison. How do we confirm what we see by what we hear? How do we confirm what we smell by what we taste? How do we confirm what we are told by what we touch? One obvious way would be to translate all perceptual inputs into a common code and then define a confirmation matrix for the code, but this is really no more than a notational variant on the set of cognitive kitchen scales discussed above (and probably deserves the epithet 'mental chemistry approach' more than Savin's own). It is interesting to note that one bold answer to the scales approach denies the premise that we do have to translate inputs into a common code. Bower's 'unity of the senses' hypothesis seems to suggest that a child starts with such a common code unspecified as to particular senses. It is however tempting to point out that this is not the first time the concept appears in the literature. In a work of speculative mechanistic psychology of 1818, significantly subtitled 'The Modern Prometheus', Mary Shelley (1818) makes her famous creation say: "It is with considerable difficulty that I remember the original era of my being: all the events of that period appear confused and indistinct. A strange multiplicity of sensations seized me, and I saw, felt, heard and smelt, at the same time; and it was indeed, a long time before I learned to distinguish between the operations of my various senses."

We can spare our victim from much more anguish, and allow the evening to pass quickly by.

Suffice it to say that he listened for a while to records (Delboeuf's and Guyau's models of auditory memory) ... and went out. He returned later to light the lamp ("In an Edison lamp the light which is emitted from the burner may be said indifferently to be caused by the number of vibrations per second going on in the carbon, or by the temperature of the carbon ... Similarly a train of thought may be said indifferently to be caused by brain-action or by mind-action; for ex hypothesi, the one could not take place without the other." Romanes), taking care to cover the caged budgie (Plato's birdcage model of memory ... "now let us make in each soul a kind of aviary"), whilst reading a book (for copious illustrations see next chapter).

It only remains then to our victim to ablute himself before retiring. Unfortunately, he is not even safe from the prying modelling mind of the psychologist even as he pulls the bath-plug. It was Francis Galton who devised a model of the mind where the unconscious was modelled by the sewage system and power lines beneath our homes: "Introspection", he wrote, "gave me an interesting and unexpected view of the number of operations of the mind, and of the obscure depths in which they took place, of which I had been little conscious before. The general impression they have left upon me is like that which many of us have experienced when the basement of our house happens to be under thorough sanitary repairs, and we realise for the first time the complex system of drains and gas and water pipes, flues, bell wires and so forth, upon which our comfort depends, but which are usually hidden out of sight, and with whose existence, as long as they acted well, we had never troubled ourselves."

This surely must be the end ... you cannot go any lower than the foundations. Well ... yes you can. We will finish off with perhaps the most backhanded of compliments to the psychologist in terms of what he studies ... that of Karl Vogt (1845-9), who wrote: "Thought stands in the same relation to the brain, as ... urine does to the kidneys."

This remark, made in a letter, was widely circulated and raised a furore of horrified criticism. Vogt himself admitted his aphorism was expressed "a bit crudely", as would have been realised had he been fully quoted, but stuck to his main point, that "all those capacities that we understand by the phrase 'psychic activities' are but functions of the brain substance"; this point was, of course, frequently missed in the uproar.

If, as Longuet Higgins (1968) has written in a different connection, there is a "lurking thought that machines which imitate our brains are for some reason to be taken more seriously than machines which imitate our arms or even our kidneys", how are we to react to the suggestion that the brain itself is a machine which imitates our bodily functions? The answer is, historically at least, very seriously indeed.

Büchner (1870) talked of the soul as "a special way of expressing the force of life, determined by the characteristic construction of the brain", and asserted that the "same force that digests by means of the stomach, thinks through the brain."

Slightly earlier (1824) however, the most explicit use of the secretion model of thought had been made by Cabanis.

Cabanis argued that "In order to form a correct idea of the operations which give rise to thought, one must look upon the brain, as a special organ, which is particularly designed to produce it, in the same way that the stomach and the intestines are designed for performing digestion, the liver for filtering the bile ..." He compared impressions reaching the brain to foods reaching the stomach ... both stimulate their respective organs to activity, and "the organic movements" by which the functions of each organ are performed, are equally unknown.

"We see the foods drop into this viscus, having certain qualities of their own; we see them emerge from it with new qualities, and we conclude that it has in fact changed them in this manner. We see equally that impressions reach the brain by the intermediary of the nerves, they are at that time, isolated, disorganised. The viscus enters into action; it acts on them, and soon they reappear metamorphosed into ideas which the language of facial expression and of gesture, or the signs of speech and writing, manifest outwardly. We conclude with the same certainty that the brain in some sense digests the impressions; that it organically produces the secretion of thought."

In 1856, Adolf Harless published a play - a satire on the whole materialist neuro-psychological enterprise. The play, Goethe im Fegfeuer, was subtitled, mockingly, Eine materialisch-poetische Gehirnssekretion (a materialistico-poetical brain-secretion) ! I hope the reader will not conclude that all I have been doing in this chapter is, like Harless, taking the thinking out of psychological models.

# CHAPTER TEN

## THE STOREHOUSE OF PAST IMPRESSIONS.

"I consider that a man's brain originally is like a little empty attic and you have to stock it with such functions as you choose. A fool takes in all the lumber of every sort that he comes across ... now the skilled workman is very careful indeed as to what he takes into his brain-attic. He will have nothing but the tools which may help him in doing his work, but of these he has a large assortment, and all in the most perfect order. It is a mistake to think that the little room has elastic walls and can distend to any extent. Depend upon it that there comes a time when for every addition of knowledge you forget something that you knew before."

Holmes to Watson
(quoted Gardner 1977).

Prologue.

Philosophers have been known to assert that 'normal' psychological phenomena need no explanation; only the unusual, the unexpected - that is, departures from the norm - can be called to give an account of themselves (Hamlyn, 1957). As remarks about the scope of theories in science, such claims are, of course, preposterous (see Grice, 1961, and Fodor, 1968, for rather more elaborate dismissals); but reconstrued as an observation about the historical development of scientific interest in human behaviour the position acquires considerable validity. As Chomsky (1968) has noted 'One difficulty in the psychological sciences lies in the familiarity of the phenomena with which they deal. A certain intellectual effort is required to see how such phenomena can pose serious problems or call for intricate explanatory theories. One is inclined to take them for granted as necessary or somehow "natural".' But this complacency is more difficult to maintain when our attention is drawn to extreme cases of a 'natural' human ability. Whilst it is by no means always true that "the psychologist who desires to analyse a faculty of the mind has his wish answered if he meets a creature who possesses this faculty to an eminent degree" (Duhem, 1914), our curiosity is aroused by individuals whose capabilities seem to be far greater, or far smaller, than our own.

Aspects of memory - Historic studies of remembering and forgetting.

So it is when, in his Natural History, the elder Pliny (23-79 A.D.) reviews the data on memory, that 'boon most necessary for life'. First, those men who excel are mentioned: "King Syrus could give their names to all the soldiers in his army, Lucius Scipio knew the names of the whole Roman people ... Charmadas recited the contents of any volumes in libraries that anyone asked him to quote, just as if he were reading them."

Next are described those cases where 'diseases and accident' have resulted in a striking diminuition of ability: "One who fell from a very high roof forgot his mother and his relatives and friends, another when ill forgot his servants also; the orator Messala Corvinus forgot his own name." Having recounted these differences between individuals, Pliny concludes by observing that analogous variation exists within an individual and is dependent upon the psychological and physiological state of the person: "Similarly tentative and hesitating lapses of memory often occur when the body even when uninjured is in repose; also the gradual approach of sleep curtails the memory and makes the unoccupied mind wonder where it is."

Such variations in the power to recall have no doubt been noticed and remarked upon by men ever since there were men. But noticing and theorizing are rather different occupations. Psychological theory in the modern sense of the term was created when, in the aftermath of the Ionian enlightenment, the Atomists and the school of Hippocrates made the psyche, its sensations and cognitions into a physical object. This was achieved by turning man's technology into a metaphor for the structure of the universe and eventually of man himself. As Hippocrates (460-375 B.C.) puts it: "Men do not understand how to observe the invisible by means of the visible. Their techniques resemble the physiological processes of man but men do not know this. (T)hough men understand the technical processes, they fail to understand the natural processes imitated by the techniques." (The translation is from Farrington, 1947). A panoramic impression of psychological theory as a set of 'machines of the mind' is given in Marshall (1977); here we will concentrate on but one aspect of that story, the elaboration of physical metaphors of memory.

The construction of the dominant models.

The foundations of the theory of memory were laid, somewhat fortuitously, when in 477 B.C. the poet Simonides identified the mangled bodies at Scopas' banquet by recalling the places at which the guests had been sitting. It was in this highly unfortunate fashion that the Ars memorie artificiali began. Later poets and orators borrowed from Simonides this technique of training the faculty of remembrance by placing on an orderly sequence of prelearnt architectural locales the images they wished to commit to memory (Yates, 1966). One should realize, however, that for Simonides his discovery was merely the lucky invention of a technique, a method. And as a technique it has proved remarkably efficient to this very day; Patten (1972), for example, has reported how the ancient mnemonic system can be used to help patients with memory defects consequent upon cerebral lesions. But the crucial step from a theoretical point of view occurred when, in the aftermath of Hippocrates' 'physicalization' of psychology, a variety of structural models were elaborated from Simonides' accidental discovery. It had become possible to think of memories as 'objects', objects having a size and a place. In order to count as memories, events, though past, must be stored, and hence be potentially available - by being taken out of store - at dates later than the time of original occurrence - when they were 'laid down'. The way to a general theory of storage was prepared by Empedocles (495-435 B.C.), who conjectured that sight is merely an extended version of touch, and by Democritus (460-360 B.C.), who guessed that the images of objects 'moulded' the air through which they travelled en route to the eyes. (An excellent discussion of this background will be found in Beare, 1906). Thus it was that Plato (427-347 B.C.), in the Theaetetus, and his graduate student Aristotle (384-322 B.C.), in De Memoria et Reminiscentia could extend and develop the Atomists' notions by suggesting that sensory images may impress themselves upon the mind as a signet ring may leave its impression in a block of wax.

Plato's metaphor - "Suppose, then, I beg, for the sake of argument, that we have in our souls a waxen tablet ..." - allows for the possibility of quantitative measurements of memory along a number of (qualitatively distinct) dimensions. The tablet in one mind may be larger or smaller than that in another, it may be more or less pliable, and it may be more or less pure (that is, contain few or many imperfections which will distort the impression of the seal). Plato summarizes his position as follows: "... whatever is imprinted, this we remember and know, as long as its image remains; but when it is effaced, or can be no longer imprinted, we forget and do not know it." The operation of the system is fairly crude, consisting merely of template matching. When we try to recognize or remember 'things that we have seen or heard, or have ourselves thought of' we simply place the previously impressed block 'under our perceptions and thoughts'. If a match can be found, then the object is recognized. False positives arise when an approximately matching stimulus is taken as a perfectly fitting one, an error Plato likens to putting one's shoes on the wrong feet.

Plato then proceeds to explain individual differences in terms of his model. If the block is small, then images will fall on top of each other and the templates will become confused with each other; similar distortions will arise when the wax is 'stony or full of earth or mixed mud'. The variable of hardness is particularly interesting because it enables Plato to draw the distinction between learning and memory. If the wax is hard, the subject will find learning difficult; many trials will be necessary to make an impression on the wax. But once the impression is stamped in it will persist with relatively little loss of definition over time. If, on the contrary, the wax is soft, the person can be easily taught, but he will also forget quickly as the pliable image become indistinct. Aristotle was later to argue that the shortness of memory in children is due to the soft, moist nature of their brains; by contrast, he claimed that the inability of the old to learn new facts is due to the rigidity and hardness of the elderly brain. Later still, Gratoroli (1562) was to propose that yet other individual differences could be interpreted in terms of the model: "There be fewe founde that are indewed with a good witte and an excellent Memorie of Nature: for because that witte betokeneth a subtile and softe substance of the braine, and Memorie a permanent substance."

This dimension of the theory was neatly captured by Alexander Pope when he wrote in his best male chauvinist pig style:

"Nothing so true as what you once let fall,
Most Women have no Character at all.
Matter too soft a lasting mark to bear,
And best distinguished by black, brown, or fair."

And a little later, James Harris (1751) added a further temporal or developmental component to the model. Harris notes that the consistency of the wax is the crucial part of Plato's theory - "... the Wax would not be adequate to its business of Signature, had it not a Power to <u>retain</u>, as well as to <u>receive</u>." The receptive power Harris calls <u>sense</u>, and the retentive power <u>imagination</u>. He then broaches the interesting question: What would it be like to have sense without imagination? And, the answer borrowed from Aristotle's <u>De Memoria</u>, is that it would be like making impressions on water, "where tho' all impressions may be instantly made, yet as soon as made they are as instantly lost". (Some more recent speculations on this form of iconic and echoic memory of very brief duration will be found in Sperling, 1960, and Crowder and Morton, 1969.) Harris accordingly calls sense "a kind of transient Imagination" and imagination "a kind of permanent Sense". The imagination, then, has the power to convert a transient state into a permanent one. Harris continues: "Now as our Feet in vain venture to walk upon the River, till the Frost bind the Current, and harden the yielding Surface; so does the Soul in vain seek to exert its higher Powers, the Powers I mean of Reason and Intellect, till Imagination first fix the fluency of Sense, and thus provide a proper Basis for the support of its higher Energies." (Later versions of 'consolidation' theory, as Harris' metaphor is now called, will be found in Muller and Pilzecker, 1900, and in Deutsch, 1973.)

Plato's first model, the wax block, provides us with a superb conceptualization of one aspect of memory - recognition memory. It expresses one way in which a stimulus seen for the second time can be recognized as the same stimulus as that presented the first time. The model is less suggestive as a way of approaching the ideas of storage and recall. Plato accordingly develops a second metaphor - the birdcage model; this is the notion of memory as an aviary: "... now let us make in each soul a kind of aviary of all sorts of birds, some being in flocks, apart from others, and others few together, and others alone ..."

This enables Plato to speak of knowledge as "possessing them (them = birds = ideas) in that aviary", to call learning the process of receiving birds into the (previously empty) aviary, and to call teaching transferring the birds from one cage ( = mind) to another. For the later development of psychology, however, it is more crucial to note that the model enables one to draw a firm distinction between storage and retrieval. As Plato puts it, the bird can be in the cage without one necessarily being able to catch it instantaneously and at will; that is to say, "catching is of two kinds, one before possessing for the sake of possessing, the other when one already obtained possession for the purpose of having in the hands what was already possessed." This second type of 'catching' is critical, of course, for it shows how errors might arise in the retrieval process, "as it were taking a pigeon that he possessed instead of a dove."

The concept of a memory store emerges, then, fully fledged in Plato, and continues in an unbroken line to the present day. Cicero (106-43 B.C.), in his De Oratore, remarks that "Memory is the treasurehouse of all things"; John Locke (1632-1704), in his Essay Concerning Human Understanding, is rather more frugal when he writes that memory "is as it were the storehouse of our ideas", and Henry Head (1920) specifies the location of memorial capacity: "The sensory cortex is the storehouse of past impressions."

By flipping backwards and forwards between Plato's two models the full range of modern answers to the question "Why does memory sometimes fail?" is generated; memory may fail because the percept or idea never got into the store in the first place (Treisman, 1964); it did get in but then disappeared, either through spontaneous decay or natural wear and tear (Brown, 1958); it got in but was then destroyed by some other object being dumped on top of it (Sperling, 1960); it got in but was pushed out by some later arriving stimulus (Cermak, 1976); it got in and is still there but the storeman cannot find it amidst the rest of the junk (Underwood, 1957).

Thomas Reid (1719-96), in his Essays on the Intellectual Powers of Man, notes that it is "very natural to express the operations of the mind by images taken from things material". Indeed it is, but Reid unfortunately fails to note the value of so doing, namely, that it enables one to conceive of making quantitative measurements of the powers of the mind. The Greeks did not themselves wish to indulge in anything quite so vulgar as this, of course, but they did provide the models which enabled later scholars to do so. In a similar fashion to Reid, Ebbinghaus (1885) has remarked that in order "to express our ideas concerning the physical basis of memory we use different metaphors - stored up ideas, engraved images, well-beaten paths". And again like Reid, he is sceptical of these models: "there is only one thing certain about all these figures of speech and that is that they are not suitable." The scepticism may be justified, but it is difficult to see why Ebbinghaus conducted the particular experiments that he did if it was not to put quantitative flesh on the bones of Plato's 'stamping in' model.

Certainly, Ebbinghaus' contemporaries took him to be proposing a new branch of psychometry, not a new theory. As Joseph Jacobs (1885) commented at the time, Ebbinghaus' experiments were "remarkable . . more for their methods than their results". The results, said Jacobs, "... scarcely seemed calculated to set the Spree on fire".

Ebbinghaus' primary concern, then, was with measurement and control; he was one of the first scholars to demonstrate that associative learning processes could be captured and quantified in a laboratory setting. Ebbinghaus' more general aim - "to get a foothold for the application of the method of the natural sciences" - involved him with all the familiar paraphernalia of the modern laboratory of experimental psychology. He was careful to establish stable, controlled experimental conditions, learning his stimulus items to the beat of a metronome; he attempted to standardize both procedure (by his invention of the 'anticipation' method) and material (by the use of nonsense syllables). The intent of this latter innovation was to ensure that subjects did not bring differing degrees of familiarity with the experimental material to the learning task. However, the contribution was not, and is not, appreciated by all psychologists, many of whom consider that asking subjects to repeat such syllables is utter nonsense. The basic problem of course, is that all subjects (with the possible exception of Ebbinghaus himself) bring to the task the ability to import meaning into nonsense by crafty mnemonic devices. Subjects may well differ in this ability more than they differ in familiarity with ordinary language material.

The nature of Ebbinghaus' results: that the number of trials required to learn increases very rapidly as the material increases in length; that the curve of forgetting falls rapidly at first, then more slowly; that overlearning is proportionally related to extent of remembering; and that repetitions separated in time ('spaced') are more effective than when crammed together ('massed') is, as Jacobs remarked, unexciting. The precise, numerical parameters of these effects are, however, of paramount importance for the construction of finely detailed, fully explicit models. (Murray, 1976, provides a superb review of this aspect of Ebbinghaus' work.)

None the less, Ebbinghaus' contributions were almost entirely quantitative and methodological. In terms of technique, his introduction of the relearning method, and the associated use of saving scores to infer 'below threshold' retention are major innovations. When learnt material had been totally lost to conscious recall, Ebbinghaus was able to demonstrate 'unconscious' retention by showing that the information could be reacquired in fewer trials than were originally needed for learning. But in terms of conceptual advance, he took not a step beyond the Platonic tablet. As Ebbinghaus writes in section 31, "If the relearning is performed a second, third, or a greater number of times, the series are more deeply engraved and fade out less easily"!

It is a curious fact about both the wax block and the storehouse metaphors that they have a mysteriously seductive power for those scholars who deny their usefulness. Thus Bartlett (1932) objected to Henry Head's use of the latter metaphor. Bartlett claimed that "a storehouse is a place where things are put in the hope that they may be found again when they are wanted exactly as they were when first stored away". That this 'hope' is not always fulfilled can be easily verified by anyone who leaves chopped liver in a warm cupboard for a few days. Bartlett did, of course, try to break out of the storehouse by revitalizing (and reinterpreting) Kant and Head's notion of 'schema'. A schema for Bartlett purported to be a "living ... constantly developing ... active organization of past reactions or of past experiences" and he believed that "the storehouse notion is as far removed from this as it well could be". Yet as Mary Northway (1940) points out, "although Bartlett states that he is considering remembering as an active reorganizing process, in many of the cases he gives, the idea of memory as a 'storehouse' seeps in only too easily." Currently, Bartlett is often cited in support of the notion that 'much of what is remembered is reconstructed from stored fragments' (Fodor, 1975). Just so ... but this is not to impugn the storehouse metaphor at all, it is merely to offer a variation on its contents.

A non-random walk through the great library.

Meditation upon the consequences of the storehouse view of memory led directly to the emergence of the major conceptual model that has guided modern studies. We have seen how Plato distinguished storage and retrieval, and was then able to imagine the possibility of retrieval failure to error. Such retrieval miscalculations are presumably dependent, in part at least, upon the nature of what is stored and the manner in which it is stored. If all the birds are black, then it will be all too easy to retrieve the wrong bird; if some are white, others gray, red, blue and so forth, then such mistakes should be less likely. If a bird is close to the entrance then it should be easy to grab it; if the bird is far away then one may fail to reach it. If all birds of one type tend to cluster together in one particular part of the aviary, then one knows 'where to look' for them; if the birds are randomly distributed, then no such reduction of the 'search' space will be possible. We see, then, how the idea of pro- and retroactive inhibition arises, and with it the notion that 'recollecting is, as it were, a kind of search' (Aristotle, _De Memoria_). Once this position is reached the concept of _organization_ becomes critical for all later theorizing. The key questions are now: "According to what principles are memories stacked in the store?" (throwing the goods into the warehouse at random can only lead to chaos); and "According to what principles do we search through the store?" (searching in a haphazard fashion is likely to be grossly inefficient).

At this point, the metaphors of the storehouse and the wax tablet on which the scholar can draw or write finally merge into one; Cicero (_De Oratore_) complete the link with the original technique of memory improvement by equating Simonides' places with the wax tablet and the images with the letters written upon it.

Such written documents both preserve past ideas and events and are themselves preserved in the great libraries of the Hellenic world (Parsons, 1952). Once a collection reaches a critical size it must be organized in a principled fashion if it is to fulfil its scholarly or bureaucratic functions. Aristotle - who must have catalogued his own large private collection (Norris, 1939) - took over the terminology of the place (topoi) theory of memory to refer to general patterns of argument, subject headings (or topics) under which a variety of instances fall (Sorabji, 1972). This new theory - that memory is a well laid-out and well indexed library - was ably summarized by Kant (1798): "Most of all, the use of topics - that is, of a framework for universal concepts, called general headings (loci topici) - makes remembering easier, by dividing the material into classes, as when we arrange the books in a library on shelves with different labels." (The translation is by Gregor, 1974).

The most elaborate version of the library hypothesis was produced by Bowen (1877); most modern accounts of 'semantic memory' build upon one or more aspects of this model, and it is therefore worth quoting Bowen at some length:

> "Many educated persons ... know enough of at least four languages, Latin, French, German or Italian, and English to be able to read any common book in either of them with about equal facility. The whole number of English words, not including purely technical terms or mere derivatives, is at least 40,000; and that portion of the vocabulary of either of the other three languages, which is at the command of a well-educated foreigner, is probably half as large. Among the treasures of memory in such a mind, therefore, must be reckoned at least 100,000 mere words, all of which, with some trifling exceptions for onomatopoeia, are symbols as arbitrary as the signs in algebra. What a countless multitude of individual facts and familiar truths in science and ordinary life are either wrapped up in these words, or exist side by side with them, in any well-informed mind! Certainly such a mind is far more richly stocked with words and ideas than the British Museum is with books. That admirably managed institution, suffering from an embarrassment of riches, maintains a full staff of well-trained librarians; and each of them, after rummaging the catalogue and the shelves for perhaps ten minutes, will triumphantly produce any volume that may be called for. But the single invisible librarian, who awaits our orders in the crowded chambers of Memory, is far more speedy and skilful in his service. A student reads a page of French or German in a minute, and for each of the two or three hundred groups of hieroglyphics printed on it, 'the Unconscious' instantly furnishes us whatever we call for, either its meaning, or its etymology, or its English equivalent, or its grammatical relations to other groups in the

> same sentence, or any of the associated ideas in a little world of knowledge of which this one word forms the centre. We have no conscious clew with which to direct ourselves in the search; it is enough that we have an interest in the point to be remembered, that we need it for the work which is at hand, and instantly it is produced out of the vast repository."

A wax tablet obviously allows either drawings or writings to be inscribed on it, and these two forms of notation are taken over into the library model. We thus find both a verbal lexicon (Morton, 1970) and a visual eidolikon (Seymour, 1973) co-existing, albeit not always peaceably (Pylyshyn, 1973), within the 'great library'. St. Augustine (in his Confessions) remarks that only some thoughts (or 'internal representations', to use the currently fashionable euphemism) are stored in the form of visual images whilst others are stored in propositional form. A considerable proportion of nineteenth and twentieth century work on the library model has been devoted to arguing that different types of stacking arrangements are used for the two types of material, verbal and visuo-spatial, that make up the contents of the library. One of the more breathtaking hypotheses of recent times has been the claim that the left wing of the library contains verbal images and the right wing visuo-spatial ones (Hughlings Jackson, 1874).

It has usually been conjectured that the most prevalent form of organization for (meaningful) verbal items is a tree structure which branches from general to more and more special terms (see Ramus, 1578, for discussion). The internal architecture of the library thus becomes an n-dimensional pyramid stretching from superordinate categories on the top floor to subordinate categories on progressively lower and lower floors. The structure of the store has always been deduced from retrieval errors, on the assumption that erroneous responses will be items that are stored close to the objects that should have been retrieved. It is widely held that cerebral damage may lead to local perturbations of stacking. Thus Thomson (1907) writes: "... after some brain shock, a person may be able to speak, but the wrong word often vexatiously comes to his lips, just as if his Broca shelves had been jumbled." A strictly spatial interpretation of the theory was, however, sometimes felt to be mildly problematic on physiological grounds; it was somewhat puzzling to imagine how the brain could achieve such an orderly addressing system. Thus Gesner (1770) noted the prevalance of semantic paraphasias in some of his patients (they, for instance, say "Good evening" when they mean "Good morning") and he also observed that certain classes of words (for instance, abstract nouns) were more liable to be lost than others. The spatial theory did not appeal to him, however, on the grounds that "The vessels of the brain are surely not arranged in accordance with categories of physical ideas and therefore it is incomprehensible that these categories should correspond to areas of destruction."

Such neurological qualms have, of course, never inhibited the experimentalist, who could see a way of converting <u>space</u> in the model to <u>time</u> that he could directly measure. And in any case, even stranger things are known to occur in the endothelium of the cornea (Bard et al., 1975), and in the optic tectum (Keating et al., 1975).

The reasoning of the experimentalists (Cattell, 1887; Cattell and Bryant, 1889; Collins and Quillian, 1969) builds upon Aristotle's idea of search and runs as follows: Imagine that places are marked in the library and that the relationships (pathways) between places are also marked. If we give a starting point and a (two-place) relationship to the subject we ought to be able to measure how long it takes him to traverse the pathway to the item(s) which is (are) linked to it. For example, given the relationship 'is a' and the term 'dog', we can measure the time taken to complete the proposition "A dog is a ..." with the term "animal". Alternatively, we can give examples of true ("A canary is a bird") and false ("A duck is a plant") propositions and measure the time taken to evaluate the truth-function of such statements. The assumption is that we can work out the Euclidian and/or city block metrics of the library's internal structure by taking reaction times in the above fashion. As Aristotle put it: "There is no need to consider how we remember what is distant, but only what is neighbouring, for clearly the method is the same." Much of the geometry and topology of the library was worked out in this fashion by Wilhelm Wundt and his colleagues at Leipzig (see Marshall, 1970, for discussion and references), and the method has recently become popular again (see, for example, Loftus et al., 1970; Meyer and Schvaneveldt, 1976). It is customary today to think of the library as incorporating a set of sub-geometries appropriate to different conceptual domains. This idea works particularly well for so called 'semantic memory'. For example, a tree structure seems appropriate for many of the set and superset relationships involved in the hierarchical classifications used in animal and plant taxonomies; on the other hand, a cube with the faces marked by $\pm$ sex, $\pm$ co-lineal, and $\pm$ descendant has often been regarded as the appropriate structure for classifying the eight basic kinship terms of English (Wood and Shotter, 1973, have provided a very neat reaction-time justification for this latter claim.)

Another beautiful example of the relationship between space and time has recently arisen in the study of immediate memory. Recall that many scholars have postulated that the library has a small entrance which can hold a limited number, $7 \pm 2$ on average (Miller, 1956a), of recent acquisitions for a limited period of time. This number, the so called 'span' of immediate memory (Jacobs, 1887), varies slightly as a function of the <u>type</u> of information which is put into store. Digit span is a little higher than letter span which is in turn higher than word span which is in turn higher than nonsense syllable span. Nonsense syllables, then, take up more space in the anteroom than do digits. From studies which utilize Sternberg's memory scanning technique, Cavanagh (1972) has discovered a linear relation between scanning rate (milliseconds per item) for these classes and the reciprocals of their memory span values. One can imagine no more elegant a demonstration of the power of the storehouse metaphor.

But let us leave the anteroom and return to the main body of the library. We are now in a position to see the paradigm as a whole. The notion of a library as a general framework must be supplemented by the details of storage-arrangements and retrieval strategies. The classical laws of association theory were intended to provide precisely those details; associationism formulated the principles which determined the distribution of elements which entered the system, and hence the principles of effective search. A sequence of associations is thus a path through the great library. Harris (1751) writes: "... the Road, which leads to Memory thro' a series of Ideas, however connected whether rationally or casually, this is RECOLLECTION."

Three classes of associative principles are usually recognized. In the terminology of Abercrombie (1857) these are:

I Natural or philosophical association

II Local or incidental association

III Arbitrary or fictitious association.

In the first class, items are connected rationally; proximity of elements is a function of logical or semantic structure.

In the second class, items are connected by accident of time or place; elements are entered together by virtue of their temporal or spatial contiguity in the flow of events. In the third class are to be found those artificial transformations or mnemonics whereby associations are produced by "... a voluntary effort of the mind". Abercrombie reports that Feinagle managed to remember the birthdate of Henry IV (1366) by changing it into letters (mff) and then transforming that string into the word 'muff'!

Aristotle drew a sharp dividing line between class I (logical or thesauric arrangements) and the other classes; these latter fortuitous connections were the only ones that he referred to as associations. It is of course, pleasing to observe that this fundamental distinction is once again coming to be honoured even in experimental investigations (Fischler, 1977).

Expansion, contraction and coding within the library.

We have stressed in the previous section the notion of efficient storage and search, in which the strategies used in the latter process should match the principles employed in the former. Many scholars have argued that rational structures are preferable to fortuitous ones and Singer (1976) has drawn attention to Locke's claim that "association was a process which interfered with rational thought and in excess could lead to madness".

Efficient search procedures are, of course, of supreme importance in the case where there resides in an ancient and extensive library the only 'copy' of a precious manuscript. (Recall Bowen's remark about the efficiency of the staff of the British Museum.) This unique object can be in only one place at one time, although it may be transferred from one place to another.

In a similar fashion, searchers for the 'engram' (Semon, 1923; Lashley, 1950), or neural locus of learnt information have conjectured that particular experiences or knowledge are stored in a unique locale in the central nervous system. Lashley (1950) remarks that many investigators, including Henschen (1920-22), had proposed " ... the location of single ideas or memories in single cells". (Searching for such an idea is presumably analogous to looking for a needle in a haystack.) If the lateral aspects of the temporal lobes are stimulated with the fine tip of an electrode it is found that recall of complex sequences of past events may be elicited from discrete points (Penfield, 1968). And some scholars believe that skills and concepts may become transferred to other parts of the brain consequent upon damage to their prior and proper locale. The general validity of the notion of "... shifting and changing place in the Repository or Organ of Memory ..." (Hooke, 1682) is still debated, however.

Considerable interest has always been aroused by the varied possibilities of transferring information within and even between brains. In the original experiments on this latter topic, mathematical knowledge was transferred in the following manner: "The proposition and demonstration were fairly written on a thin paper, with ink composed of a cephalick tincture. This the student was to swallow upon a fasting stomach, and for three days following eat nothing but bread and water. As the water digested, the tincture mounted to his brain, bearing the proposition along with it" (Swift, 1735).

It has been reported that advances in technique now permit "fear of the dark", "left-right discrimination", "suppression of the startle response", "black-white discrimination" and many other simple concepts or behaviours to be moved from one brain to another by means related to

Swift's proposal. A complex nucleic acid or protein is extracted from the brain of a donor animal (which, for example, already fears the dark) and injected into the brain of a recipient, which as a consequence is itself now afraid of the dark. (A calm appraisal of some of this work will be found in Gurowitz, 1969.) To some extent, such fancies have been rendered more plausible by Hering's hypothesis (1870) that "... the phenomena of memory and of heredity have a common source" (Butler, 1880).

Finding a single nerve cell or molecule in the brain presupposes that the search strategies employed are highly efficient. There is, however, one situation where clever searches may, as it were, not be worth the effort. The situation is where there are many identical objects of search, multiple copies, any one of which will meet the requirements of the seeker. Theories of memory can accordingly 'trade off' the efficiency of search against the number of replicas that are stored, and so preserve on the right-hand side of the equation a constant, empirically determined time for retrieving a certain piece of information.

The notion of verbal facsimiles arises when (conscientious) scribes copy a manuscript; the notion of pictorial replicas arises when artists take a series of prints from a woodcut or an etched plate. More modern methods of mass reproduction which have resulted in the printed book, photographs and gramophone discs serve to weaken yet further the idea that a stored record must be a unique object. These new cultural objects were incorporated into psychological theory in the nineteenth century. Delboeuf (1880) suggested that 'L'ame est un cahier de feuilles phonographiques' and Guyau (1880) elaborated an extensive theory of auditory memory from this metaphor; Burnham (1888) summarized a range of models for visual memory based upon an analogy with photography.

It was Freud (1896) who then seized upon the idea of multiple copies that was so important in the commercial possibilities of the new technology. He writes to Fliess that " ... what is essentially new about my theory is the thesis that memory is present not once but several times over ... ".

This attractive concept does, however, have drawbacks of its own. Recall that for Freud the memories in question are physical objects - configurations of neural nets - and that the brain is finite in size. (Earlier notions to the effect that the soul was infinitely extended in space and time held little attraction for nineteenth-century materialists.) The disadvantage, then, is that if very many copies are made there may be too little space in which to store them whilst yet allowing room for the registration and storage of new information. The problems of over-printing are already explicit in Plato's wax block; proponents of mnemonic techniques similarly warned against the dangers of overcrowding: "We must take heed that we overcharge not our memory ... the number of things that may be committed at once unto a man's memory by this Art, are six and thirty, which are abundantly sufficient for the memory to be charged withall at once" (Willis, 1621).

Hooke's lectures to the Royal Society in 1682 contained some elaborate calculations on the number of impressions that the organ of memory might contain (Waller, 1705). Hooke postulates a generative centre for concepts, and around it "... a certain Sphere of Capacity fill'd with adapted Matter, for the Formation, Reception, and containing of all the Ideas which shall be emitted from the said Center." He continues: "These Ideas I will suppose to be material and bulky, that is, to be certain Bodies of determinate bigness, and impregnated with determinate

Motions, and to be in themselves distinct; and therefore that no two of them can be in the same space ...".

Hooke computes an upper bound on the number of thoughts and impressions that "... a Man of ordinary Constitution" may form and "store up in his Repository". The figure he arrives at is "a thousand Millions of distinct Ideas"; but this is based on a life span of a century and the formation of 2.25 ideas a second throughout that time. A more realistic lower bound is needed. Hobbes considers what proportion of time is lost through Sleep, Infancy, Old Age, Sickness, and Inadvertency, and what proportion of the ideas that are formed nevertheless fail to obtain a permanent place in the repository. His final conclusion is that the number of ideas stored is somewhere between 3,652,500 and 10,957,500, smaller admittedly than the original estimate but still a formidable library to house in one little brain.

Hooke mentions the fact of individual differences in memory capacity, but it was left to Gall (1791) to suggest the most straightforward explanation. Gall hypothesized that the size of the cortical organ responsible for the faculty was directly correlated with the individual's talent for remembering. Inferences from the size of a cortical area in a particular person to his relative psychological ability have recently become popular again (see Geschwind, 1972).

But a more interesting solution to the problem of limited space may be found in the notion of coding. Whenever space, time, or money is at a premium it is usually possible to find a more compact version of a message - a short code - which nonetheless enables all the information in the original to be retrieved. One of the first short codes was intended to be deployed on the wax tablets that inspired Plato's first model of memory. We have already mentioned that in an attempt to merge Simonides' technique with Plato's theory, Cicero suggested that the places could be regarded as a writing tablet and the images as the letters inscribed thereon. The inscription of full forms is, however, rather inefficient when time and space are limited; Cicero therefore introduced to Rome the new technique of writing shorthand symbols or notes (notae), and he took this innovation over into the theory and practice of remembering.

Tachygraphy accordingly became the first model of a short code that both speeded the intake of information and economized on the storage capacity required in the organ of memory (Milne, 1934; Yates, 1966).

Cicero was also involved in the development of the other great class of techniques that led to modern concepts of information-transmission and storage; the methods deployed in the art of cryptography (Pratt, 1939). Transposition ciphers were known to the Greeks and the substitution cipher was invented in the time of the late Roman Republic. Ciphers, however, do not economize upon either transmission-time or storage-space; the history of warfare, diplomacy, and espionage has accordingly seen the rapid rise of special purpose codes which "... express more in less space" (Pratt, 1939). For example, a single letter or digit in a code may represent a word, phrase, sentence, or even longer message. In order to avoid the arbitrariness of simple substitution codes, it was necessary (as it was in the evolution of shorthand) to develop the notion of redundancy, and to devise procedures whereby frequently occurring words or letters can be expressed by short sequences of symbols. (The code developed, for other purposes, by Samuel Morse is a good example of the principle involved.) As Oldfield (1954) puts it, "... recording with consequent economy is only possible if there are some recurrences, uniformities or common patterns in the incoming messages."

Subsequent advances were made by telephone engineers (Shannon, 1949) and quickly incorporated into the main stream of psychological theory (Miller, 1956b).

In the period between Cicero and the advent of modern telecommunication systems, the idea of encoding was also kept alive by the mnemonists (Grey, 1730) in an amusing inversion of the original idea. Typically in cryptography, a sequence of words is encoded into a digit string; but the mnemonists, of course, reversed this process, encoding random strings of digits into meaningful words and phrases in order to display their virtuosity in memorial power.

Once the metaphors of shorthand and cryptography had taken root (sic), we see that recall may be construed as a 'reconstructive' process rather than the literal retrieval of an exact copy of the original impression (Bartlett, 1932). Thus, for Bartlett, the rememberer (as opposed to the memorizer) "... has an over-mastering tendency simply to get a general impression of the whole and, on the basis of this, he constructs the probable detail". This notion has recently been rediscovered by members of the Artificial Intelligence community who have renamed Bartlett's schema 'frames'.

In practice, all messages are liable to error in encoding, transmission or decoding; messages that are recalled by 'imaginative reconstruction' (Bartlett, 1932) are presumably more prone than others to such error. In order to obtain data on message distortion Bartlett introduced two variations on the theme of presenting subjects with a story or picture for recall. In the first variation - repeated reproduction - a subject would be asked to recall the same event over and over again at varying time intervals; in the second variation - serial reproduction - a chain of subjects would pass a message from one to another at varying time intervals. With the former procedure, Bartlett found that frequent recollections at short time intervals would rapidly become fixed in form; recollections more widely spaced in time underwent a process of transformation, however. The material would, for example, become simplified and familiar detail would be substituted for unfamiliar. Similar distortions were also found with the second procedure; stories would become shorter, more coherent, and appeared to retain only those characteristics readily assimilable into the background of past experience and culture held in common by all members of the chain. Bartlett's work thus provides an elegant commentary on the old story in which the message from the front line - "Send reinforcements, we're going to advance" - arrives at General Headquarters as "Send three and fourpence, we're going to a dance".

The ubiquity of errors leads to the notion of error-detecting and correcting codes (Hamming, 1950), and the proposal that one solution to the problem of coding errors is to have the 'clear' in a number of (uncorrelated or only partially correlated) encoded forms; the ease with which simple codes can be broken (and must hence be changed by the sender) leads likewise to the idea of multiple representations in which the <u>same</u> clear can be expressed in a variety of codes.

In the years between 1891 and 1896, Freud studied in considerable detail the implications of thinking of the peripheral and central nervous system as a vast ciphering and coding machine (Marshall, 1974). In his monograph on aphasia, Freud (1891) characterizes the mapping of peripheral nerve tracts to higher cortical representations by saying that the cortex contains "... the body periphery ... as a poem contains the alphabet, i.e. in a completely different arrangement serving other purposes, in manifold association of the individual elements, whereby some may be represented several times, others not at all." He further notes that language is "overdetermined" in the sense that the system is represented by a variety of codes that safeguard speech against error caused by "... loss of one or the other element".

In his studies of hemiplegia, Freud (1893) distinguishes explicitly between projection paralysis in which there is a "... point by point and element by element" mapping between the periphery and the spinal cord (i.e. a simple cipher), and representation paralysis in which the mapping between spinal cord and cortex is many-to-one or one-to-many (i.e. the final set of representations can be thought of as context-sensitive codes). In letter 52 to Fliess (Freud, 1896), he proposes that memory-traces are "... subjected from time to time to a re-arrangement in accordance with fresh circumstances - to a re-transcription".

Freud refers to these successive transcriptions of psychical material as "translations", and he calls the phenomenon of repression "a failure of translation" from an unconscious memory-trace to conscious verbal recall. The memory-model outlined in the Project (Freud, 1895) is one of qualitatively distinct levels of representation which are differentially available to other levels and to consciousness, and differentially affected by error.

A record of the past is no use if it cannot be 'read'. (Recall Freud's interest in the decipherment of unknown scripts and unidentified languages.) A record is worse than useless if it has been falsified or incorrectly deciphered. (Recall the evil which resulted from the treachery of Captain Esterhazy and Colonel Henry.)

Conclusions and preface.

Whilst it is obvious that we have now collected far more facts about memory than were available to the Greeks, it is less clear that any new theories have been put forward in recent years. Experimental research, both psychological and physiological, in the nineteenth and twentieth centuries has been primarily devoted to working out and quantifying the empirical consequences of a small number of metaphors that were formulated at the birth of formal psychology. At times it has seemed that various aspects of the models have been forgotten only to reappear again in a flurry of data-gathering. The idea of dual-coding - verbal and visual - is a good example of such reemergence (Paivio, 1974), as is the notion of 'levels of processing' (Craik and Lockhart, 1972). New deductions are constantly drawn from the old models. That objects which are particularly important to on-going reasoning can be placed in an easily accessible reading room leads to the construct of 'working memory' (Hitch and Baddeley, 1976); likewise, the enterprise of making précies of documents immediately they are received leads to the hypothesis that changes in form may result not only from reconstruction at retrieval but also from abstracting at input (Gomulicki, 1956; Zangwill, 1972).

The possibility of combining the different models has provided some entertaining mixed metaphors. For example, in the mid-1960s a combination of the architectural and the coding model was briefly considered. The theory consisted of a large anteroom, a very small corridor, and an exceptionally large storehouse; as objects were passed from front to back, raw sensory impressions were transmuted into a phonological code, recycled through the corridor a few times whence they finally emerged into the storehouse in semantic form. (Belief in this solera-system for the making of long-term memories is no longer widespread.)

As is usual in psychology, particular aspects of memory models have often been given a new lease of life by advances in technology. Thus the notion of multiple copies has recently come to the fore-front again as a consequence of developments in optical holography. The virtues of holograms as memory devices are transparent. Chopping (1968) lists three: "(1) Many patterns in the storage area can be superimposed and decoded independently. (2) Ablation of large portions of the storage area will not drastically impair function. (3) Part of the input pattern will serve as the code for extracting the remainder of the input pattern." Holographic models of memory thus provide one solution to the problem of storing multiple representations in a limited space; their current popularity vindicates the fruitfulness of Lashley's conjecture (1929) that neural interference patterns are a basic form of memory coding in the brain.

Insofar as theoretical advances have been made, they seem then to have taken place in the area of coding theory (Von Neumann, 1958). Of particular importance are those aspects of coding theory that are concerned with content-addressable memories (Bigelow, 1965; Miller, 1968), and with the attempted specification of holographic processes realizable in the neural realm (Westlake, 1970; Cavanagh, 1975).

Throughout the modern period, philosophers from Plotinus (205-270 A.D.) to Malcolm (1977) have tried to undermine Plato's metaphors, without any great success even within the terms of their own discipline (Rosen, 1975). The stature of Plato's work is amply attested by its power to have sustained empirical research for two millennia. Psychology is not a young science; it is merely a difficult one. Yet one may hope that eventually someone will hit upon a radically fresh idea about the dynamics of memory. We conclude then with this preface that looks forward to ...

# CHAPTER ELEVEN

## THREE CONTEMPORARY DOMESTIC MODELS.

"Against Solemnity, the best weapon is wit. Most other weapons produce only another dogmatic sectarian solemnity. I have tried to avoid this danger, though I must confess that I have not always been successful in this endeavour."

Bertrand Russell.

As far as theoretical cognitive psychology is concerned, then, the organism, being a computational system, is fundamentally a machine; and to explain it is to give the principles underlying those aspects of its functioning construed as handling information.

This, it is herein suggested, is the commonly held conception of explanation which unifies the work of all cognitive simulators, whether they worked in 18th century France or 20th century America.

Another, inverted but equally valid, way of putting it is to say that all machines are computational systems.

This is not an outrageously modern view. In 1887, Peirce wrote: "every machine is a reasoning machine, in as much as there are certain relations between its parts, which relations involve other relations that were not expressly intended ...it does not depend on the laws of the human mind, but on the objective reason embodied in the laws of nature." Fodor (1975) has submitted this type of view to close scrutiny and he too concludes that any complex system which changes physical state in some way determined by physical laws is a computer "just insofar as it is possible to devise some mapping which pairs physical states of the device with formulae in a computing language in such a fashion as to preserve desired semantic relations among the formulae."

The view under consideration suggests, then, that at the core of the cognitive process lies a machine. Which machine? Any machine?

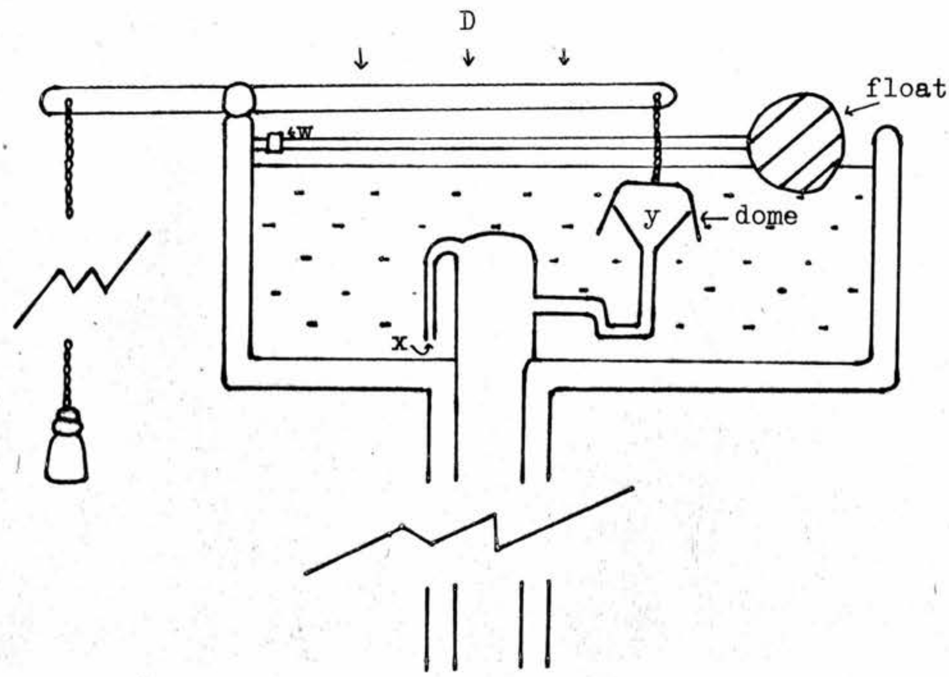
As Fodor & Block (unpublished ms.) have written: "if we were willing to tolerate very complicated and unnatural functions of sentences onto physical states, practically any middle sized physical object ... could, in principle, be employed to compute any of indefinitely many proofs; but ... in such cases designing the semantic function would be the real work in designing the computer."

In this chapter, we take a passing glimpse at three everyday middle sized objects with an eye to their suitability as mechanical models of cognitive capacities.

1. A Finite Chain Reasoning Machine.

A simple machine which we each use many times each day is the lavatory cistern. How much reasoning is it possible to persuade such a machine to perform?

Imagine an idealised cistern:-




Now we can label:

'A' ... 'The handle is pulled.'

'B' ... 'The dome rises.'

'C' ... 'The pressure at y decreases.'

'D' ... 'The air pressure on surface is constant.'

'E' ... 'Water is forced into x.'

'F' ... 'Siphon action commences.'

'G' ... 'Water level drops below x.'

'H' ... 'The float lowers.'

'J' ... 'Valve w releases water.'

'K' ... 'The float rises.'

'L' ... 'Valve w stops releasing water.'

'M' ... 'Pressure rises in y.'

'N' ... 'Siphon action stops.'

etc.

We are now in a position to give (an approximation towards) the (an) underlying or logical form of the flushing loo.

i.e.

$$\text{I}\begin{cases}(A \rightarrow B)\\(B \rightarrow C)\\(D \,\&\, C \rightarrow E)\\(E \rightarrow F)\end{cases}$$

and transitively $(A \rightarrow C)$.

therefore $(D \,\&\, A \rightarrow F)$

$$\text{II}\begin{cases}(F \rightarrow G)\\(G \rightarrow N)\end{cases}$$

$$\text{III}\begin{cases}(H \leftrightarrow J)\\(K \leftrightarrow L)\end{cases} \quad \text{or} \quad \begin{matrix}(\neg K \leftrightarrow \neg L)\\(\neg H \leftrightarrow \neg J)\end{matrix}$$

$$\text{IV}\begin{cases}(F \rightarrow H)\\(H \rightarrow J)\end{cases}$$

n.b. a self regulating system.

But:

$$\begin{cases}(J \rightarrow K)\\(K \rightarrow L)\end{cases}$$

Also to be found in the system is:-

$(J \vee \neg J)$ $\quad [(H \,\&\, J) \quad \vee \quad (K \,\&\, L)]$ etc.

(As can be seen, the ontology can be somewhat reduced by using negatives e.g. $K = \neg H$. etc.)

Notice, we can now illustrate in the loo cistern:-

1) The logical operators ... $\&, \vee, \rightarrow, \leftrightarrow, \neg$ .

2) Propositional signs A....N.

n.b. Wittgenstein (1921) wrote: "The essence of a propositional sign is very clearly seen if we imagine one composed of spatial objects (such as tables, chairs and books) instead of written signs.

Then the spatial arrangement of these things will express the sense of the proposition."

A New Cistern of Thought.

The machine is now ready to reason. All we have to do is feed in the propositions relevant to the matter in hand, flush the cistern calculus and see what it churns out.

In a sense, of course, it is limited, for there is only one well formed sequence of events (operations) through which it can pass. In another sense, though, this is its strong point, for insofar as we are certain of the physics of the machine, we are certain of its reasoning ability.

The well formedness of its strings are guaranteed by plumbing!

Illoostration of Functioning.

It is by no means easy to find a reasoning task for which this unique computer is (in its entirety) suited, but we can do so for fragments of its process, with the aid of a little story-telling.

Suppose we were visitors from Mars ... entirely naive as to terrestrial characteristics: gravity, climate, horticulture etc. ... but wanted to take up gardening! Luckily, along with all the implements, seeds etc., we find five illustrations torn from a gardening book but in random order. We might conclude that there are five essential elements in terrestrial horticulture (corresponding to the pictures):-

i) Sunshine etc. obtains.

ii) The ground gets wet.

iii) The plant grows.

iv) The water soaks down to the roots.

v) We water the garden.

Our reasoning (logic) being Martian is totally inappropriate (on Mars we water the plants after the harvest !), our experience of terrestrial growing nil, our supply of seeds too meagre to allow trying all permutations;

what shall we do?

Happily, but as always with fairy stories, a kindly Percy Thrower has left behind in his potting shed a cistern reasoning machine and a translation manual. Now by means of translation functions:-

A ⟷ 5 C ⟷ 4 E ⟷ 3

B ⟷ 2 D ⟷ 1

we can work out (or rather the machine can) what to do. We simply pull the chain and watch.

One further point to note, Part IV instantiates the logical form of many self regulating systems e.g. a thermostatic immersion heater, where:- F = Water is cool; H = Heater is on; J = Water is warm; K = (¬ H); L = (¬J).

<u>Notice</u> (cf. Fodor.)

1) The conceptual power of the machine is (in principle) limited. It is as impossible to illustrate a piece of predicate calculus on a propositional lavatory cistern, as it is to represent predicate calculus in propositional calculus.

To belabour the point; imagine that we are all born with an innate lavatory cistern cunningly disguised as cerebral hemispheres; and that learning a language is a matter of hypothesising (and confirming) states of the cistern as metalinguistic truth-conditional translation functions of the object language; then, manifestly, we cannot learn any language which transcends in conceptual power (i.e. can't be represented in) our cistern.

2) It is possible to change the logic (underlying form) of our system not by concept learning (which uses the old logic) but by using a spanner or hammer.

Concluding Remarks.

It would not do to get too bogged down in this analogy. It may seem to some that the suggestion that we are born with a siphon action lavatory cistern in our heads is too potty for words. But notice, strained as it may seem to suppose we learn propositional calculus by mapping it onto machine states of a flushing loo, we do, of course, more mundanely, learn how flushing lavatory cisterns work ... at least I hope we have just done so! Naturally, if we can come to learn how they work (and we surely do), we must have an internal language for representing those very hypotheses that have been confirmed.

Still, my hopes for a cisternmatic account of thought are, I confess, low; for, even should other problems be solved, someone is bound to come up with Wittgenstein's conclusive refutation ... the privy language argument.

2. A modern Soap Opera(tion).

I have argued earlier that the prevailing technology of the day and the psychology of that day are closely linked ... that is to say that technological and psychological theory go hand in hand. But who is to say which leads which? Mere chronological precedence, even if conclusively demonstrable, cannot convince those prepared to entertain the possibility of backwards causation (Beloff, 1977).

To come clean from the outset, the possibility has suggested itself to me that psychological theory dictates technological advance.

If this is so, we may be indebted to William James (1890) for that lyrical instrument, the harp. Long contemplation of the phenomenon of dualism, led him to speculate that, if his theories were correct, a string should vibrate releasing a non-interactive melody:

"The mind-history would run alongside of
the body-history of each man, and each
point in the one would correspond to,
but not react upon, a point in the other.
So the melody floats from the harp-string,
but neither checks nor quickens its vibrations."

Similarly, we may be indebted to Plato for the invention of writing. Consideration of memory; its psychological problems - the way our ideas melt into and become confused with each other when the brain is overheated by fever - and physical instantiation - the evidence of impressions led Plato, in The Theaetetus, to postulate that if we could find something in the outside world sufficiently similar to brain matter, we might have an 'external memory'. Thus he hit upon wax tablets and the notion of recording by making impressions upon them (and of erasing by melting.) C.f. in this connection the notion of imprinting.

However we need not delve far into the past to prove the thesis, but may consider an everyday object, which embodies the influence of psychological theory on technology.

The object in question is a cuboid of dimensions (approximately) 4' by 3' by 2'. It is constructed of rubber, plastic and steel, scantily covered in white enamel, and stands on castors. The bulk of the model consists of two compartments with hinged lids, complicated machinery (in line with current mechanistic views of brain function) and pipes for conveying fluids (hydraulic models last attained popularity with Lorenz' model of motivation) complete the gross picture.

Some little wit is needed to decide just which aspect of psychological theory has influenced this artefact. The connection is revealed in an enigmatic, not to say cabbalistic, fashion in the phrase "Machine Eliminating Mud OR Yellowness." As can be seen, the initial letters form the word: "MEMORY".

Confirmation is supplied by other clues often stored near the model ... cartons containing biological washing powder.

In what sense is the machine a simulation of memory? The twin tubs are obviously illustration of the two hemispheres and the fact that the machine works as an electro-chemical process alludes to the current beliefs about the brain's mode of functioning. The brain model also instantiates a relatively sophisticated simulation of the homeostatic principle in its thermostatic heating control.

The more interesting part of the analogy, though, is that specifically relating to memory. Obviously, as with ideas, one puts objects (in this case, articles of clothing) into the machine, processes them, temporarilystores and then retrieves them. Just like ideas remembered, the clothes tend to shrink and become faded. Moreover, just as in a fever memory is more likely to be impaired, if the water in the machine gets too hot, articles of clothing are even more likely to shrink, become distorted or even destroyed. The colours of clothes of different hues put in together run, thus simulating the confusions amongst memories as a result of association. Partly for this reason, some semantic grouping of clothing is required ... whites, linen and woollens, for example, being grouped together and kept to some extent separate ... just as ideas, or more strictly those structures which represent them, are semantically grouped in the brain.

When the cycle of operations is over, the machine switches itself off and sounds an alarm ... thus explaining that mystifying phrase from memory folklore: "it rings a bell."

There are always those who will refuse to be convinced by any number of examples however clean cut those examples are. These sceptics will refuse to be dazzled by these few sheets, preferring to swim against the tide and will cling to the more conservative view that it is the technology which influences the psychological theory ... for these this brief note will be a wash-out.

It is true, I own, that argument by analogy is not foolproof (some might say that is the only thing this author demonstrates) but we may do worse than note Samuel Butler, who wrote:

> "Though analogy is often misleading,
> it is the least misleading thing we have."

3. Are Computer-Programmers Knit-wits?

Arguments against the possibility of machine intelligence are often woolly and emotive, but occasionally an apparently more formal argument is offered. For example, it is frequently said of computers, as if to prove they cannot transcend the capabilities of their makers, that they can only do what, and as much as, they are told or programmed. This pearl of wisdom can be expanded thus: to solve a problem, a computer must be programmed; and to write a successful program a programmer must (barring lucky accidents) not only understand the principles of solution, but understand them in some explicit way and be able to formalise them so that they can be represented as a series of instructions. Plainly, the role of the machine is, then, limited to mere brute force and blind following of some algorithmic routine.

The news that computers are completely slavish and obedient to their desires and intentions often comes as a surprise to, and usually needles, novice programmers, whose experience of computers and the difficulties of debugging invariably suggests that the programmed machine will do anything but what they intend it to. However, there is only a trivial point at issue here - of course the machine will only do what it is programmed to do, the difficulty lies in translating your intentions and desires into a program.

The more substantive point is that it is not strictly true that a computer can only do what its programmer tells it to do, rather it can only do what its program tells it to do. Admittedly, in our experience, programs are usually written by programmers, so these two assertions are often contingently true - but this is not necessarily so! Realising this, leads one to allow the possibility of a naturally occurring programmed computer - due to accident or natural process e.g. evolution (the human brain), or of a naturally occurring program which transcends the power of the human programming. The assertion that the human brain is a computer, then, does not entail the existence of some great programmer in the sky (although, presumably, it must be initially - biologically - programmed, and in view of its flexibility, must be initially programmed to subsequently program itself!)

How though, more precisely, are we to understand the activity of programming?

The essence is to take a complicated task and to show how it can be accomplished by performing simple tasks in a specified order. Many analogies can be, and are, drawn. Recipes, for example, are programs. Given a complicated task like the making of a soufflé, the effective cookery book reduces it to a series of simple operations (beating of eggs etc.) in a specified order. Of course, what counts as a simple or complicated operation varies according to the expertise and aims of the trainee - we may need to consult another egg-beating program, or on the other hand, may count making the soufflé as only one operation in a five-course-dinner-preparation program.

In her recent book, Margaret Boden (1977) chooses an interesting alternative analogy ... the knitting pattern as a program. Typically, a knitting program takes a complex task, e.g. the knitting of a baby's boottee, and reduces it to a well-formed string of simple operations or primitives; typical primitives are "st", "S1", "k", "wlfwd". Anyone unfamiliar with these primitives must consult a beginners manual for still simpler programs explaining e.g. "k" (knit) in terms of yet simpler operations e.g. finger/needle manipulations. For further elaboration of this metaphor, see Boden '77.

Note, in both of these analogies, and in programming itself, "the exercise of a complex capacity is analysed as the organised exercise of relatively less problematic capacities"(Cummins, 1977). However, what is of particular interest here is the extent to which the manufacturing of knitwear and textiles occurs as an analogy of the mental process.

During the Second World War, work, still today officially secret, was done on early electronic digital computers at Bletchley Park, where the core of the British Government code breaking effort was also centred. Refusing to have the wool pulled totally over his eyes, Randell (1976) has pieced together available fragments of information to reveal that, at Bletchley Park, there were built "machines called bombes which were prototype computers" (Calvocoressi, 1974). A.C. Brown (1976) also refers to a machine called "The Bomb" designed by Turing:"a copper coloured cabinet some 8ft. tall and perhaps 8ft. wide at its base, shaped like an old fashioned keyhole. And inside the cabinet was a piece of engineering which defied description." This machine, still shrouded in secrecy, is an enigma as regards its functioning and purpose, but significantly, as its Chief Engineer, Harold Keen, said: "It's initial performance was uncertain, and its sound was strange; it made a noise like a battery of knitting needles as it worked ..." (in Brown).

Sherrington (1940) refers to the brain as an "enchanted loom" and in a similar vein Bennett talks of "the facts of today" as "so much raw material from which my brain has to weave a tissue of life that is comely." Analogously, Longuet-Higgins (1968) says "the computer is like a loom" and takes us back to the issue of programming: "a loom will weave nothing until the weaver comes along and supplies it with the pattern of the carpet his client has ordered."

In his Gifford Lecture of 10th March, 1978 ('Learning and Memory'), J. C. Eccles compared the parts of the hippocampus concerned with memory with antique oriental rugs, which his wife collects. "We live with patterns", he said, and proceeded to draw the analogy between patterns on rugs and patterns in cells, even being so specific as to compare the estimated two million knots on an antique oriental rug with the estimated four million functional units of the hippocampal area!

The link between looms and computers (artificial and natural) is again stressed by Peirce (1897): "The study of how to pass from (a hand loom simulation of logical reasoning) to one corresponding to a Jaquard loom would be likely to do very much for the improvement of logic." Actually, as has been pointed out earlier, the attribution of the invention of the automated loom to Jaquard is quite spurious - the actual inventor was, of course, Vaucanson (Bedini 1964). Early computers and automated looms actually shared some of their technology. The former made use of punched tape which allowed (or hindered) crucial pins to pass through and the latter, of course, also used punched tape but replaced the pins by pulses of light. (Vaucanson's apparatus for the automatic weaving of brocade of around 1750 was not even the first use of such methods - records exist of punched tape being used as a memory store as early as the 16th century in Augsburg hodometers.)

As can be appreciated from this brief note, it does not take a warped mind to see links between textile manufacture and intelligent functioning. However, no attempts are made here to stretch this metaphor beyond its bounds and lead the reader surreptitiously into too close an identification of looms and brains. This would, no doubt, be foolhardy. As Scott has pointed out: "O, what a tangled web we weave, When first we practise to deceive."

# CHAPTER TWELVE

## SOME MODELS WHICH GET ON ONE'S NERVES.

> "The history of ... (the) nervous impulse reflects the history of the most fundamental stages of medicine, if not of human thought and knowledge in general, passing from metaphysics and speculation to physics, observation, and experimentation."
>
> Walter Riese (1959).

> "There was no burst of good work in scientific physiology right after Harvey nor in fact in the 18th century."
>
> Edwin Boring (1950).

In the immediately preceding seven chapters, we have examined first some formal, and then some intimate, models in some detail. For the purposes of exposition and ordered discussion, it has been useful to divorce these models somewhat from their respective research areas, and to discuss them in relative isolation. Naturally however, in the course of actual research on particular phenomena, simulators do not rigidly adhere to the use of either an intimate or a formal model - typically both sorts of model are used by different modellers at the same time; the same modeller at different times; different modellers at different times etc., etc.

In this chapter, we deal with models ... from the point of view of the phenomenon under study, rather than from that of the model.

We take a look at the influence of models on the accounts given in a domain of particular interest to physio-psychological theory ... that of how impulses are conducted along nerves and, indeed, of how we are to conceive of nerve impulses at all. Concerning this topic, it is natural to start with René Descartes - not because he was the first to propose a machine model of man, others had already beaten him to that e.g. Aristotle (De Mota Animalium): "The movements of animals may be compared with those of automatic puppets ... or with the toy wagon." - but rather for two other reasons. Firstly, because it was only from the time of Descartes that the question of how nerves conduct began to be asked in isolation from the question of how muscles contract ... earlier these two questions were dealt with as one (further details later); secondly, because of the almost incredibly powerful influence that Descartes' direct contact model of causality has had in neuro-physiological theories of nerve conduction.

In this chapter, then, after a brief exposition of the foundations of Descartes' influence, we discuss some interesting models of nerve impulse conduction during the century and a half commenced by Descartes' seminal Treatise of Man.

As Huxley (1874) has observed "the fundamental proposition of the whole doctrine of scientific Physiology", "the idea that the physical processes of life are capable of being explained in the same way as other physical phenomena", is owed to René Descartes, that "great and original physiologist", who did for the mechanical theory of motion and sensation what Harvey did for the blood.

The Cartesian influence has indeed been profound and far-reaching; the basic model he proposed has had universal appeal and has virtually dictated the subsequent 300 years of physiological research. Let us be rather more explicit: Descartes' works are not a series of isolated if brilliant forays into disparate intellectual fields. Although it is true that he made advances of genius in mathematics, physics, metaphysics and physiology, one profitably views Descartes' efforts as expressing a coherent vision of the nature of man and the universe, a vision based mainly on his theory of optics, which, for Descartes, had replaced the classical conception of celestial mechanics as the key to nature (Caton, 1973).

As is now well recognised, Descartes (1733) conceived of the body as a thinking machine, all of whose "functions occur naturally ... solely by the dispositions of its organs, not less than the movements of a clock or other automaton." A central problem of this thinking-machine account was that of how the mediation between sensation and movement was to be explained.

It is crucial to realise that in his theorising Descartes was starting with a completely clean slate. His theory was intended to explain the interaction between two apparently quite different spheres ... the psychobiological one of neuro-mental events and the natural scientific one of optical physics, and, although Descartes owes something to Ibn al-Haytham's (1039) visual optics (as did Witelo and Kepler), he was able to let his optical and psychobiological theories mutually constrain each other's development. (Since Descartes the situation has been much different, as Young and Burtt lament, for psychologists have been compelled - and in the main been content - to let a pre-established physics decide their ontology, leaving themselves the task of labouring merely on epistemological issues. The position is now beginning to change with the Gibsonian challenge - for elaboration see Reed, 1978.)

It is interesting that both Traité de l'homme and Traité de la lumière were completed in 1633, as parts of Traité du monde, sharing the same gestation period and fertilised by common sources.

Caton has persuasively argued that Descartes' theory of optics was his foundation stone - that having once fixed his theory of physics ... concerning the nature of light ... Descartes had to tailor his theory of man to fit in with this ontology i.e. Cartesian man had to be the sort of entity who, when struck by single rays of light composed of particles in motion (a percussive view of light with its accompanying contact model of causality), could behave (move) appropriately.

However, it is at least conceivable, and the possibility is interesting to consider, that in fact the influence worked the other way, i.e. that Descartes' theory of man as a machine influenced his optics. As early as 1619 (Cogitationes Privatae) Descartes proposed the construction of a man-machine worked by magnets and according to Père Poisson, he actually drew blue prints for automata in an attempt to verify his theories. We know (Additions A.T.XI 669) he was fascinated by the hydraulic statues at St. Germain en Laye and in L'homme he refers to such automata frequently. For example: "truly one can well compare the nerves of the machine that I am describing to the tubes of the mechanisms of these fountains, its muscles and tendons to divers other engines and springs which serve to move these mechanisms, its animal spirits to the water which drives them, of which the heart is the source and the brain's cavities the water main." It is at least conceivable, then, that it was at least partly the influence of the hydraulic machine model of man which led Descartes to propose a theory of the environment (optics) commensurate with such a model ... and that the only way of which he could conceive to affect a (hydraulic) automaton was by the involvement of direct contact causality ... hence his percussive optics and its direct métaphysical consequence - that extension is the only property of matter.

Be that as it may, from this point on the priority of optics can hardly be disputed. On the organismic level, optics certainly become the model for the other senses, while Descartes' entire physics is plausibly seen as an attempt to generalise his theory of optics i.e. general physical laws become rules describing the motion of percussing rigid bodies. Finally, on a metaphysical level, Descartes forges the bond between thinking (ideas) and being (the world) by the conception of corporeal ideas which occur at the interception of the principles of nature with the principles of knowledge ... thus identifying sensation and cognition.

For Descartes, the brain consisted of a cortex surrounding a cavity in which was situated the pineal gland. The cavity could expand or contract according to the influx of animal spirits - reaching its greatest dimensions during our periods of greatest awareness, contracting during sleep and closing altogether on death. (This explains why Descartes was unconcerned at never having discovered such a cavity. He spent several years dissecting - but never, of course, live animals.) The cortex itself was a 'tissue' of nerve fibres, pores and tubes which led to and from receptors and effectors to the inner surface of the pineal cavity. Some terminated, however, in the fibrous cortex where they could cause 'memories' by distending fibres to leave an impression.

The pineal gland itself was maintained in its central position in the cavity by being buoyed up on a fresh supply of animal spirits which were fed in by special pipes ... just as balls were buoyed up on jets of water in the hydraulic pleasure gardens which Descartes visited and by which he was so intrigued. The core of the model lies in the fact that there are orifices on the gland itself from which the spirits could gush (again like jets of water under pressure) and, crossing the cavity, enter into the tubes of the cortex and so eventually find their way to the effectors. In the case of perception, as opposed to action, the impulse caused by the percussion of optic particles on the retina is conducted to the inner surface of the cavity by a 'filament' composing the marrow of the nerve and surrounded by animal spirits. This filament is attached at one end to an 'organ of sense' such that it is "very easily moved by the objects of that sense" (e.g. bombarding light particles) and at the other to valves on the tubes surfacing on the inside surface of the cavity, such that the slightest sensation opens a valve allowing the animal spirits to flow. This then, results in a pattern of open tubes isomorphic to the pattern of retinal impact (itself isomorphic to the perceived object). This pattern is communicated intact across the void to the pineal gland and there results in a "figure ... which is traced in the spirits on the surface of the gland." (A.T. XI 176-7). This pattern is, transitively, isomorphic to the object, which allows Descartes the title of realist. Also in this way Descartes makes either a sensation or an imagination count as an idea, which he identified unambiguously with a brain process (it is truly corporeal).

The problems still remain, though, as to what "the conduction via the animal spirits" amounts to and of the relation between animal spirits and effectors (muscles). As Descartes (1649) puts it (P.S. Art VII): "We know ... that all these movements of the muscles, as also all the senses, depend on the nerves, which resemble small filaments, or little tubes, which all proceed from the brain, and thus contain like a certain very subtle air or wind which is called the animal spirits", but what we now need to know is "what is the corporeal principle which causes them to act" and "in what way (do) these animal spirits and these nerves contribute to the movements and to the senses.?" (P.S.VIII). As Harmon and Lewis (1966), amongst countless other commentators, have observed, Descartes "relied on previously existing hydraulic automata as models of his system to settle the issue." Descartes is quite explicit about this: "... spirits enter the cavities of the brain they also leave them and enter the pores (or conduits) in its substance, and from these conduits they proceed to the nerves. And depending on their entering (or their mere tendency to enter) some nerves rather than others, they were able to change the shapes of the muscles into which these nerves are inserted and in this way to move all the members. Similarly you may have observed in the grottoes and fountains in the gardens of our kings that the force that makes the water leap from its source is able of itself to move divers machines and even to make them play certain instruments or pronounce certain words according to the various arrangements of the tubes through which the water is conducted." (T.M. 21).

However, whilst both hydraulic automaton and man are, at an interesting level, isomorphic, they were not instantiated, of course, in the same material substance. The animal spirits and the automaton's water were both fluids, its true, but, to be more specific, the animal spirits were "the most animated and subtle portions of the blood which the heat has rarified in the heart" (P.S.X.) - i.e. it is worth stressing "nothing but material bodies ... of extreme minuteness" (P.S.X). (This notion of animal spirits as the product of the application of heat to blood perhaps reached its culmination in the 'explosion' model of Thomas Willis (1672). Willis, who actually coined the word "neurology" and who taught John Locke medicine at Oxford, observed that rectified oil of Vitreol mixed gently with Alcohol Vini produced a sudden effervescence releasing sufficient force to explode the vessel in which the mixing took place and conjectured accordingly that it was a mixture of acid animal spirits with oily blood which produced sufficient explosion to 'inflate' a muscle.)

As mentioned at the beginning, at least from the Ptolemaic period up to Descartes, the problem of the action of the nerve and that of muscle action were counted as a single issue. Even by the time of Eristratus (more than 1400 years before Descartes) the idea was well established that contraction of muscle was due to a swelling in its volume, an account even favoured by Galen, ... we can call this a 'balloon' model (possibly drawn from some natural balloon like a bladder or intestine?). Descartes: "these spirits ... although they are very mobile and subtle, they lack not the strength to inflate and tighten the muscles in which they are enclosed, even as the air in a ball hardens it and stretches the skin that contains it." (T.M.28).

The central conception is of the muscle as an inflatable flexible balloon filled with animal spirits which "distend and shorten it". The main point for us to notice is that in the traditional model the role of the animal spirits as inflat<u>ors</u> and of the muscle as inflat<u>ed</u> are indissolubly linked. However, a seldom commented upon, but important facet of Descartes' theory is that he was beginning to move away from this indissoluble link towards separating the question of how the nerve conducted from that of how the muscle contracted. He did this by making the muscle more autonomous than it had ever before been ... and thus paved the way for a notion of irritability.

He did this by insisting that the inflation of the muscles was not totally accomplished by the 'pumping' of the nerves: "Not that the spirits which proceed immediately from the brain" - of which there are "but very few" - "suffice in themselves to move the muscles," rather they bring about the movement of the muscle by affecting "the other spirits which are already in the two muscles." (1643). In other words, the animal spirits from the brain are transferring information rather than power and, in a sense, are telling the animal spirits in the muscle what to do.

Muscles are arranged in pairs such that: "when the spirits that come from the brain to one of them have ever so little more strength than those that proceed to the other, they open all the entrances by which the spirits of the other muscle can pass into this one, and at the same time close all those by which the spirits of this last can pass into the other" (1643). The main role of the brain animal spirits, then, is to open and close valves, i.e. to pass on a message rather than actually themselves pump up the muscles. This departure from the traditional balloon account, with its suggestion of the active self-contained nature of the muscle rather than a passive recipient of inflation, smoothed the path for a notion of intrinsic irritability.

Passions of the soul was published in 1649. In 1680, Borelli, leader of the so-called iatro-mathematicians, published his De Motu Animalium in Rome, supporting Descartes' view. Borelli's support took two forms. Formally, he offered some mathematico-mechanical models of muscle based on the rhombohedron, in which he demonstrated that if the edges of a rhombohedron are fixed in length, as the volume is increased, the distance between the opposite vertices decreases. On the more practical side he offered an alternative suggestion as to the nature of the animal spirits ... an alternative which can perhaps be best described to the modern mind as a 'toothpaste' model. (In fact, Borelli was thinking of the soft marrow which can be squeezed out of the sumach twig). As Borelli remarked: "nervous fibres ... may be hollow (but) appear to be solid and full to the simple vision ... therefore it is not impossible that the nerves are hollow fistulas ... tubes which are filled with some moist and spongy substance like the marrow of sumach twigs. We may suppose that the spongy hollows of the nerve fibres are always moist and filled to swelling with a juice or spirit which is received from the brain. And just as we see in an intestine, when it is filled with water and closed at both ends, that if one end is compressed and lightly struck, the motion is quickly communicated through the turgid intestine to the other end; so it would seem that a light compression or irritation made at one end of the canal, where it ends in the brain itself would cause the discharge of a small amount of that spongy substance or juice into the meat of the muscle itself." In this way, presumably, the 'toothpaste' became the signal.

The great Boerhaave (1744), too, added his support to the notion of both a "fluid body, very easily or quickly moved ... forcibly thrust into or applied to the muscle" (i.e. animal spirits) and to that of the dilation of muscle fibres "to reduce them from an oblong to a rounder figure, increasing their diameter and diminishing their length, so as to bring the tendons nearer to each other."

Even as late as 1897, McDougall was arguing that the contraction of the wing muscles of insects was caused by an increase in their volume due to their absorbtion of some of the fluid sarcoplasm which surrounds them. Meigs (1905) valiantly defended McDougall by offering, in answer to the question of how an increase in the volume of the contents of the fibril could cause it to become shorter, a model of a model - or more correctly the original model itself.

Meigs constructed a model of the fibril. It consisted of a thin rubber tube encircled by wire rings. Along this tube were inelastic cotton threads firmly attached to each of the rings. To the single open end of the tube Meigs applied an air pump - the air intended to model the sarcoplasmic fluid (but it was just as good a model of animal spirits or sumach marrow). When air was forced into the tube, each of the rubber sections between the rings changed shape from an approximation to a cylinder to an approximation to a sphere. As the cotton threads would not stretch, but were forced into curves, the overall length of the model perceptible shortened. Meigs proclaims triumphantly that: "A good deal of McDougall's argument is based on the fact that the form of contracted fibrils is somewhat like that of the contracted model" and then goes on to give histological evidence that fluid does pass from the periphery towards the centre of voluntary muscle fibres of frogs during contraction.

It is a tribute to the power of the balloon model that it has exhibited such a persistent and universal appeal ... and this in spite of the fact that, even by Descartes' time, it had been conclusively shown that muscle volume simply does not increase during contraction - as any Archimedean flexing his/her toes in the bath could have demonstrated. Francis Glisson (1677) actually did experiments which showed that, rather than inflating, muscle actually decreases in volume when flexed. Glisson also commented on the 'absurdity' of the idea of spirits running backwards and forwards in a system of tubes. Later critics tended however to ignore the former devastating point and concentrated on the latter technique of ridicule. Cheyne (1733) mocked the very idea of basing a theory of man on the model of an automaton: "The Similitude of a Machine put into Action and Motion by the Force of Water convey'd in Pipes, was the readiest Resemblance the Lazy could find to explain Muscular Motion by. It was easy, from this Resemblance to forge a thin, imperceptible Fluid, passing and repassing through the Nerves, to blow up the Muscles, and thereby to lengthen them one of their dimensions, in order to shorten the other. On such a slender and imaginary Similitude, the precarious Hypothesis of Animal Spirits seems to be built." A year later, another critic appealed to the more noteworthy experimental evidence. Robinson (1734): "But it does not appear from any Experiments that the Nerves are Pipes; or that such a Fluid as they conceive Animal Spirits to be is separated from the Blood of the Brain; and therefore these Opinions are without any just Foundations. The Nerves are not only impervious to the smallest Stylus, but when viewed with a Microscope, evidently appear to have no Cavity." This final point was echoed by Bertrand (1756) together with the further experimental finding that affixing a ligature to a nerve produces no swelling either above or below the ligature (to be expected on the animal spirits or toothpaste model).

The survival of the inflation theory of muscle action, despite all the contrary evidence, can only be attributed to the power of the motivating model - the balloon. The attraction of this model, and the way it suggested a single problem of nerve and muscle rather than two, even obscured Descartes' attempts to tease the two apart. However, Glisson did seize on the idea that the muscle was not purely passive and dependent on an inflating fluid, but contained within itself the wherewithal to actively contract (irritability) and therefore needed only stimulation to be set in action. This realisation in itself raised the further problem - of how the stimulus was to be transmitted by the nerves. It is worth making clear, however, that the concept of irritability in no way goes against the man-machine thesis. As Lups put it: "Irritability is the principle of movement in the human body, as the pendulum is in the clock." It is interesting to note that Lups' book De irritabilitate was published in the same year (1748), by the same publisher (Luzac), as La Mettrie's L'Homme Machine, and is dedicated to Haller, traditionally the originator of the concept of irritability, as, again, was La Mettrie's book - in the latter's case without Haller's permission. This may go some way towards explaining why Haller, unjustifiably, accused La Mettrie of having himself claimed the authorship of the concept. Meanwhile, Haller had his own theory about the passage of the nerve stimulus - a theory based on what Harmon and Lewis described as a croquet model! To be more precise, Haller (1768) compared the nerve itself to a row of small rigid balls, each in contact with the next on either side, and the passage of a nerve impulse is compared to that of a physical impact such as occurs when the first of a series of balls is struck and the last of the series flies off 'instantly'.

The models of nerve action as inflation, carrying of exploded gases, toothpaste or croquet are clearly what, towards the beginning of this paper, I called contact models of causality. It was part of the function of that earlier section to suggest that the contact model of causality was not just an aspect of Cartesian psychobiology, but, via his optics, a fundamental principle of his entire world picture. The antithesis of the contact principle is, of course, that of action at a distance. As a principle of physics, this was most importantly put forward by Newton, against his own inclinations for he was far from happy about it, in his concept of gravity. As for his views on psychobiology, however, Newton was fully if somewhat eccentrically within the Cartesian fold. Thus, Newton accepts the basic terms of the Cartesian argument; he postulates a "most subtle Spirit" (M.P.N.P.) and asks (Opticks Qu. 23): "Is not Vision perform'd chiefly by the Vibrations of this Medium, excited in the bottom of the Eye by the Rays of Light, and propogated through the solid, pellucid and uniform Capillamenta of the optick Nerves into the place of Sensation?" Then, in true Cartesian fashion, he gives an identical analysis to Hearing and finally generalises the account completely ... "And so of the other Senses." Moreover, these "Vibrations" travel along the nerves to the "Muscles, for contracting and dilating them." (Qu. 24).

However, although the basic design of the Newtonian nerve-muscle system is Cartesian, his notion of "Spirit" is rather different. It was this aspect that appealed e.g. to Robinson (1734) who, quite apart from his empirical researches, was unsatisfied as to the suitability of animal spirits to do the job required of them ("such a Fluid is altogether unfit for this Work") in the Cartesian sense of "spirit". Newton's conception of "Spirit", at least as expounded in <u>Opticks</u>, appears to analogise nerve impulses to waves of light travelling along transparent rods, reminiscent of modern fibre optics. Thus the "Capillamenta of the Nerves and the Muscles" are "solid pellucid and uniform" so that "the vibrating Motion of the aetherial Medium may be propagated along them from one end to the other uniformly and without interruption." Newton explains that nerves may not <u>look</u> transparent only because a nerve is actually a bundle of capillamenta whose "cylindrical Surfaces may make the whole Nerve ... appear opake and white." (Qu. 24). There is, however, a fascinating, if enigmatic, reference to the "subtle Spirit" which suggests that the notion of action at a distance <u>was</u> playing some part in Newton's thought about nerve conduction. In M.P.N.P. (1729), where Newton talks of "this electric and elastic Spirit" (Galvani didn't publish until 1791) "by the force and action of which", he writes, "the particles of bodies mutually attract one another at near distances, as well repelling as attracting the neighbouring corpuscles; and light is emitted .. and all sensation is excited, and the members of animal bodies move .. by the vibrations of this Spirit, mutually propagated along the solid filaments of the nerves." Unfortunately, Newton declines on grounds of lack of space and adequate experimentation, to discuss this tantalising reference to action at a distance further. Certainly this was <u>not</u> the aspect which was seized upon and echoed by other neuroscientists who were typically captivated by the notion of <u>vibration</u>.

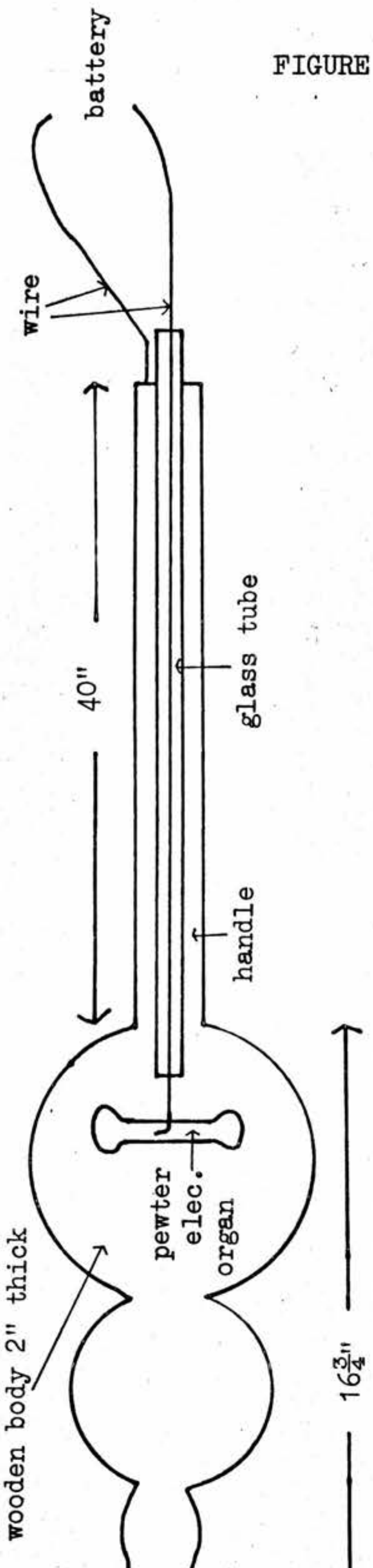
An example is Cheyne (1733): "May not the infinite Windings, Convolutions, and Complications of the Beginnings of the Nerves which constitute the Brain, serve to determine their particular Tone, Tension, and consequently the Intestine Vibrations of their Parts? ... May not these Vibrations be propogated through their lengths, by a subtle, spiritus, and infinitely elastick Fluid, which is the Medium of the Intelligent Principle?"

Newton's reference to electricity need not evidence such a degree of prescience as may initially be supposed. Electricity, as a phenomenon in which a piece of amber when rubbed attracts objects to itself, had been reasonably familiar since Sir Thomas Browne described it the previous century, and it was, no doubt, of this attraction that Newton was thinking. However, the role of electricity in neural functioning was near recognition - and models play an interesting part in bringing that recognition about.

The ability of the fish, the Torpedo, nowadays better known as the electric ray, to impart an unpleasant sensation has been known at least since Aristotle. In the 18th century, however, the question of how the torpedo achieved this was a matter of contention. Lorenzini (1678) for example, thought the assault was due to corpuscles (or "effluvia") which the torpedo gave off and which entered the body. Reaumer (1714) on the other hand, thought the assault due to a mechanical blow which the torpedo gave by sharply contracting its relevant muscles. The debate warmed up and eventually in 1772 sparks started to fly when, after several independent suggestions that the phenomenon was due to electricity, Walsh in some experimental work claimed to demonstrate that the shock was indeed conducted through electrical conductors in the organism. Scientists were generally happy with the basic notion that the phenomenon was due to electricity but two oddities raised grave doubts: i) no electrostatic repulsion or attraction could be detected; ii) no sparks could be observed across gaps in the circuits.

Four years later Cavendish (1776) set out to prove there was "nothing in the phenomenon of the Torpedo at all incompatible with

FIGURE D.




Cavendish's Artificial Torpedo MK I (not to scale).

electricity," by conducting a fascinating series of simulation experiments: "In order to examine more accurately, how far the phenomena of the torpedo would agree with electricity, I endeavoured to imitate them ... "

To be even more precise; he actually built a couple of artificial torpedos. To make his first model torpedo he cut, out of a piece of wood, an approximately torpedo-shaped 'body' (see Figure D) with an additional handle. On each side of the body he fixed a simulated "electric organ" of pewter, wires from which passed up the handle insulated in glass tubes. The whole body was then covered in "sheep skin leather" and soaked thoroughly in a saline solution of one part of salt in thirty parts of water, intended to simulate sea water ("It appeared to conduct electricity not sensibly better or worse than some sea water procured from a mineral water warehouse.") The artificial torpedo was then placed in its habitat, a trough 19" by 14" by 13", and the wires were connected to a battery consisting of 49 Leyden jars. The logic of the experiment is clear; the circuit can only be completed if the electricity finds a way through/across the body of the torpedo or through the water. The intrepid Cavendish placed his hand on the 'fish' and received the expected shock. By varying the charges on, and number of, jars in the battery, Cavendish was able to produce some conclusive results. He found the shock was greater from a large number of weakly charged jars than from a small number of strongly charged ones (i.e. the phenomenon was due to static rather than voltage); he also showed that, if enough jars were used, you could receive a torpedo-like shock even though there was not enough voltage to produce a spark across even a small gap.

However Cavendish was not satisfied. In particular, he could not understand why, when "the shock, when received in air, was as strong as ... that of the real torpedo commonly is" it was "just perceptible when received under water." (By increasing the charge he could get an equivalent shock under water but this then produced a massive shock in air).

To resolve the problem, he decided to build a second improved version of the artificial torpedo. Reasoning that the difference in strength of shock in and out of water might be due to the conductivity of the torpedo's body, he decided to build the second torpedo exactly the same as the first except that the body was to be made of "several pieces of thick leather, such as is used for soles of shoes", again thoroughly soaked with sea water. This alteration had the added appeal of making the artificial resemble the real torpedo even more, for Cavendish saw "no reason to think, that the real torpedo is a worse conductor of electricity than other animal bodies." Success attended the completion and trial of this Torpedo Mk. II: "The event answered my expectation ... the difference between its strength when received under water and out of it, was much less than before, and perhaps not greater than in the real torpedo."

It is quite clear that Cavendish was indeed simulating ... and not just indulging in a fanciful pastime. We find plenty of completed instances of the following sentence schema: "Mr. Walsh found ... I accordingly tried the same experiment with the artificial torpedo ... I got the same results." As a general conclusion "all the foregoing experiments ... seem to agree very well with Mr. Walsh's experiments."

Cavendish conducted yet another simulation experiment with the artificial torpedo which is too charming to fail to relate. Referring to stories of people who were allegedly knocked from their feet by stepping on a torpedo which had semi-buried itself in the sand of a beach, Cavendish "filled a box, 32" long and 22" broad, with sand, thoroughly soaked with salt water to the depth of 4" and placed the torpedo in it, intirely covered with the sand, except the upper part of its convex surface."

Cavendish, perhaps not quite so dauntless after all, decided it was "too troublesome" to actually tread on this loathsome lurking artificial fish and, instead, cut some saline-impregnated leather "intended to represent shoes" (especially "the shoes of persons who walk much on the wet sand"), which he placed on the torpedo and then touched with his hands. All in all, Cavendish concluded it "likely" that a person treading with wet shoes (or even more without shoes at all) on a semi-submerged torpedo would "be thrown down" ... especially as the effect of the animal "would be aided by the surprize"!

Apart from some minor inconsistencies, possibly due to the battery used being underpowered ("a compleat imitation ... would require a battery much larger than mine"), Cavendish showed by his models (which "agree very well with the natural one") that there is "nothing in the phenomenon of the torpedo at all incompatible with electricity". (He also turns, in his report to the Royal Society, to a consideration of the instantiation of such an electrical system, maintaining that although "it is not necessary that there should be anything analogous to a battery within it", in fact, there is "room in the fish for a battery of a sufficient size".

By this solid experimental work Cavendish directly prepared the way for an electric account of the animal spirits, and so brings us into the era of modern neurophysiology. This influence can clearly be seen in the work of Fontana (1781): "The pretended very great velocity of the nervous fluid seems to be contradicted by that inert, viscous fluid or matter with which the primitive nervous cylinders appear filled. The considerable size of the nervous cylinders ... when compared with the primitive fleshy threads, leads me to suspect that these threads are not put in motion in any immediate way however ... by the nerves ... we are not only ignorant of muscular motion, but we cannot even imagine any way to explain it, and we shall apparently be driven to have recourse to some other principle ... if it be not common electricity, may be something however very analogous to it. The electrical gymnotus and torpedo if they do not render the thing very probable, make it at least possible."

Alexander Monro (1783), too, was on the same track: "Most authors have supposed that the nerves are tubes or ducts conveying a fluid secreted in the brain, cerebellum, and spinal marrow. But, of late years, several ingenious physiologists have contended, that a secreted fluid was too inert for serving the offices performed by the nerves, and, therefore, supposed that they conducted a fluid the same as, or similar to the electrical fluid."

From here on, the story is well known, or at least well documented. Galvani's classic experiments of 1791, showing that electricity could induce contraction in muscles, met with some opposition from Volta. (Incidentally both Galvani and Volta discussed analogies between living tissues and electric devices.) The controversy culminated eventually in du Bois Reymond's triumphant declaration of 1848-9: "If I do not greatly deceive myself I have succeeded in realising in full actuality ... the 100 years dream of physicists and physiologists, to wit, the identification of the nervous principle with electricity."

The models discussed by no means exhaust the weird and wonderful devices which neuroscientists found useful in their research in the 150 years between the publication of Descartes' seminal Treatise of Man and Fontana's intimations of the electrical principle, nor was that attempted. Rather I have tried to trace a sort of continuity of theory, centred around the Cartesian concept of animal spirits and furthered by the use of mechanical models. This theme can be restated and reinforced by a liberalised version of an idea of Harmon and Lewis: "While the present utility of ... these models can be doubted, their historical value cannot ... a model or a theory that leads to a dead end is of limited interest, but one that forms a link in the continuing chain is extremely valuable regardless of whether or not subsequent events far outreach it." Amen.

We have now completed our detailed investigation of particular models. In the course of these investigations, various advantages and problems of model use have been raised and discussed.

In the next three chapters, we go on to consider, and attempt to meet, some other, more general problems that have been raised regarding model use as a means of explanation.

# CHAPTER THIRTEEN.

## FOOD FOR THOUGHT : A LOOK AT FODOR AND CRAIK.

"... the diagram of a hydraulic system for motivation, or comparison of the nervous system with a telephone switchboard, or the human brain with an electronic computer, ... do not explain motivation, the nervous system, the brain ... they may only make them more familiar. They do not explain because they, in turn, must be explained either by another science or by another model."

Jay N. Eacker (1972).

A recent work (Raphael, 1976) perhaps somewhat fancifully suggests that several prominent computational theorists, Marvin Minsky, Johnny von Neumann and Norbert Wiener, were/are all direct descendants of a 15th century golem maker, the chief rabbi of Prague, the Maharal. (Sussman dedicates his A computer model of skill acquisition (1975) to the same Rabbi, attributing to him the realisation that 'And God created man in his own image' is recursive!)

It is true that the history of automaton-building-simulators is by no means lacking in bizarre and interesting characters. As mentioned earlier, Jacques de Vaucanson reputedly left the order of Minims of Lyons after his creation of mechanical angels, which flew around the room, incurred the wrath of the provincial. He gave himself up to a life of debauchery in Paris before getting involved in the construction of his more famous androids. He was later suspected of black magic during a voyage by a sea captain who discovered in one of Vaucanson's trunks (and threw overboard) a life-size female automaton (allegedly) constructed to while away the long evenings at sea. To take another example, Julien Offray de La Mettrie added to his translation of Boerhaave's 'Aphrodisiacus' a treatise on venereal diseases (1734), wrote others on vertigo (1736) and smallpox (1740?) and later offended almost the entire literary, religious and scientific world with his brilliant mechanistic philosophy as expounded in L'Homme Machine, which he conceived whilst suffering from a fever during the siege of Freiburg and which was later consigned to the flames by the Parliament of Paris on July 9th 1766. He died, typically, from an unfortunate gastronomic accident ... having consumed a prodigious amount of pate de faison aux truffes! It is to this illustrious tradition that Kenneth Craik, though separated by many intervening generations, clearly belongs, combining academic brilliance, technical virtuosity and a streak of the ludicrous.

Originally a student of philosophy at Edinburgh, he did a little psychology before being sent to Cambridge by James Drever, who forewarned Bartlett (1946) (the then Professor): "Next term I am going to send you a genius." Craik was renowned for his skill in devising and constructing apparatus and gadgets and delighted in a minute working internal combustion engine, which he used to carry around in his waist-coat pocket in a test-tube. Typically, whilst at Cambridge he also attended the Cambridge Technical College for a course on plumbing and welding. As Bartlett remarked in a case of brilliantly insightful (but probably unintended) juxtaposition: "He went to the Cambridge Technical College for a course on plumbing and welding. He did a lot of physiology and, later on, brain anatomy."

The A.P.U. was founded for Craik and he built (as well as designed) there the famous first experimental training cockpit. A representative selection of his published papers are: Origin of Visual After Images ('40); Instrument Lighting for Night Use ('40); Orange Self Luminous Paint ('43); Electrical Simulation of the Ear ('44); White Plumage of Sea Birds ('44). Masses more of unpublished papers are stored away in the A.P.U. Archives, Cambridge, but in his lifetime he published only one actual book: The Nature of Explanation (1943). Craik who was by all accounts a man of tremendous energy and pace, was also somewhat impetuous; during one experiment at the Pleasance, he collapsed his own eyeball and was rushed off to hospital and he eventually died after being knocked from his bicycle into the path of an oncoming vehicle, when he rode full-pelt into an open car door.

The main point of this section is not however to discuss Craik's connections with earlier mechanists (I don't assert a direct lineage!) but rather with one later one.

In particular, I want to discuss some of Craik's ideas together with some of those of Jerry Fodor in 'The Language of Thought' (1975). My aim here is neither to show, as is often done in the purported history of psychology, that (to Fodor's chagrin) Craik had said it all before, nor (to Craik's credit) that some of his ideas are confirmed by Fodor. It is rather to engage in a creative extension of the arguments of both psychologists.

I try to do this in the following way: first I sketch in enough of each man's theory to demonstrate that their quite outstanding similarities of view do provide a common ground between them; secondly I argue that Craik pointed to the isomorphism of the position of the individual coming to know about the world (building a model of it) and that of the psychologist coming to know about the individual (building models of his/her psychological processes); thirdly, I describe and discuss some issues and problems raised by Fodor's work, concerning the way the individual comes to know his/her world; fourthly, I show how, by means of Craik's bridge, some of these considerations are relevant to the problem of the psychologist qua model-builder and that they expose (apparently paradoxical) inherent limitations of the model as an explanatory tool.

1. Craik and Fodor.

Craik: "... thought models, or parallels, reality ... its essential feature is ... symbolism ... this symbolism is largely of the same kind as that which is familiar to us in mechanical devices which aid thought and calculation ..." (1943)

Fodor: "The general structure of psychological theories of cognition presupposes underlying computational processes and a representational system in which such processes are carried out ... When we think of an organism as a computer, we attempt to assign formulae in the vocabulary of a psychological theory to physical states of the organism." (1975).

Craik: "It may be that in genuine generalisations apparently abstract properties of objects are really recognised as the same because, acting on the brain mechanisms of the animal they produce the same effect, ... Thus a balance 'recognises' a weight of 1 grm., whether it be the weight of a piece of brass or of lead. In the same way we can recognise colours apart from the nature of the coloured objects." (1943).

Fodor: "A sensory mechanism is a device which says 'yes' when excited by stimuli exhibiting certain specified values of physical parameters and 'no' otherwise. In particular, it does not care about any property that environmental events fail to share so long as the events have the relevant physical properties in common, and it does not care about non-physical properties that environments have in common so long as they fail to share the relevant physical properties." (1975).

Craik: "... when any calculating machine is in operation, there are objects or events ... the number of teeth projecting from an Odnes wheel (etc.) ... which represent numbers, and ... to the man who observes the machine, these have a greater degree of 'thinghood' and conceptual definiteness than the interconnecting parts and the continuous transmission of motion from one part to another. In the same way in a neural calculating machine there may well be patterns of excitation in the cortex

... and so forth, which, to a physiologist sufficiently skilled, would 'represent' concepts or sensations of objects." (1943).

Fodor: "The criterial property of the machine language of computers is that its formulae can be paired directly with the computationally relevant physical states of the machine in such a way that the operations the machine performs respect the semantic constraints on formulae in the machine code. Token machine states are, in this sense, interpretable as tokens of the formulae." (1975).

Craik: "... the machine will parallel ... those phenomena whose mechanisms most resemble itself ... the structure of the brain ... sets a limit to causal explanation ... because it is ill-suited to representing ... strange processes." (1943).

Fodor: "Nothing can be expressed in a natural language that can't be expressed in the language of thought. For if something could, we couldn't learn the natural language formula that expresses it." (1975).

Craik: "Our thought ... has objective reality because it is not fundamentally different ... from that objective reality but is specially suited for imitating it." (1966 No. 28).

Fodor: "The device is so constructed that its use of the predicate (e.g. in computations) comports with the conditions that such a representation would specify." (1975).

2. Craik.

For Craik the central question was: "What do we mean by explaining anything?" (1943) and his answer was, to a first approximation, "A phenomenon is unexplained if it strikes us as totally unique - uncorrelated and incomparable with anything else in our experience" (1966, No. 28). "We understand something by comparing it with something else in our experience which we <u>do</u> understand ... we model it." Our thinking is ... part of a series of natural events, capable of paralleling and imitating them just as one natural phenomenon ... can resemble or imitate another." (1966, No. 29). Craik, as Bartlett rightly says, "was searching for and using physical analogies", and he looked long and hard at machines "trying to see them as evidence that insofar as they are successful they show how the mind works, not in inventing the machines and using them, <u>but in actually solving the problems</u>." (Bartlett, 1946). Indeed, even the experimental training cockpit was a "very brilliant and beautiful application of a calculating machine principle to a complex psychological problem."

For Craik, one of the most fundamental properties of thought is its power of predicting events but "this process of prediction is not unique to minds, ... A calculating machine, an anti-aircraft 'predictor' and Kelvin's tidal predictor all show the same ability ... the physical process which it is desired to predict is <u>imitated</u> by some mechanical device or model." (1943).

Memory too and the power to learn have their analogies in mechanism: "A piece of sprung steel may bear signs of its previous history of strain" (1943) and "'thought' is a term for the conscious working of a highly complex machine ... this mechanism ... has the power to represent ... certain phenomena in the external world as a calculating machine can parallel the development of strain in a bridge" (1943). Ideas are "labels on parts of a continuously acting maching" (1966, No. 27) and ambiguous or infinite concepts are "perhaps to the mind, what a continual movement of the figure setting keys on a calculating machine would be ... they prevent the problem from ever being set unambiguously to the machine." (1943). On the issue of consciousness, Craik closely followed La Mettrie (1747) ("Thought is so little incompatible with organised matter that it seems to be one of its properties") in not regarding "thought as an inactive halo round mechanical brain processes" but rather attributing "consciousness and conscious organisation to matter when it is physically organised in certain ways" (1943).

However the reader will no doubt have observed that there are really two similar but separate processes being discussed here. Craik is saying both that the individual understands the world by constructing (or using) an internal model of it, and that the theorist (for example the psychologist) understands (psychological) processes by constructing (or using) an external model. (Although he will need, of course, to understand his external model by building (or using) an internal model of it.) Thus, on an individual level, "thought models symbolise external processes and so allow a human being to predict and forestall events", while on quite a different explanatory level, Craik is able to write: "The eye resembles a multi range meter."

There is indeed a very strong formal analogy between the theoretical process of constructing simulation models of natural phenomena and the cognitive processes by which human beings instantiate their knowledge of their environments and their place in them. (Shaw, 1971).

To adapt Minsky's (1968) schema:

1) To an organism B, the set of cognitive structures A* is a model of A to the extent that B can use A* to answer questions that interest him about A.

2) To a theorist B, the machine A* is a model of a psychological phenomenon A to the extent that B can use A* to answer questions

that interest him about A.

(due to Shaw, 1971).

It follows that the task of both the (higher) organism and the simulation theorist is to construct a theoretical model which reproduces the important features of what it is trying to understand, ... but the tasks are logically the same process. Craik pointed to this isomorphism and I shall use it as a bridge, to bring to bear considerations developed by Fodor, in connection with the modelling of the higher organism, on the issue of simulation as a scientific explanation.

3. Fodor.

The idea that higher organisms' knowledge of themselves and their environment is instantiated in cognitive representational systems is common to such diverse writers as: Hebb ('49); Piaget ('67); Miller, Galanter and Pribram ('60); Koffka ('35); McCulloch and Pitts ('43); Lashley ('42); but Fodor ('75) goes much further when he writes that the only psychological theories of cognitive processes "that seem even remotely plausible represent such processes as computational" and that "computation presupposes a medium of computation: a representational system."

The Language of Thought is an attempt to characterise, or at least to show how we could go about characterising, the internal system of representation to which Fodor thinks we are committed by accepting the only plausible psychological theories. He seeks to show, then, that plausible psychological theories do commit us to a 'language of thought'; that this notion of a 'language of thought' is a coherent notion and that it is not empirically totally inaccessible. He admits that his arguments may act as a reductio ab adsurdum (once we know what we are letting ourselves in for by accepting certain theories, we may no longer be so keen to accept them), but he insists that if we <u>do</u> reject such theories, we are left literally with nothing.

Fodor rehearses the same argument in several different scenarios ... considered action, perception and concept learning. In the case of the latter, he (plausibly to my mind) argues that the only serious theory of concept learning we have is that of hypothesis formation and confirmation. But, "concept learning presupposes a format for representing the experimental data, a source of hypotheses for predicting future data and a metric which determines the level of confirmation that a given body of data bestows upon a given hypothesis." The bitter pill to swallow is that representation presupposes a medium of representation - - there is no internal representation without an internal language. In the case of learning a natural language (the classic case of concept learning) it follows that you cannot learn a language whose terms express semantic properties not expressed by the terms of some language you are already able to use. Nothing can be expressed in natural language that cannot be expressed in the language of thought, for, if something could, we could not learn the natural language formula that expresses it (we would have no hypothesis to put forward to be confirmed). You cannot increase your expressive powers by learning .. or, as Wittgenstein put it; "we predicate of the thing what lies in the method of representing it."

Note that the claim <u>is</u> that you cannot learn a language unless you (at least) already understand another one of equal expressive power. This formulation obviates two apparent claims that a vicious regress is involved. Fodor is <u>not</u> saying that you cannot learn a language unless you have already learned one, nor is he saying that understanding (as opposed to learning) involves hypothesising. Understanding may just consist in the fact that one's use of the unlearned language is in fact appropriate e.g. conformable to a truth rule. The patent fact that we do in fact have to learn natural languages e.g. English, necessitates the single regression, but we are not committed to an infinite regression.

Thus, at least one of the languages one knows without learning is as powerful as any that one can ever learn and not all languages one knows are languages one has learned ... i.e. one (at least) is innate.

With this brief outline of '<u>The Language of Thought</u>', let us pause and take stock of our situation. We now know enough of Fodor's thinking, to expect some pretty tough problems concerning how a (higher) organism B could construct a model (symbolic representation) A* of A in order to understand A e.g. where B is a child, A is a sentence of some natural language and A* is a hypothesis as to its semantic properties. But by virtue of our discussion of Craik and Shaw, we found that many problems should transitively apply to the case of a theorist B, who constructs a machine A* as a model of a psychological phenomenon A.

4. <u>Craik's Bridge</u>.

It is instructive to substitute terms in Fodorian schemas:

<table><tr><td>The general structure of</td><td>philosophical theories of cognition<br>psychological theories of modelling</td></tr></table>

presupposes underlying computational processes and a representational system in which such processes are carried out.

The consequences of such a view are (however) as awful for us as for Fodor:

<table><tr><td>You can't</td><td>learn a language<br>simulate a system</td><td>whose</td><td>terms<br>parts</td><td>express</td><td>semantic<br>structuro-functional</td></tr></table>

<table><tr><td>properties<br>relations</td><td>not expressed by the</td><td>terms<br>parts</td><td>of some</td><td>language<br>model</td><td>you are</td></tr></table>

<table><tr><td>able to</td><td>use<br>understand.</td></tr></table>

*Note*: the claim is that you cannot simulate a structure unless you already understand a structure, not that you cannot unless you have already simulated one. Moreover, modelling a structure involves representing the structuro-functional properties of that system, *understanding* the model needn't ... that might just consist in the fact that one's use of the model is always conformable to the isomorphism rule.

Thus, there are two ways in which it can come about that a device (including presumably a person) understands a predicate/system:

1) The device has and employs a representation/model of the extension/system of the predicate where the representation/model is itself given in some 'language' that the device understands.

2) The device is so constructed that its use of the predicate/system comports with the conditions that such a representation would specify.

It is difficult to see, then, how modelling extends the understanding in the strong sense of increasing the power of one's explanatory framework. One only understands the modelled insofar as one realises both that it is isomorphic with the model and, more importantly, insofar as one understands the model itself. The gain seems to be merely that one knows as a result of modelling that, as a matter of fact, the range of application of a particular explanation is actually greater than one previously had thought ... as, for example, when the wave theory of transmission of energy (in water) was extended to cover cases of transmission of sound. Suppose one's inner language of thought were propositional calculus, then, given that learning is by hypothesis formation and confirmation, it would be impossible to learn predicate calculus ... because the latter systematically goes beyond the former in expressive power.

You cannot improve your grasp of English by looking in an English dictionary ... you can only make it more concise, but you only make concise what you already know!

Similarly, it is impossible to fully understand a system of greater intrinsic qualitative complexity by modelling it with a model of lesser intrinsic qualitative complexity ... for that by which the former exceeds the latter will be forever out of reach.

There seem to be two options at this point. Either one can throw up one's hands in horror and simply throw out the practise of modelling as a futile occupation; or, accept some form of extreme nativism ... that we are somehow endowed with an innate understanding of certain models and accept that the development of that part of science which is modelling is simply the extension of the cases of application of these models

There is, however, a point about performance, as opposed to competence, which is, it seems to me, important ... there may be computational advantages associated with knowing a language (being able to model) which are not <u>principled</u>, but which result from performance parameters - memory, fixation of attention etc.

Thus whilst it is true that:

The child / The theorist cannot use the fragment / model of the language / system that he/she knows to increase the expressive power of the concepts at his/her disposal; he/she may use it for mnemonic purposes ... to abbreviate and thus reduce demands on computing memory whilst hypothesising / modelling.

Thus, although it is true that

For every predicate in natural language / understandable structure it must be possible to express a $\frac{\text{coextensive predicate}}{\text{model}}$; it does not follow that for every $\frac{\text{natural language predicate}}{\text{understandable structure}}$ that can be entertained, there is an entertainable $\frac{\text{predicate}}{\text{model}}$.

In other words, it does not follow that an articulate (modelling) organism has *no* cognitive advantage over an inarticulate (non-modelling) one ... just that there is no *principled* advantage.

There certainly *does* seem to be a paradox here, but it should be made clear that this problem is not limited to, i.e. is not a difficulty specific to, modelling, but rather is a general one attaching to notion of understanding. How can it ever be (strictly speaking) that we understand something at time t+1 that we did not understand at t? How can one ever extend one's understanding, as opposed to realising that something new is only a case of something one already knew?

It may be, then, that this argument, like its twin in *The Language of Thought*, merely acts as a reductio ad absurdum ... to make sense of the notion of modelling we appear committed to accepting an extreme form of nativism; we must have wired-in, innately or maturationally developed, a complete set of models of a complexity equal to anything we can ever understand by modelling. This may be simply too much for some to swallow (it being little comfort that *any* account of explanation which relies on a notion akin to hypothesis is subject to the same commitment). But the penalty for rejecting it is to be left with precious little. It is better, in this author's opinion, to live, if a little uncomfortably, with this apparent paradox for the moment, believing with Wittgenstein (1953) that "problems are solved, not by giving new information, but by arranging what we have always known."

# CHAPTER FOURTEEN

## SOME POTENTIAL DIFFICULTIES OF TRYING TO PEEK INSIDE THE BLACK BOX.

"Miracles ... lie between his collar
and his hat, in that queer box he calls
his head."

Arnold Bennett (1908).

The 'black box' is today one of the most widely used explanatory devices in cognitive psychology. Given some process which we cannot sufficiently comprehend, we often construe the situation as one involving a black box which, when given a certain input, yields a certain output. Given the input and the output, we may attempt to deduce (some of) the properties of the (contents of the) black box. Additionally and/or alternatively, we may choose to design or even build a device which yields the same output given the same input - and use this device as a model of the contents of the black box. This, of course, would only be a first step - once minimally, i.e. observationally, adequate the model would have to be rigorously tested for descriptive adequacy.

A definitive statement of this black box approach is given by Chomsky (1965) in a psycholinguistic context: "We may think of the theorist as given an empirical pairing of collections of primary linguistic data associated with grammars that are constructed by the device on the basis of such data. Much information can be obtained about both the primary data that constitute the input and the grammar that is the 'output' of such a device, and the theorist has the problem of determining the intrinsic properties of a device capable of mediating this input-output relation."

An alternative mode of statement makes even clearer the connections between black boxes and the modelling strategy: "we must ask how on the basis of the limited data available to him, the child is able to construct a grammar of the sort that we are led to ascribe to him, ... What, in other words, must be the internal structure of a learning model that can duplicate this achievement?" (Chomsky, 1969) (my underlining).

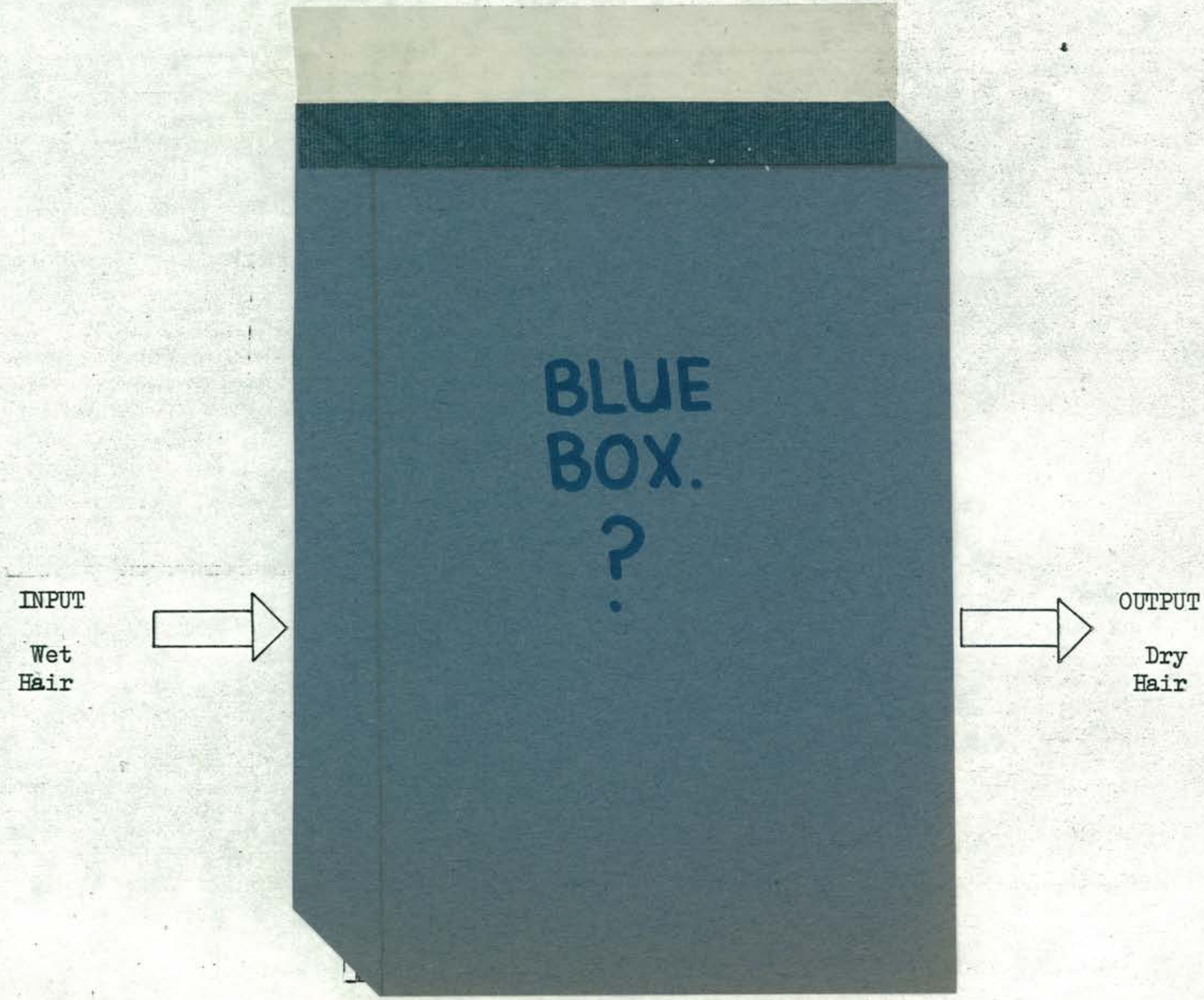
In modelling this 'internal structure' by designing a device which 'can duplicate this achievement', several options are open to us. We may describe such a device linguistically i.e. theorise conventionally; draw it; or even actually build it. As has repeatedly been pointed out, this latter option was one frequently taken up during the 17th and 18th centuries with the concomitant exercise of an amazing amount of technical skill and ingenuity. More recently computer simulation has found more favour, yet the principles are the same: "We postulate that the subject's behaviour is governed by a program organised from a set of elementary information processes. We encode a set of sub-programs (subroutines) for a digital computer, each of which executes a process corresponding to one of these postulated information processes. Then we undertake to write a program, compounded from these subroutines, that will cause the computer to behave in the same way that the subject behaves - to emit substantially the same stream of symbols - when both are given the same problem. If we succeed in devising a program that simulates the subject's behaviour rather closely over a significant range of problem-solving situations, then we can regard the program as the origin of the behaviour." (Newell and Simon).

Clearly though, the possible pitfalls in this approach are legion. One may be simply described: Chomsky's device capable of mediating the input-output relation between linguistic data and grammar has often been called a Language Acquisition Device or L.A.D. for short. Now one fairly sure way of creating such a device has been suggested, in conversation, by John C. Marshall: Problem - create a device which, given the limited and degenerate linguistic input available to any normal human infant, outputs a grammar of the language of which the input was a sample; Solution - take a man and a woman, give them privacy, wait nine months etc. This solution would, naturally, deliver the goods - we would achieve a model matching in structure and function that which we were trying to model. The trouble is, of course, that we would understand the model no more than the modelled; while we asked for a L.A.D. we actually got a LAD (or perhaps a LASS). There is then a preliminary constraint on the contents of the black box which we can dream up. As Pylyshyn (1974-5) has put it: "The scientist must substitute for the 'real thing' a system built on principles which he can understand." It is perhaps worth mentioning that this constraint is still not universally appreciated by all researchers, some of whom continue to postulate (or worse assume) homunculi in the head in their attempts to explain problematic psychological processes or issues in the philosophy of mind.

A second (slightly) more serious possible pitfall would occur were the contents of the black box literally so bizarre that the chances of a modeller achieving descriptive adequacy were virtually nil. This could occur if, for example, the contents of the black box were chosen arbitrarily or on a random basis or if it contained a structure created by an intelligence radically different from that of the modeller or by an arational or even irrational one.

As an illustration, consider the following simplified case of a blue box - and the problem of figuring out its contents. The problem is ... given the initial 'input' of wet hair, to design a device to yield the 'output' of dry hair. Obviously many devices could satisfy such an input-to-output conversion (be observationally adequate) e.g. a towel; a Saharan breeze; an electric hair dryer etc. But would they achieve descriptive adequacy i.e. do what the contents of the blue box do in the same way?

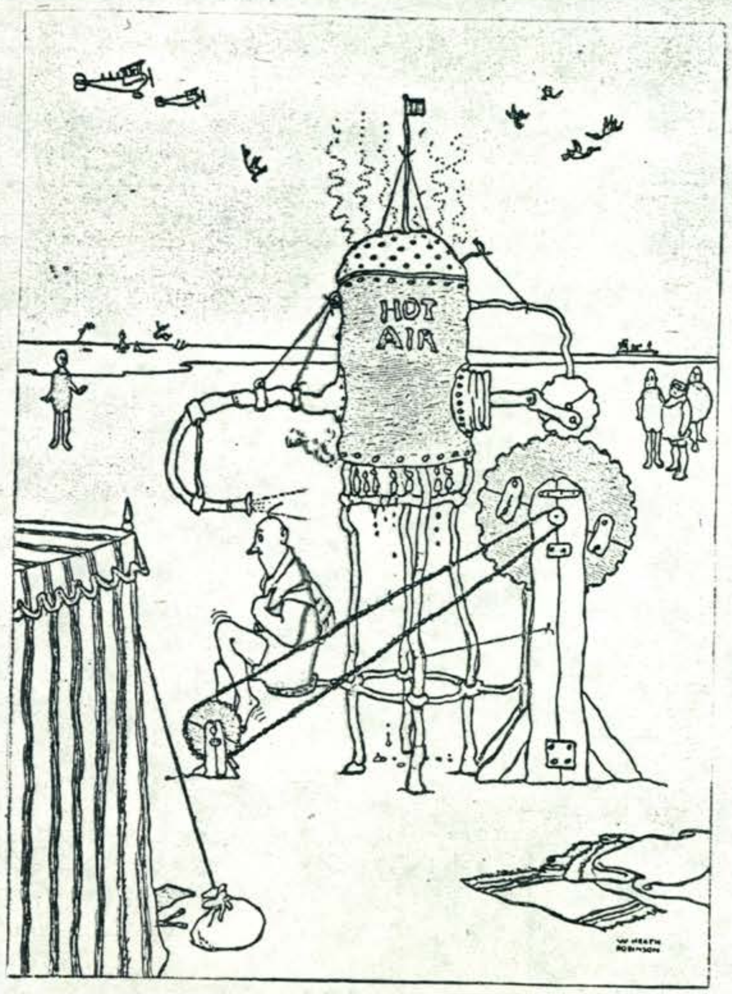
Consider, then, the blue box overleaf. When you have made your hypothesis, check it by lifting the lid of the blue box.




Problem: think of a mechanism which will turn the given input into the given output. How likely is your guess to coincide with what is really in the blue box?

INPUT

Wet
Hair




OUTPUT

Dry
Hair

Problem: think of a mechanism which will turn the given input into the given output. How likely is your guess to coincide with what is really in the blue box?

If you actually _did_ guess the contents of the blue box, may I recommend a long restful holiday and plenty of fresh fruit. However, it is unlikely that you did so for, in our modelling, we tend to assume that which we are modelling to be efficient, parsimonious, lacking in redundancy and in some sense 'reasonable': in short, we tend intellectually to idealise that which we are seeking to understand. This point has been well made by Ghiselin, who wrote of "the tendency of physiologists to anticipate perfection in organic structures ... (which) ... leads them to overlook the imperfection." (1969).

This is no isolated abstract case of splitting hairs without real bearing on scientific methodology. In 1862, Charles Darwin published _On the Various Contrivances by which British and Foreign Orchids are Fertilised by Insects_ in which he intended to supplement _On the Origin of Species_ with detailed substantiation of the belief that "it is apparently a universal law of nature that organic beings require an occasional cross with another individual."

The book has conventionally been received as merely a rather charming guided tour of some eccentricities of nature ... and this was in line with the generally held view of Darwin as a man of rather moderate intellect but with a prodigious appetite for work. However, as Ghiselin (1969) has made clear, Darwin's philosophical competence was not negligible and he was not so artless as to be incapable of satire. Read in this context, the Orchids work does indeed seem "a sort of biological Candide, which, albeit with the greatest restraint, holds up the very idea of organic design to ridicule and contempt."

The general structure of Darwin's argument was to show that the structure of organs was not deliberately designed to fit a particular function, but rather was a fortuitous collection of parts originally adapted to completely different functions. At the core of this argument is a distinction between purposes which do and functions which do not presuppose (according to the Darwinian tradition) a designer. The contrivances of orchids have functions but not purposes.

Most of the book is a description of the amazing variety of these contrivances whereby orchids ensure that any one flower is fertilised by the pollen of another. When Darwin began his researches, he was immediately struck by the apparent bizarreness of these devices. As he put it: "The flowers of Orchids, in their strange and endless diversity of shape, may be compared with the great vertebrate class of Fish, or still more appropriately with tropical homopterous insects which seem to us in our ignorance as if modelled by the wildest caprice." However, in the course of his researches, he came to believe that: "the endless diversity of structure" was not just a bizarre functionless freak of nature, nor even less had "been created for the sake of mere vanity and beauty, - much as a workman would make a set of patterns", rather they represented a "prodigality of resources, for gaining the very same end, namely, the fertilisation of one flower by the pollen of another."

However, Darwin leads us to ask, if God, the ultimate designer, had designed orchids so that each could transfer pollen appropriately, why were there so many devices - why didn't he install the most efficient mechanism in all of them?

Of course, in one sense, all the mechanisms are equally efficient ... all the ones we can examine have after all survived! It is true that some orchids thrive in greater numbers than others, but there is no reason to attribute this to their pollen transferring mechanisms rather than environmental conditions. However, in spite of this, there is a tendency to attribute inefficiency to such natural mechanisms ...

apparently on the intuitive conviction that a designer could come up with a simpler more parsimonious way of achieving the same end.

It has been suggested in the literature stemming from Darwin that machines are efficient in that they are designed to do a job with a minimum of waste, whereas adaptations are merely efficacious in that they just happen to do a job, regardless of waste. As our experience of the blue box has just shown us, this is hardly credible at least in the case of Heath Robinson machines ... and it is doubtful whether a detailed study of the history of automata would support it either, even granting that the suggestion is meant to apply to the intended aims, rather than actual achievements of such machine builders.

However, one can see some truth in the suggestion. A skeleton, for example, is an adaptation which has achieved its structure "through a bizarre history of accretions and deletions" (Reed, 1978b), so that not only do aspects of its structure survive which did not actually hinder (at least in the short term) the survival of its host but were coincident with adaptive aspects (cf. 'superstitious' behaviour in conditioned rats) but also so did aspects which did actually give the organism a selectional advantage although in an outlandish or grotesque fashion. The instructive point for us is that had we, for some reason, practical or ethical, had denied to us the opportunity to make detailed observations of the spine but had had to treat it like a black box, deducing its structure from its load bearing etc. properties and had had to build a model of it, it is most unlikely that we would have come up with anything so 'inefficient' as a human spine!

Darwin did indeed think that because of our relatively poverty stricken imaginations, we were unlikely to be able to model many aspects of nature - but this was not because they were inherently bizarre, but because of their complexity. As Darwin wrote: "The more I study nature, the more I become impressed with ever increasing force with the conclusion, that the contrivances and beautiful adaptations slowly acquired through each part occasionally varying in a slight degree but in so many ways, with the preservation of natural selection of those variations which are beneficial to the organism under the complex and ever varying conditions of life, transcend in an incomparable degree the contrivances and adaptations which the most fertile imagination of the most imaginative man could suggest with unlimited time at his disposal." It is, then, only our ignorance which makes us think nature "modelled by the wildest caprice." What appears to be merely capricious on first glance turns out on further investigation to have some functional contribution. Darwin mentions his own conversion: "I, for one, have often and often doubted whether this or that detail of structure could be of any service; yet, if no good, these structures could not have been modelled by the natural preservation of useful variations." The problem which Darwin points up then is that of the difficulty of generating sufficiently rich, complex and unique models, i.e. a psychological one in the methodology of scientific discovery.

There is then a very real problem facing modelling cognitive simulators. The conviction that the contents of the black box we are trying to understand by modelling is a coherent, parsimonious, efficient reasonable system, is, alas, only an act of faith, buttressed by the (mistaken) belief that because most of our hypotheses as to the contents of the black box are reasonable etc., the contents themselves must be.

On the positive side, considerations such as these do not imply that modelling, even of efficacious contraptions, is impossible - only that it is more difficult than it would be were the structure of the world merely an embodiment of the structure of our mental processes.

The case of the Heath Robinson machine shows that the fact that a machine is designed does not guarantee it is efficient; conversly, the fact that a machine is <u>not</u> made (but evolves) does not imply it is <u>not</u> efficient. As it is not laid down in advance, via the information as to how the device has come to be, which parts of the natural world are or are not efficient, how a machine came to be cannot act as a methodological recommendation for or against modelling. The natural recourse is to try to model it, achieve a model which is observationally adequate and then, by subtle and ingenious experiments, to see whether it is or is not descriptively adequate. Who knows; by innovation, accident or good fortune we might even be able to model the bizarre.

# CHAPTER FIFTEEN

## OUGHTOMATA?

"There are some human functions for which computers <u>ought</u> not to be substituted. It has nothing to do with what computers can or cannot be made to do."

Joseph Weizenbaum (1976).

If there is to be criticism, let it be informed. 'Computer Power and Human Reason' ("an eloquent argument for the sanctity of the human spirit") is dedicated to combatting the dehumanising metaphor, as Joseph Weizenbaum sees it, of man the machine. Its particular interest, for us, lies in the fact that Weizenbaum, a distinguished Professor of Computer Science at M.I.T., past Fellow of the Centre for Advanced Studies, Stanford, and creator of ELIZA (one of the most famous of all natural language processing systems) i.e. a man exactly at the centre of A.I. research, is forced to conclude that, in the debate on 'Computers and the Mind', "the relevant issues are neither technological nor even mathematical; they are ethical."

Thus, to combat the man-machine metaphor, Weizenbaum has to go beyond his own area of considerable expertise on an informal venture into ethics. (This is a trip often made in reverse, of course, humanists arguing against the analogy on sinking ethical grounds are prone to make appeals impugning the logical or epistemological status of the analogy.) So let materialists take heart that a highly motivated Professor of A.I. is unable to come up with powerful formal criticisms of the machine simulation enterprise from within A.I. itself.

One might be tempted to suppose that Weizenbaum's twin roles, as computer scientist and apocalyptic spokesman, would sit uneasily together and this suspicion is substantiated by the structure and style of the book. The first three chapters form a beautifully lucid explication of the basic operations and notions of computing, whilst the last seven gradually decline through wishy-washy humanism and well-intentioned cliches towards the grandiose ("the salvation of the world - and that is what I am talking about.") Yet the whole is punctuated with not infrequent sparks of humour and piercing insights.

Weizenbaum notes, for example, that the operating machine not only abides by but embodies, or instantiates, laws. Thus it is impossible for a machine to behave in an unlawful manner (or "capriciously" as Weizenbaum puts it) - although it might embody laws different from the ones we want at a particular time. It is nonsense, strictly speaking, to say a machine is malfunctioning - rather we should say it is a well functioning machine, but not the one we want. William James (1890) made this same point some 90 years ago: "A machine in working order acts fatally in one way. Our consciousness calls this the right way. Take out a valve, throw a wheel out of gear or bend a pivot, and it becomes a different machine, acting just as fatally in another way which we call the wrong way ... A brain with part of it scooped out is virtually a new machine, and during the first days after the operation functions in a thoroughly abnormal manner."

Because computers, like brains, "interact with the real world", the laws which they embody, unlike abstract machines, must of course, be consistent with the laws of the physical universe. However, a computer, again like a brain, is "not completely characterised by only its manifest interaction with the real world." Whilst it is true that, looked at one way, machines are concerned with the transmission of power (or, in the case of modern electronic digital computers, energy) an alternative but much more useful way to look at them is, as Weizenbaum points out, as concerned with transfer of <u>information</u>: "Our image of the machine (has changed) from that of a transducer and transmitter of power to that of a transformer of information." This realisation has very significant implications for psychology. As Fodor (1975) puts it: "it is the essence of cognitive theories that they seek to interpret physical (causal) transformations as transformations of information, with the effect of exhibiting the rationality of mental processes."

We cannot therefore completely characterise a machine (brain or computer) by talking about its physical workings and interactions with the physical world. Its physical states are symbols which can stand for virtually anything and its symbolic transformations are "regulated by systems of ideas whose range is bounded only by the limitations of the human imagination." In short, although considered physically a computer (brain) cannot violate natural law, considered symbolically it is not thus restricted; one thinks of Atkin's interpretation of chess-playing as the expansion and contraction of two geometrical structures in 53 dimensional space.

Weizenbaum is led to the conclusion that "whatever else man is, then, and again he is very much else, he is also a receiver and transmitter of information", indeed, since we can all learn to imitate universal Turing Machines; "we are by definition ... at least universal Turing Machines." On the other side, he accepts "the idea that a modern computer system is sufficiently complex and autonomous to warrant our talking about it as an organism ... a kind of animal." This modern 'computer-organism' can be "socialised" (modified by experience of the world), "can form a model of itself which could, in some sense, be considered a kind of self consciousness" and the degree of intelligence it can in principle attain is limitless. We could sum up Weizenbaum's thoughts on this then as ... men are at least machines and machines (some at least) are at least organisms.

Weizenbaum raises a criticism of man-machine simulation however which is apparently fairly powerful when he says it is "not obvious that all human knowledge is encodable in 'information structures', however complex."

One purportedly glaring difference between a brain and a computer is the flexibility of the former as to the nature of the information which it will accept and upon which it will act. Dreyfus (1972) is a good example of a holder of this view: "the data with which the computer must operate if it is to perceive, speak, and in general behave intelligently, must be discrete, explicit and determinate; otherwise it will not be the sort of information which can be given to the computer so as to be processed by rule."

As can be seen, there are really two points at issue; what information must be like if it is to be given to a computer; and what information must be like if it is to be operated on by a computer.

As for the former, it is now notorious that in programming we must be totally precise and unambiguous; the set of instructions and their order of operation must be made totally explicit. But this is a pragmatic point - it concerns getting the computer to do what we want, (to be a particular machine), not the nature which the information must (in a strong logical sense) have, if the computer is to be able to operate at all. In this former sense it is, alas, only too easy to put indeterminate and ambiguous instructions into a computer - but the result is that we have no idea what machine the computer then is; or, if you like, what operations it is going through and what problems it is solving. The explicitness and non-ambiguity at this level is thus for our benefit - so that we might know what is going on and be able to utilise it.

The second issue is rather more weighty, as it purports to show that the only sort of information which a (digital) machine can possibly handle and manipulate is inappropriate for solving the sorts of problems which humans solve.

Typically the argument goes as follows:

1. (i) The basic components of a digital computer are binary states (yes/no choices), which can be either on or off, i.e. in one of two mutually exclusive, discrete and determinate states.

   (ii) The total information manipulable by a digital computer is but a summation of such states.

Therefore

   (iii) The total 'knowledge' of a computer is a set of discrete and determinate elements.

(Dreyfus (1972) states in one place that there is an "ontological assumption underlying A.I. work" that "everything essential to intelligent behaviour must in principle be understandable in terms of a set of determinate independent elements" and in another place (1969), even more explicitly writes: "digital computers, composed of flip/flops, must ultimately contain a model of the world represented as a structured set of facts or propositions which are either true or false." Note also, that this objection is extended by Dreyfus (1972) to <u>all</u> discrete state machines - all the machines and automata we have dealt with so far - as well as digital ones: "the flip/flops are only a technical convenience, as is the binary system which they dictate. Any finite state machine whether with three state elements, ten cog gears, or any other set of discrete states would dictate the same ontological conditions."

2. (i) At best, to humans individual elements only become determinate subsequent to an awareness of the whole.

   (ii) Often, humans make a 'configurational response' which is indivisible into elements.

   (iii) Configurational response is essential to <u>all</u> intelligence.

Therefore

3. (i) At best, (according to 1 and 2 (i)), computer intelligence presupposes human intelligence to previously analyse the configurational data.

   (ii) Worse, (according to 1 and 2 (ii)), the computer is incapable of making a human like (configurational) response.

   (iii) Worst of all, (1 and 2 (iii)), the computer is incapable of intelligence.

Strictly speaking the validity of (2) is an empirical matter upon which Dreyfus little elaborates apart from vague gestures towards Gestalt psychology, introspection and phenomenonological doctrine. He cites perspicuous grouping, our recognition by touch of silk and fringe consciousness as the sort of thing he means by indivisible and holistic capacities. However we need not bother ourselves too much with empirically questioning (2), for the argument in (1) is quite clearly fallacious.

Digital computers <u>are</u> discrete state machines but note that in saying that the operations of a digital computer are <u>discrete</u>, we are saying something about the way it <u>represents</u> information ... not about the way it goes from one physical state to another, for in this sense there are <u>no</u> discrete state machines (which would go instantaneously from one state to another without passing through any intervening state). Rather we are saying that states in between have no significance as regards information representation. The difference between a digital and an analogue computer is not that one clicks and one creeps (Williams, 1973), but that, in the latter case, any points crept through represent something but in the former any states clicked through do not.

Because of this, it is quite acceptable and not at all contradictory, to admit that a computer processes information in terms of discrete and determinate states and yet assert that the information itself is not at all discrete and determinate. This is because the way information is conveyed and the information itself are simply not identical - the information is conveyed in a code or language but the information itself is not identical with that code (Sayre, 1968). In particular, the fact that the code is in terms of a set of discrete determinate elements does not entail that the information expressed is - in fact it is not easy to make sense of this claim. Fairly uncontroversially, the code token "the brown cow" is composed (lexically) of three discrete items; the referent of the phrase i.e. the brown cow, is not, at least not three elements in any sense corresponding to the lexical elements. The case of the information contained in the phrase is more problematical. It seems we must await an acceptable theory of semantics. However, note that, for familiar reasons, as we can understand, derive information from, indefinitely many new code tokens, there must be primitives of some kind. This sort of consideration leads us further to doubt Dreyfus' claims in (2) above. Although there are not indefinitely many chess configurations, there are so astronomically many that we couldn't possibly learn a 'configurational response' to each one we might conceivably come across - it seems vastly more likely that chess ability, like language ability, is reliant upon primitives and a syntax of moves rather than a completely novel and holistic response to each individual board layout.

To return to the main theme, as many commentators have pointed out, Dreyfus himself "uses the 26 letters of the Roman alphabet as a symbolic code to represent his 'ambiguous and indeterminate' information" (Boden, 1977). To press this point right home, it is a trivial matter to number the letters of the alphabet from one to twenty-six, express these in binary notation (on a base of two), write anything in English in terms of bits ... and then store them in a digital computer. Thus the digital computer is quite capable of handling representations of Shakespeare sonnets or works of Heidegger - surely ambiguous and indeterminate enough - without change of content.

Of course it remains true that the code with which the nervous system operates may turn out not to be digital ... though the information it handles is codifiable in digital terms to any required degree of accuracy, but this is an empirical matter. To break this code is of course a supreme challenge - one to which the machine simulation research is rising. Whilst it is true, as Farrell (1968) has said, and as I hope to have here demonstrated, "the history of psychology is strewn with models that have exhausted themselves ... without cracking the code of organismic functioning", there is no a priori reason to expect that the story will always be the same.

There is then no convincing argument, presented here, as to why digital computers, because of their binary constitution should be unable to operate on any information whatsoever - so long as that information can be codified. (As Wittgenstein (1921) said in a different connection: "What we cannot speak about we must pass over in silence.") Moreover, on the positive side, as McCulloch (1955) puts it: "To the theoretical question, Can you design a machine to do what a brain can do?, the answer is this: If you can specify in a finite and unambiguous way what you think a brain does with information, then we can design a machine to do it. Pitts and I proved this constructively. But can you say what you think brains do?", thus substantiating Miller, Galanter and Pribram's assertion that "History suggests that man can create almost anything he can visualise clearly." The interesting point is that in their work McCulloch and Pitts showed that any exhaustive and complete description of the brain was "ipso facto realisable by a suitable finite neural network" (von Neumann) - where the neural network was composed of abstract neural modules, of simpler structure than living neurones and capable of assuming only one of two states! It is interesting to note that Minsky's Ph.D. thesis of 1954 was devoted to showing that "almost any of the proposed models of the neuron is a universal computing element" (McCarthy, 1974).

To return again to the main theme. We have discovered that a highly sophisticated and motivated critic of A.I. was unable to come up with any powerful formal arguments against the possibility of a digital simulation of human intelligence. His apparently strongest argument - based on formalisability - fails to achieve its purpose even in Dreyfus' elaborate form. To which arguments does Weizenbaum get thrown back?

Whilst a person's behaviour is generally observable, the functioning of his brain is not. One way to get over this problem is to treat the brain as a 'black box', which yields a certain output given a certain input, and try to build a model, which one understands, which duplicates the output given the same input. Weizenbaum recognises the usefulness of this approach but is worried that we might be overwhelmed by the seductiveness of the model: "the computer is a powerful new metaphor for helping us to understand many aspects of the world, but ... it enslaves the mind that has no other metaphors."

The root of Weizenbaum's worry is that because the computer provides such a good metaphor for man as an information processor, it leads us to ignore aspects other than these abilities. In other words, because cognitive psychology is powerfully aided by machine simulation it may lead us to forget that man has moral qualities, emotions and motivation etc. In fact however, the mechanist's vision is more all embracing than Weizenbaum seems aware. Taking his cue from La Mettrie, who wrote: "We think we are, and in fact we are, good men, only as we are gay or brave; everything depends on the way our machine is running", the modern mechanist can assert that <u>all</u> aspects of a person's psychology (not only the cognitive ones) are, in principle, simulable by a machine. Since, as Turing showed, a digital computer (being an instantiation of a Turing machine) can mimic any other machine, people in toto are in principle simulable psychologically by digital computers, though in practice we have at present little idea how to go about simulating many areas of them.

Of course, on a purely methodological level, the computer metaphor is merely a tool and it would be foolish to restrict ourselves arbitrarily to a single tool. However, Weizenbaum seems to be attributing to machine simulation workers what Margaret Boden (1977) has called: "the mistaken epistemological assumption that the use of one interpretive scheme to extend our understanding excludes the use of others." However, the real issue is not poverty of imagination. We do not <u>have</u> to compare man to a machine, we could compare him to an angel or magician ... the objection to these latter models is not so much methodological monism as the conviction that they would not extend our understanding.

Weizenbaum, then, is left with his final reservation: "What I conclude here is that the relevant issues are neither technological nor even mathematical; they are ethical. They cannot be settled by asking questions beginning with "can". The limits of the applicability of computers are ultimately statable only in terms of "oughts"." Thus, for example, he rejects Colby's suggestion that we could use computers as surrogate psychotherapists "not on the grounds that such a project might be technically infeasible, but on the grounds that it is immoral."

This argument based on morality is difficult to evaluate. The hypothesis that man is a machine is a scientific one, justified only by its fertility in suggesting illuminating experimental hypotheses and systematising generalisations and, as such, is neither moral nor immoral but amoral. Moreover, in a situation in which psychotherapists are in short supply and where people could benefit from psychotherapy, it would seem to me to be immoral to prohibit the use of automaton psychotherapists, providing the patients knew with whom (or if you prefer, with what) they were interacting.

It is difficult to be certain, but Weizenbaum's rationale for this type of statement seems to be that such uses of computers are subtly dehumanising. The computer used in this way is "an instrument pressed into the service of rationalising, supporting and sustaining the most conservative, indeed reactionary components of the current Zeitgeist." This complaint can be amplified as follows: the use of machines to explain, or in place of, people suggests to those people that they _are_ machines and as such should treat each other as, and expect to be treated as, machines. There is some force to this criticism, I admit. In one of the most important points of her recent book, Margaret Boden has pointed out that psychological theories do not only _describe_ but to some extent also _constitute_ psychological reality - because people tend to become what they think they are (as the mistaken belief, whilst walking a high wire, that you are falling might make you fall). As she puts it: "technological analogies (often of a crudely debased popular form) can enter deeply into the personality and self image of individuals ... if the public believes - rightly or wrongly - that science regards people as 'nothing but clockwork', then clockwork-people we may tend to become." It is a strange paradox, if in fact it is true, that man is such a sensitive machine that he can turn himself into an insensitive machine.

Even the sophisticated simulator may be in some danger. Seymour Papert, writing the forword to Sussman's A computer model of skill Acquisition (1975), a model called Hacker, wonders whether Sussman has "in creating Hacker in his image, also, to some extent, recreated himself in Hacker's image?" However, even if the working simulation psychologist escapes this recursive fate, and even if his/her concept of a machine, a computer, may be extremely rich, there is always the real possibility that, in ignorance of, or unpersuaded by, the researcher's actual beliefs, the subject of the research, or even more the casual bystander, may turn himself into what he thinks he is - an impoverished machine. In consequence, it may be important to enrich the generally held concept of machine, so that it becomes no longer an insult to be compared to one. But this is precisely what current machine intelligence studies are going! It is generally complained that it is dehumanising to compare man to a machine, but one seldom hears that it is 'demachinising' to compare a machine to a man! It is assumed without warrant that in comparing man and machine we have to bring man 'down to the level of' a machine, it is equally possible, and less potentially offensive, to think of bringing a machine 'up to the level of' a man. This is what machine intelligence studies are doing.

It is thus significant that the real conclusion of Margaret Boden's recent book is precisely that: "the prime metaphysical significance of artificial intelligence is that it can counteract the subtly dehumanising influence of natural science ... by showing, in a scientifically acceptable manner, how it is possible for psychological beings to be grounded in a material world and yet be properly distinguished from 'mere matter!" i.e. she suggests that, in a sense, we <u>ought</u> to build automata.

# CHAPTER SIXTEEN

## CONCLUDING MACHINATIONS.

Boring (1950) attributes to Ebbinghaus the remark that psychology has a long past, but only a short history. If by 'history' we mean a formal record of the past, a chronicle, Ebbinghaus' aphorism is still today both true and false of experimental psychology. It is true in the sense that the history of experimental psychology is still, by and large, regarded as beginning in 1862, false in the sense that Boring's monumental A History of Experimental Psychology, with its close on 800 densely packed pages, cannot, by any stretch of the imagination, be considered short. This latter sense is by no means as trivial as may initially appear, nor is it unconnected to the former sense. Let me expand: the prevailing view of the recent nature of the origin of experimental psychology is due, almost single handedly, to Boring's magnum opus, to which there are no serious rivals, as opposed to abridged imitations, whatsoever. The reason for this unchallenged supremacy is, however, nothing more than the colossal nature of the task of rewriting the history of experimental psychology more adequately. Of the few psychologists interested in the history of their subject, still fewer are satisfied with Boring's efforts to do it justice. However, by the time one becomes thus interested, one has already been grist for the Boring mill and a considerable mental exertion is required to detach oneself and see the Boring myth for what it is. Of the psychologists who have achieved this relatively objective pristine state, few any longer have the inclination to devote the unavoidably many years of scholarly research demanded by a serious rival to the History.

Objection can be taken to two unifying ideas of the History. Boring's general vision of the history of psychology, indeed of science, is that it is "intensely personal". He appears to believe that the way to write a history of psychology is to write an account of the personalities of men who have played a dominant role in it ... indeed "rapid scientific progress" is, one way or another, made possible by the operations of the nervous systems of "Great Men".

In this work, although occasional reference has been made to the (often eccentric) characters of the main figures behind research, little explanatory significance is read into this. Far more emphasis, as regards responsibility for the progress of the science of psychology, has however been placed on developing technology. Again and again, scientific psychological advances have been made as a result of the use of neutral analogies of implicit models suggested by progress in technology, and the expression and development of theories in the form of formal models has often only been made possible by technological breaksthrough. Indeed, the grip of mechanical models on the scientific imagination is so tenacious that they are seldom abandoned, except for yet other models, themselves suggested by technological developments.

However, what about the aetiology of technology itself? This question trespasses on a whole new field of enquiry, and I do not intend here to follow it. Suffice it to say that it is at least likely that the aetiology of technology involves questions of technological demand, hence industry, hence economics, hence politics, hence power. We seem in danger, then, of circuitously explaining theoretical developments in psychology by (ultimate) appeal to socio-economic factors. This does not, in itself, disturb the present author unduly, although I feel it is not the whole story, but there are scientists, and even more philosophers of science, who hold such a view to be an expression of total epistemological anarchy.

This view, which allegedly reduces science to a mere epiphenomenon of economic and social conditions, is often labelled externalism, as opposed to internalism, which sees scientific development as being governed by its own immanent laws ... whose inexorable progress can only be expedited or hindered by 'peripheral' socio-economic factors.

As usual, with such extreme characterisations, the 'truth' lies somewhere in between. In classic dialectical fashion, internalism and externalism, rather than being warring mutually-exclusive alternatives, actually jointly provide an answer. This is because psychology, or indeed science in general, may have external causal factors operating on it and yet still retain its own internal rationality. How is this conjuring trick accomplished? At any stage in the development of a science, there is never one theory alone which is uniquely determined by the data ... there are always (indefinitely) many theories consistent with (making evidence of) the available data. Which of these theories is in fact chosen, is itself a phenomenon worthy of explanation ... and it is here that causal factors can enter the explanation without overwhelming it. This has important implications for the philosophy of science, for, as Mary Hesse (1977) has pointed out, it explains how causal explanation need not entail causal determinism.

To return to the particular case in hand, in this thesis we have seen how it is possible for technological developments to causally influence psychological theory without rendering it totally dependent on external factors. Boring, in his History, places little or no emphasis on the role of technology in the history of experimental psychology, and (far too) much emphasis on the personality of the main researchers; in this he has his priorities radically misplaced.

My other principle objection to the History is, of course, Boring's oft repeated myth that experimental psychology was "founded" (p.21) in 1860, with the publication of Fechner's Elemente der Psychophysic.

This assertion has been echoed since Boring's first edition of 1929, in the pale offspring spawned in its shadow. Flugel (1933), for example, fixes the birth of psychology "about the middle of the last century", and repeats the Boring line that the parents were the philosophy of mind and sensory physiology and the midwives the "sturdy trio of English empiricists Locke, Berkeley and Hume." It is to be hoped that this thesis has at least established that an experimental psychology, which attempted to explain by simulating, was thriving long before Fechner was even thought of (conceived), let alone forsook his legendary armchair for the laboratory.

Perhaps, however, something should be said, if not justifying at least explaining, the noted lacunae in the History. Artists are notoriously bad at verbalising about their creations. This may not be a result of intrinsically inadequate verbal skills in general however, so much as the inaccessibility to the linguistic brain of the processes upon which the artistic products are contingent. (I will resist any specific talk of hemispheres). The fabrication of machines, apart from being an art form in its own right, may similarly draw upon non-verbal skills, and the ideas they embody may be extremely difficult, if not impossible, to describe and communicate in words. Mayr (1976) attributes the unique charm of machines to the difficulty of grasping their intellectual substance in terms of conventional forms of communication. He also notes that the ideas instantiated in clocks and mechanical automata at least tend to be related to those of geometry and kinematics and thus be more easily verbalisable than those instantiated in thermal and electrical machines which "transcend the boundaries of the existing mathematics and philosophy."

Be that as it may, it is certainly true that adequate descriptions of automata and simulacra of the past are not easily found. This may be, as suggested, due to the difficulty of describing linguistically entities incorporating ideas radically different from any verbal concepts. How often would the requisite mechanical ingenuity and verbal virtuosity needed to create a machine model worthy of insightful description and also to insightfully describe it, be combined in one person? Presumably seldom.

Of course, the builders of mechanical simulacra might have thought that the actual machines themselves were just as, if not more, elucidatory and indeed durable as any written account. Sadly however, relatively few of the machines in question have survived the ravages of time. This was partly due to storage problems ... the objects usually took up a lot of room and thus tended to be pushed into attics, damp storehouses or vast warehouses, when their initial appeal waned; and it was partly due to the lack of mass production ... as each was handmade, they tended to be unique. When Barnum's museum in Philadelphia burned down, not only the Automaton Chess-Player was destroyed, but also literally thousands of other automata. Some simulacra, of course, struggle through any tribulation. Charles Roberts, a mechanic at the Research Museum of the Franklin Institute, recently painstakingly restored an anonymous fire-damaged writing automaton. When he had completed his task, he set the machine in motion and it wrote out a poem, signing it "Written by Maillardet's automaton" (in French, of course). The Maillardets were a famous 18th century automaton building family.

A third possibility is that the lack of texts describing the efforts and motives of mechanical simulators of yore is due to a policy, conscious or unconscious, on the part of the major libraries as to which books were, and which were not, desirable or even worthy of being shelved. As we have seen, the machines themselves, and so too no doubt works discussing them, have been considered by turns literally works of the devil or trivial playthings for the wealthy and gullible, neither much of an incentive for a hard pressed, and, often in the past, clerical, chief librarian.

Whatever the reason, the vitality and importance of mechanical simulation in early psychology does not lie exposed on every library shelf or in every science museum showcase, but requires some teasing out and making explicit. I hope this too has been accomplished in these pages.

So much for historical contributions, what about issues more relevant to theoretical psychology?

For a long time, philosophy of science saw its role as elucidating general principles of scientific method, which when followed would guarantee the production of 'true science'; indubitably certain knowledge. It is now generally accepted that there are no such algorithms, or effective procedures. We have to explain nature's riddles by hook or by crook, seizing and exploiting opportunities to extend our understanding as available. The use of models, it has been argued, _is_ such an opportunity, and one not to be missed, although, of course, it does not guarantee an increase in understanding. Mechanical models have thus mainly been recommended for their heuristic value in the furthering of our understanding.

Strictly speaking, this recommendation is _neutral_ with regard to an answer to the question "Are people machines?". Even if people were to turn out _not_ to be machines, whatever that may mean, the heuristic value of treating them as if they were machines, for the purposes of explanation, would not be impugned. Indeed, treating people as if they were machines, in the above sense, may be one good way of showing they are not. Of course, people may turn out to be partly machines and partly not, and, in that case, that in which they were not mechanical would be forever beyond the reach of mechanical models i.e. a mechanical model would be necessarily incomplete, but this seems to me to be an empirical question, best resolved by continuing mechanical simulation, rather than one to be settled a priori.

I do not adopt a pessimistic attitude towards this question of completeness. I believe the extension of the mechanistic conception of the world into the cognitive domain, which I see as an indication of the gradual clarification of human thought about such issues, to be, in principle, completable. The question is not whether people are machines, but what kind of machines people are! As Patricia Smith Churchland (1978) puts it in terms entirely appropriate to this thesis: "we humans are basically information processors, albeit of a remarkably intricate and splendidly complex variety. Au fond, we are epistemic engines, whose epistemic states are a function of environmental influx and existing internal states."

I do not intend to get involved here in, largely fruitless, metaphysical wrangles about the human-machine question. However, one comment is in order; it is a merit of this approach that it, whilst being entirely materialist, avoids the threat of reductionism.

Following Fodor and Putnam, I take materialism to be the belief that all real events (ones that can figure in causal sequences) that fall under the laws of a special science, also fall under the laws of physics and are therefore physical. The subject matter of psychology is thus part of the subject matter of physics.

Every science has a vocabulary of predicates such that events fall under the laws of that science by satisfying those predicates ... these are the 'kind' predicates, which pick out the events talked about by it.

Reductionism is the belief in materialism plus the belief that there are kind predicates in (an ideally completed) physics, or whatever reducing science is chosen, corresponding to each kind predicate in (an ideally completed) psychology, or whatever reduced science is chosen.

The point is that whilst special sciences make generalisations about events, each of which is physical, these events need have nothing in common except that they fall under the generalisation of the special science; a fortiori, they need have nothing physical in common.

Anything which is an intelligent system will necessarily be a physical object (materialism), but it is not laid down in advance what its physical constitution must be (neurones, cams, flip-flops, valves, transistors etc.). What makes several things (brains, computers, etc.) be intelligent systems is not anything physical they have in common, but the fact that they fall under some predicate (" - is an intelligent system") of theoretical psychology. Consequently, theoretical psychology is not reducible to physics (there are no kind terms of physics corresponding to the kind terms of theoretical psychology, in particular, to the kind term "intelligent system"). Incidentally, it is not even possible to give a brute enumeration in terms of physics of what it is to be an intelligent system (e.g. x is an intelligent system if and only if x is a brain or x is a computer or ... etc.), for there are indefinitely many possible instantiations in physics of the kind term of theoretical psychology 'is an intelligent system', and an enumeration of the infinite is impossible. (Strictly speaking, whether or not physics actually has an adequate vocabulary for stating the laws of the special sciences is an empirical question, but, at present, we have no grounds for optimism that it has.)

Theoretical psychology then, thus construed, deals with events which are physical, but it is not in any interesting sense reducible to physics. In this manner, it not only avoids the ogre of reductionism, whilst being wholly materialist, but has the complementary advantage of avoiding the spectre of the homunculus in the machine. In mechanical models, we really do have wheels, rather than men, within wheels. The existence of machines which really do function entirely mechanically, yet perform as intelligent beings, provides persuasive grounds for abandoning dualistic notions. As Büchner (1870) wrote:

> "Why even man is able to form out of coarse metal or bits of wood, by the aid of very imperfect tools, instruments that play many tunes, timepieces that tell the hour, machines that weave, knit, sew, write, run, and outstrip the speed of the swiftest animals. In these we see nothing marvellous. But just put a savage, or a man who has never heard of mechanics, in our place; would he not imagine that these machines were living things, moving by their own volition? and would not one of the imbecile aborigines of New-Holland have as good a right, as Virchow remarks, to maintain that these machines do not act on mechanical principles as the partisans of the spiritualistic theory have to maintain that the mind cannot be explained by material motions?"

Of course, these grounds are not compelling. The philosophy of dualism is a complex and all-embracing framework within which any counter argument or demonstration may be defused (as, to be fair, is monism). Thus a dualist, if pressed by apparent advances in cognitive simulation, could admit that machines can be intelligent ... but only in the sense that brains (mere physical entities) can, i.e. by interacting with the quite separate realm of mind. Any alleged success on the part of cognitive simulators might, then, be explained, by their having produced a material entity with which, because of its sufficient complexity, mind has been tempted to interact. No empirical demonstration will then, persuade a committed dualist. However, for the rest of us, such machines as have been discussed above do provide encouragement for the belief that intelligence is possible without spectral interaction.

On the other hand, the theoretical psychology of epistemic engines stresses the importance of internal states as well as environmental influx, i.e. it is no simplistic S-R theory. Indeed, the intervening mechanisms are so richly interwoven that the likelihood of extensive pre-natal or maturational wiring is looming ever larger on our explanatory horizons.

The uses that mechanical models have actually had in research have, I hope, been amply demonstrated above. Some of them are conceptually trivial but experimentally important. Experiments can be done torrentially on mechanical models: they do not get tired; do not complain of boredom; ethical restrictions do not apply to them; hypotheses and the implications of assumptions can be tested at a speed unthinkable in humans ... potentially months of tests can be run in a few hours. Some of these advantages, the ethical ones will, of course, have to be reassessed as mechanical systems grow in intelligence. There may come a day when mechanical subjects (legitimately) demand the right to join a union.

More interestingly, models have the capacity to focus disparate evidence into a single coherent view, the explicitness they demand in their construction is of inestimable value to the clarity of thought and the use of their neutral components gives some rationally heedable advice for the further prosecution of research.

Finally then, what is the relationship between models and theories?

Theories are formal statements only of (and formal models are material expressions only of) what we have been calling the positive component but as such they allow no principled justification of new testable predictions. That in which we, as experimental psychologists at least, are interested are not static correlations of data with respect only to economy and elegance, but dynamic explanations which offer positive predictions into new domains. (Mary Hesse goes so far as to define a model as ..."any system, whether buildable, picturable, imaginable, or none of these, which has the characteristic of making a theory predictive.")

Intimate models are both more and less than theories. We have seen how they are more (positive and neutral components as opposed to positive alone.) In what sense are both they and formal models less? Consider a model A and its modelled B. In so far as A does model B, A and B will have, or instantiate, the same structure; therefore there is some level of theoretical description which will describe them both. But this is to say that A and B are particular instantiations of a more general theory - as it may be that brains and computers will both instantiate the (ideally completed) theory of cognitive psychology, where the material of construction constitutes the negative component. But necessarily a theory is more general than an instance covered by the theory. A theory states but a model instantiates general laws ... this constitutes the less.

Modellers, of course, must not merely generalise from an instance of one (the model) to the theory, and thus commit the inductive fallacy. The interaction is far more subtle ... there is continuous readjustment of the positive in the light of the neutral components and vice versa (in other words theory and model interact). As von Kempelen wrote, he aspired both "to establish a complete system of human speech" and to build a simulating speaking machine, but the two became an integrated concern: "my speaking machine and my theory of speech have made equal progress, the one serving as a guide for the other."

Where does all this leave us with regard to the relations between intimate and formal models and theories? The relative usefulness of formulation in terms of model or theory depends partly on the level of development of the particular science. The domain of application of the positive components of both intimate and formal models and of theories coincides - but in an ideally completed explanation there should be no neutral component. Moreover, in an ideally completed science, a theory will always be more general than an equivalent model which (together with its modelled) instantiates the theory, and, insofar as the theory carries no negative component, it will be (in some sense) correspondingly simpler.

On the other hand, when (as now) one is nowhere near a completed science of psychology, what one needs are dynamic instruments of explanation. At this stage, in respect of their licensed suggestions for further research, intimate models are more useful than formal models or theory; whilst in respect of enforced explicitness, formal models are most useful ... all the notorious difficulties of programming, but where you not only have to make your programming language formulae precise and unambiguous, but also actually design and structure the machine language.

It may well be, then, that Black (1962) is right when he writes: "Perhaps every science must start with metaphor and end with algebra; and perhaps without the metaphor there would never have been any algebra." Certainly, for all his vitriolic attacks on models, even Duhem actually makes use of intimate models, where his knowledge is slight, as in psychology. Of Napoleon, he wrote: "The slightest and most fleeting detail ... did not escape Napoleon's scrutiny, and his visual memory fixed it once and for all as would an instantaneous photograph."

—

BIBLIOGRAPHY.

Abercrombie, J. (1857). Inquiries concerning the Intellectual Powers. London. J. Murray.

Aquinas St. Thomas Summa Theologiae Prima Secundae, Quaestio 13. Art 2. Ottawa. Studii Generalis. 1941 p. 787a.

Arbib, M. A. A Simple Self-Reproducing Model in C. H. Waddington (ed) Towards a Theoretical Biology. U.E.P. 1969.

Arbib, M. A. Kilmer & Nico Spinelli: Neural Models and Memory. in Rosenzweig & Bennet: Neural Mechanisms of Learning & Memory. M.I.T. Press, 1976.

Aristotle. De Motu Animalium. trans. Farquarson 701b 1-4.

Aristotle. De Memoria et Reminiscentia.

Atkin, R. H. (1972). Multidimensional Structure in the game of Chess. International Journal of Man-Machine Studies, vol. 4, pp. 341-362.

Augustine St. Confessions.

Baly, W. & Kirkes W. S. (1848). Recent Advances in the Physiology of Motion and Senses, Generation and Development. Supplement to 2nd vol. of Müller's Elements of Physiology. London. Taylor and Walton.

Bard, J. B. L., Hay, E. D. & Meller, S. M. (1975). Formation of the Endothelium of the Avian Cornea: A study of cell movement in vivo. Developmental Biology 42 334-61.

Bartlett, F. C. (1932). Remembering. London. C.U.P.

Bartlett, F. C. (1946). Obituary Notice on Kenneth Craik. B. J. Psych. 36 109.

Battisti, E. (1960). Automata in: Encyclopaedia of World Art. Vol. II. New York. pp. 182-193.

Beare, J. I. (1906). Greek Theories of Elementary Cognition from Alcmaeon to Aristotle. Oxford. Clarendon Press.

Bedini, S. A. (1964). The role of automata in the history of technology. Technology & Culture, 5, pp. 24-42.

Beloff, J. (1973). Psychological Sciences. A Review of Modern Psychology. London. Crosby Lockwood Staples.

Beloff, J. (1977). Backward Causation. Parapsy. Rev. Vol. 8 No. 1.

Bennett, A. (1908). The Human Machine. London. Hodder & Stoughton.

Berlinski, D. (1972). Philosophical Aspects of Molecular Biology. J.Phil. LXIX pp. 319-335.

Bertrand, N. (1756). Élémens de Physiologie. Paris. Guillaume Cavelier.

Bigelow, J. H. (1965). Theories of Memory. In D. L. Arm (ed.) Science in the Sixties. Albuquerque. Uni. of New Mexico Press.

Black, M. (1962). Models & Archetypes. In Models & Metaphors. New York. Cornell U. P. pp 219-243.

Block, N. J. & Fodor, J. A. Cognitivism & and Analog/Digital Distinction. Unpub. Ms.

Boden, M. A. (1977). A.I. & Natural Man. Harvester Press.

Boden, M. A. (1977b). The Artificial Intelligentsia. T.H.E.S. 9.12.1977.

Boerhaave (1744). Academical Lectures on the Theory of Physic. Vol. 3. London.

Borelli, A. (1680). De Motu Animalium. Rome.

Boring, E. G. (1929)(1950). Appleton-Century-Crofts, Inc. New York.

Born, M. (1953/4). Physical Reality. Phil. Quart. 3. pp. 139-149.

Bougeant, G. H. (1739). Amusement philosophique sur le langage des bestes. La Haye. Van Dolen.

Bowen, F. (1877). Modern Philosophy from Descartes to Schopenhauer and Hartmann. Sampson Low, Marston, Searle & Rivington. London.

Breguet, A. L. (1896). La loupe de l'horloger. Besancon. Bortensdorff.

Brewster, D. (1832). Letters on Natural Magic. London. John Murray.

Bronowski, J. (1965). The Machinery of Nature. Encounter. November.

Brown, A. C. (1976). Bodyguard of Lies. W. H. Allen Ltd. London.

Brown, J. (1958). Some tests of the decay theory of immediate memory. Quart. J. of Exp. Psy. 10: 12-21.

Brown, S. C. (1974). Philosophy of Psychology. Macmillan.

Bruce, R. L. (1977). Fundamentals of Physiological Psychology. N.Y. Holt, Rinehart & Winston.

Büchner, L. (1870). Force & Matter. Empirico-Philosophic Studies, Intelligibly Rendered. London. Trubner & Co.

Burnham, W. H. (1888). Memory, historically and experimentally considered. Am. J. of Psy. 2: 39-90, 225-70, 431-64, 568-622.

Butler, S. (1912). Music, Pictures & Books. Thought & Word. Notebooks.

Butler, S. (1880). Unconscious Memory. London. Cape.

Calvocoressi, P. (1974). The Ultra Secrets of Station X. Sunday Times 24th November. pp. 33-34.

Carroll, C. M. (1975). The Great Chess Automaton. Dover Publications Inc. New York.

Carroll, L. (1865). Alices Adventures in Wonderland. Ch. 10.

Cattell, J. McK. and Bryant, S. (1889). Mental association investigated by experiment. Mind 14. 230-50.

Cattell, J. McK. (1887). Experiments on the association of ideas. Mind 12. 68-74.

Caton, H. (1973). The Origin of Subjectivity. Yale U.P. New Haven.

Caus, S. de. (1615). The Relations of Motive Forces with various Machines. Frankfurt am Main.

Cavanagh, J. P. (1972). Relation between the immediate memory span and the memory search rate. Psych. Rev. 79. 525-30.

Cavanagh, J. P. (1975). Two classes of holographic processes realisable in the neural realm. Lecture Notes in Computer Science. 14-40.

Cavendish, H. (1776). An Account of some Attempts to imitate the Effects of the Torpedo by Electricity. Philosophical Transactions of the Royal Society of London. Vol. 66, Part I. pp. 196-225.

Cermak, L. S. (1976). The encoding capacity of a patient with amnesia due to encephalitis. Neuropsychologia 14. 311-26.

Chapuis, A. and Jaquet, E. (1955). The history of the self-winding watch 1770-1931. Neuchatel Editions du Griffon.

Chapuis, A. and Droz, E. (1958). Automata, an historical and technological study. Neuchatel Editions du Griffon.

Cheyne, G. (1733). The English Malady or A treatise of nervous diseases of all kinds; as, spleen, vapours, lowness of spirits, hypochondriacal and hysterical distempers and C. London.

Chomsky, N. (1965). Aspects of the Theory of Syntax. M.I.T. Press.

Chomsky, N. (1968). Language and Mind. New York. Harcourt Brace.

Chomsky, N. (1969). Linguistics and Philosophy. In Hook, S. (ed.). Language and Philosophy. New York Uni. Press.

Chopping, P. T. (1968). Holographic model of temporal recall. Nature 217. pp. 781-2.

Churchland, P. S. (1978). Fodor on Language Learning. Synthese Vol. 38. No. 1 May.

Cicero. De oratore.

Cipolla, C. M. (1967). Clocks and Culture 1300-1700. Collins. London. see p. 122 note 6.

Clauberg, J. (1664). Physica, quibus rerum corporearum vis et natura ... explicanteur. Third ed. Amstelodami, D. Elzevirium.

Cohen, J. (1966). Human Robots in Myth and Science. London. Allen and Unwin.

Cohen, J. (1970). Human Robots and Computer Art. History Today.

Colby, K. M., Watt, J. B. and Gilbert, J. P. (1966). "A Computer Method of Psychotherapy, Preliminary Communication". The Journal of Nervous and Mental Disease, Vol. 142, 2. pp. 148-152.

Collins, A. M. and Quillian, M. R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behaviour 8. 240-7.

Cotes, R. (1729). Preface to: The mathematical principles of natural philosophy of Sir Isaac Newton. trans. A. Motte. London.

Craik, F. I. M. and Lockhart, R. S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behaviour 11. 671-84.

Craik, K. J. W. (1943). The Nature of Explanation. C.U.P.

Craik, K. J. W. (1966). On the workings of the mind. No. 27 in The Nature of Psychology ed. Isherwood. C.U.P. 1966.

Craik, K. J. W. (1966). On the inductive study of language and the meaning of particular words. in 'The Nature of Psychology' ed. Isherwood. C.U.P. 1966 No. 28.

Craik, K. J. W. (1966). Induction, coherence, relational systems, and the objective validity of thought. Isherwood. No. 29.

Crowder, R. G. and Morton, J. (1969). Precategorial acoustic storage (PAS). Perception and Psychophysics 5. 365-71.

Cummins, R. (1977). Programs in the Explanation of Behaviour. Philosophy of Science. Vol. 44. No. 2. June 1977.

Darwin, C. (1862). On the Various Contrivances by which British and Foreign Orchids are fertilised by Insects and on the Good Effects of Intercrossing. London. John Murray.

Darwin, E. (1806). The Poetical Works. Vol. III. J. Johnson. London.

Delboeuf, J. (1880). Le sommeil et les rêves: leurs rapports avec la theorie de la memoire. Revue Philosophique 9. 129-69.

Dennett, D. C. (197 ). Critical Notice on Fodor's Language of Thought. Mind. 1976.

Descartes, R. (1633). Traité de l'homme. trans. T. S. Hall. Treatise of Man. Camb. Mass. Harvard U.P. 1972.

Descartes, R. (1637). Discourse on the method of rightly conducting the reason and seeking for truth in the sciences. Leiden, Maire. trans. Haldane and Ross. Vol. one, p. 117.

Descartes, R. (1649). Passions of the Soul. Amsterdam.

Deutsch, J. A. (1960). The Structural Basis of Behaviour. C.U.P.

Deutsch, J. A. (1973). The Physiological Basis of Memory. New York. Academic Press.

Diamond, D. (1974). The Roots of Psychology. Basic Books. N.Y.

Digby, K. (1664). Two treatises: in the one of which the nature of bodies; in the other the nature of man's soule is looked into. Paris, Blaizot. 1664. Part one; Ch. 36-38.

Dijksterhuis, E. J. (1961). De Mechanisering van het Wereldbeeld. trans. C. Dikshoorn. Oxford. Clarendon Press.

Dresher, B. E. and Hornstein, N. (1976). On some supposed contributions of A.I. to the scientific study of language. Cognition 4. pp. 321-398.

Dreyfus, H. (1967). Why Computers must have bodies. Rev. Metaphysics. Vol. 21. No. 1 Sept.

Dreyfus, H. (1972). "What Computers can't do". Harper and Row, New York.

Dudley & Tarnoczy (1950). The Speaking Machine of Wolfgang von Kempelen. J. Ac. Soc. Am. Vol. 122. No. 2 March.

Duhem, P. (1914). The Aim and Structure of Physical Theory. Princeton U.P. 1954.

Eacker, J. N. (1972). On some elementary philosophical problems of psychology. Am. Psych. June. p. 553.

Ebbinghaus, H. (1885). Über das Gedächtnis. Leipzig. Duncker and Humblot.

Encyclopaedia Britannica. (1929). fourteenth edition (London, Encyclopaedia Britannica Co. Ltd. pp. 835-840.

Encyclopaedia of World Art. (1960). N.Y. pp. 182-193.

Ettinghausen, R. (1960). Automata. In. Encyclopaedia of World Art.

Farrell, B. A. (1968). Some thoughts on the use of Models in Psychology. Logic, Methodology and Philosophy of Science. III. ed. Rootselaar & Staal.

Farrington, B. (1947). Head and Hand in Ancient Greece. London. Watts.

Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. Memory and Cognition 5. 335-9.

Flugel, J. C. (1933). A Hundred Years of Psychology. Duckworth.

Fodor, J. A. (1968). Psychological Explanation. New York. Random House.

Fodor, J. A. (1968). "An Appeal to Tacit Knowledge in Psychological Explanation". Journal of Philosophy.

Fodor, J. A., Bever, T. G. & Garrett, M. F. (1974). The Psychology of Language. McGraw-Hill. N.Y.

Fodor, J. A. (1975). The Language of Thought. Hassocks. Harvester Press.

Fontaine, N. (1738). Memoires pour servir a l'histoire de Port Royal. Cologne, Vol. two, fifty two - three.

Fontana (1795). Treatise on the venom of the viper etc. Florence, 1781 in English. London. John Cathall.

Fontanelle, B. le B. de. (1764). Oeuvres. Amsterdam, F. Changuion. See Lettres Galantes, part one, letter eleven.

Frege, G. (1970). On Sense and Reference. In Philosophical Writings of G.F. ed. Geach and Black. Blackwell. Oxford 1970.

Freud, S. (1891). Zur Auffassung der Aphasien. Vienna. Deuticke.

Freud, S. (1893). Some points for a comparative study of organic and hysterical motor paralyses. Standard Edition. Vol. 1. London. Hogarth Press.

Freud, S. (1895). Project for a scientific psychology. Standard Edition, Vol. 1. London. Hogarth Press.

Freud, S. (1896). Letter 52. Standard Edition. Vol. 1. London. Hogarth Press.

Fryer, D. M. & Marshall, J. C. (1977). A Timely Note on Nervous Ticks. Paper read at XV International Congress of the History of Science. Edinburgh.

Fryer, D. M. & Marshall, J. C. (1979). The Motives of Jaques de Vaucanson. To appear in Technology and Culture, 1979.

Haller, (1768). Anfangsgründe der Physiologie. Vol. IV C.F. Voss Berlin 1768. p. 585.

Hamlyn, D. (1957). The Psychology of Perception. London. Routledge and Kegan Paul.

Hamming, R. W. (1950). Error-detecting and error-correcting codes. Bell System Technical Journal 29. 147-60.

Gall, F. J. (1791). Philosophisch-medicinische Untersuchungen uber Natur und Kunst. Vienna. R. Graffer.

Galvani, L. (1791). De Viribus Electricitatus in Motu Musculari. Bologna. Ex Typographica Instituti Scientarium.

Gardner, H. (1977). The Shattered Mind. R.K.P. London.

Gauld, A. and Shotter, J. (1977). Human Action and its Psychological Investigation. R.K.P. London.

Geschwind, N. (1972). Disorders of higher cognitive function in children. Clinical Proceedings, Children's Hospital National Medical Center 28. 261-72.

Gesner, J. A. P. (1770). Die Sprachamnesie. Nordlingen.

Ghiselin, B. (1952). The Creative Process. Mentor.

Ghiselin, M. T. (1969). The Triumph of the Darwinian Method. Uni. of California Press.

Gilson, E. (1913). La doctrine Cartesienne de la liberté et de la theologie Alcan, Ch. one.

Glisson, F. (1677). Tractatus de ventricule et intestine. Amsterdam.

Goethe, J. W. von. (1797). Letter to Charles August.

Gomulick, B. R. (1956). Recall as an abstractive process. Acta Psychologica 12. 77-94.

Goncalez, C. (1609). Consideraciones sobre et psalmo 133 para los seis domingos de la quaresma. Madrid.

Goodman, N. (1969). Languages of Art. O.U.P. London.

Gratoroli, G. (1562). The Castel of Memorie. London. R. Hall.

Gregor, M. J. trans. (1974). Anthropology from a pragmatic Point of View. The Hague. Nijhoff

Gregory, F. (1977). Scientific Materialism in 19th c. Germany. Reidel. Studies in the History of Modern Science.

Grey, R. (1730). Memoria Technica. London. C. King.

Grice, H. P. (1961). The causal theory of perception. Proceedings of the Aristotelian Society. Suppl. Vol. 35. 121-52.

Gurowitz, E. M. (1969). The Molecular Basis of Memory. Englewood Cliffs, New Jersey. Prentice-Hall.

Guyau, M. (1880). La memoire et la phonographe. Revue Philosophique 9. 310-22.

Harless, A. (1856). Goethe im Fegfeuer. Eine materialistische-poetische Gehirnssekretion. Stuttgart.

Harmon, L. D. & Lewis, E. R. (1966). Neural Modeling. Physiological Rev. Vol. 46. No. 3. pp. 513-591.

Harris, J. (1751). Hermes. London. H. Woodfall.

Head, H. (1920). Studies in Neurology. London. Hodder and Stoughton.

Helmholtz, H. (1901). On the Interaction of the Natural Forces. In Popular Lectures on Scientific Subjects. trans. E. Atkinson. London. Longmans, Green and Co., p. 137.

Henschen, S. E. (1920-22). Klinische und anatomische Beiträge zur Pathologie des Gehirns. Stockholm. Nordiska.

Hering, E. (1870). On Memory as a Universal Function of Organised Matter. Vienna. K. Gerold's Sohn.

Hero of Alexandria : see Woodcroft (1851).

Hesse, M. (1966). Models and Analogies in Science. University of Notre Dame Press.

Hesse, M. (1977). Internal and External Causation of Scientific Ideas. In Proceedings of XVth International Congress of the History of Science. Edinburgh.

Higgins, C. Longuet (1968). In Our Image? In The Humanist Outlook ed. Ayer A. J. Pemberton Pub. Co.

Higgins, C. Longuet (1976). Perception of melodies. Nature, 263, p. 653.

Hill, D. R. (1974). Introduction to Jazzari.

Hillier, M. (1976). Automata and Mechanical Toys. Jupiter Books (London) Ltd.

Hindenburg, C. F. (1784). "Über den Schachspieler des Herrn von Kempelen nebst einer Abbildung und Beschreibung seiner Sprachmaschine." Leibzig. pp. 11-47.

Hitch, G. J. and Baddeley, A. D. (1976). Verbal reasoning and working memory. Quarterly Journal of Experimental Psychology 28. 603-21.

Hobbes, T. (1651). Leviathan. London.

Hoffman, E. T. A. (1884). The Sandman.

Hooke, R. (1682). An Hypothetical Explication of Memory; how the organs made use of by the mind in its operation may be mechanically understood. Lecture to the Royal Society, 21 June (printed in Waller 1705).

Hopkins, G. M. (1970). No worst, there is none. In the Poems of G.M.H. 4th ed. (ed.) Gardner and MacKenzie. Oxford paperbacks. London 1970.

Huxley, T. H. (1874). On the hypothesis that animals are automata and its history. Address Brit. Assoc. Advancement of Science. Belfast.

Huxley, T. H. (1898). Collected Essays. Macmillan. London.

Inglis, B. (1976). Parapsychology and Rationalism. New Humanist Nov/Dec.

Jackson, J. H. (1874). On the nature of the duality of the brain. Medical Press and Circular 1. 19-31.

Jacobs, J. (1885). Review of Ebbinghaus. Mind 10. 454-9.

Jacobs, J. (1887). Experiments on 'prehension'. Mind 12. 75-9.

James, W. (1891). The Principles of Psychology. Macmillan and Co. London.

Jamieson, A. (1827). A Dictionary of Mechanical Science, Arts, Manufactures and miscellaneous knowledge. Vol. 1. London. Henry Fisher, Son & Co. see entry under Automata.

Janvier, J. Journal Encyclopaedique (See Chapuis and Jaquet).

Jazzari, ibn al Razzaz al. (1206). The Book of Knowledge of Ingenious Mechanical Devices. trans. & annotated D. R. Hill. D. Reidel. Dordrecht. 1974.

Jung, C. G. (1955). In an interview with Stephen Black (B.B.C.Copyright) but see E. A. Bennett. C. G. Jung. Barrie & Rockliff, London 1961. p. 149.

Jeans, J. (1937). The Mysterious Universe. Pelican.

Kaden, S. E. et al (1955). J. Psychology 39. pp. 61-70.

Kant, I. (1798). Anthropologie in pragmatischer Hinsicht. Koningsberg. Nicolovins

Keating, M. J., Beazley, L., Feldman, J. D. and Gase, R. M. (1975). Binocular interaction and intertectal neuronal connexions; dependence upon developmental stage. Proceedings of the Royal Society of London B, 191. 445-66.

Keller, A. G. (1975). Review of Jazzari in British Journal of the History of Science, Vol. viii. Part 1. No. 28. March.

Kempelen, W. von (1791). Le Mecanisme de la Parole suivi de la description d'une Machine Parlante.

Kepler, J. (1605). Opera. ed. Frisch C. Vol. 12. Frankfurt. Heyder & Zimmer, 1859. pp. 83-84.

Knauss, F. (1787). Selbstschreibende Wundermaschine auch mehr andre Kunst und Meisterstücke.

Köhler, W. (1947). Gestalt Psychology. New York.

Larsen, E. (1960). Ideas and Invention. London. Spring Books.

Lashley, K. S. (1929). Brain Mechanisms and Intelligence. Chicago. University of Chicago Press.

Lashley, K. S. (1950). In search of the engram. Symposia of the Society for Experimental Biology 4. 454-82.

Levy, D. (1976). Chess and Computers. London. B. T. Batsford Ltd.

Locke, J. (1690). Essay Concerning Human Understanding.

Liberman, A. M., et al. (1959). Minimal Rules for Synthesising Speech. J. Ac. Soc. Am. Vol. 31. No. II. pp. 1490-1499.

Lodge, O. (1890). Modern Theories of Electricity: Essay in a New Theory. London.

Loftus, E. F., Freedman, J. L. and Loftus, G. R. (1970). Retrieval of words from subordinate and superordinate categories in semantic hierarchies. Psychonomic Science 21. 235-6.

London Gazette. (1686). tenth to thirteenth of January.

Lucretius. Lib IV v533. trans. Lathan, R. E. Penguin 1951. p. 147.

Lups, J. (1748). De irritabilitate. Leyden Luzac.

Mackay, D. M. (1974). In Brown, S. C. (ed.) 1974.

Malcolm, N. (1977). Memory and Mind. Ithaca. Cornell University Press.

McCarthy, J. (1974). Book review of Lighthill. Artificial Intelligence 5. pp. 317-322.

McCulloch, W. S. (1955). Mysterium iniquitatis of sinfulness aspiring into the place of God. The Science Monthly. Vol. 80. No. 1. Jan.

McDougal, W. (1897). J. Anatomy & Physiology. 31. p. 410.

Malebranche, N. (1678). De la recherche de la vérité. Paris. Garnier Frères. Book six, part two, chapter seven, p. 133.

Marr, D. (1977). A.I. a personal view. A.I. 9. pp. 37-48.

Marshall, J. C. (1970). Review article. Mentalism and objectivism in linguistics. Semeotica 2. 277-93.

Marshall, J. C. (1974). Freud's psychology of language. In R. Wolheim (ed.) Freud: A Collection of Critical Essays. New York. Anchor Books.

Marshall, J. C. (1977). Minds, machines and metaphors. Social studies of Science 7. No. 4. pp. 475-488.

Marshall, J. C. & Fryer, D. M. (1978). Speak Memory! An Introduction to some historic studies of remembering and forgetting. To appear in Aspects of Memory. ed. Gruneberg. Academic Press.

Mayr, O. (1976). Philosophers and Machines. N.Y. Science Publications.

Meigs, E. B. (1905). A mechanical theory of muscular contraction and some new facts supporting it. Am. J. Physiology, XIV.

Mercier, C. (1885). A Classification of Feelings. Mind Vol. X.

Mettrie, J. O. de la (1748). L'Homme Machine. Luzac. Leyden. English edit. trans. G. C. Bussey Open Court 1912.

Meyer, D. E. and Schvaneveldt, R. W. (1976). Meaning, memory structure, and mental processes. Science 192. 27-33.

Mill, J. S. (1891). A System of Logic. London. Longmans, Green & Co. Book III. Ch. 20.

Miller, G. A. (1956a). The magical number seven, plus or minus two; some limits on our capacity for processing information. Psychological Review 63. 81-96.

Miller, G. A. (1956b). Information and memory. Scientific American 195. 42-6.

Miller, G. A. (1968). Encoding the unexpected. Proceedings of the Royal Society of London B, 171. 361-375.

Milne, H. J. M. (1934). Greek Shorthand Manuals. London. Cassell.

Minsky, M. (1968). Matter, Mind and Models. in Minsky (ed.): Semantic Information Processing. M.I.T. Press.

Monro, A. (1783). Observations on the Structure and Functions of the Nervous System. London. Creech, Cadell, Elmsley, Murray & Longman.

Morton, J. (1970). A functional model for human memory. in D. A. Norman (ed.) Models for Human Memory. New York. Academic Press.

Motte, A. H. de la (1721). A discourse on a fable. trans. Samber. London Curel & Jaunay.

Müller, G. E. and Pilzecker, A. (1900). Experimentalle Beiträge zur Lehre vom Gedächtniss. Zeitschrift für Psychologie. Erganzungsband 1. 1-300.

Müller, J. (1842). Elements on Physiology, Volume 2. trans. Wm. Baly. London, Taylor and Walton.

Murray, D. J. (1976). Research on human memory in the nineteenth century. Canadian Journal of Psychology 30. 201-20.

Needham, J. (1960). Automata. See Encyclopaedia of World Art.

Neumann, J. von (1958). The Computer and the Brain. New Haven. Yale University Press.

Newell, A. & Simon, A. Computer Simulation of Human Thinking. Science 61. Vol. 134. p. 2011.

Newton, I. (1724). Mathematical Principles of Natural Philosophy. trans. A. Motte. London. Vol. 1. 392-393.

Newton, I. Opticks. 2nd ed. Queries 23 & 24.

Norris, D. M. (1939). A History of Cataloguing and Cataloguing Methods. London. Grafton.

Northway, M. L. (1940). The concept of the schema. British Journal of Psychology 30. 316-25, 32. 22-36.

Oldfield, R. C. (1954). Memory mechanisms and the theory of schemata. British Journal of Psychology 45. 14-23.

Oresme, N. (1370). Le livre du Ciel et du Monde. ed. Menut, A. D. & Denomy, A. J. trans. Menut. Wisconsin Uni. Pub. in Medieval Science. Madison, 1968.

Paget, R. (1930). Human Speech.

Pascal, B. (1669-70). Pensées.

Paivio, A. (1974). Images, propositions, and knowledge. University of Western Ontario, Research Bulletin No. 309.

Parsons, E. A. (1952). The Alexandrian Library. London. Cleaver-Hume Press.

Patten, B. M. (1972). The ancient art of memory. Usefulness in treatment. Archives of Neurology 26. 25-31.

Pearson, K. (1892). The Grammar of Science. London. A. & C. Black.

Peirce, C. S. (1870). Logical Machines. Psychological Literary and Philosophical Transactions.

Penfield, W. (1968). Engrams in the human brain. Proceedings of the Royal Society of Medicine 61. 831-40.

Peterson, G. E., Wang, W. S-Y. and Sivertsen, S. (1958). Segmentation techniques in speech synthesis. Journal of the Acoustical Society of America 30. 739-742.

Plato. Theaetetus.

Pliny. Natural History. Book 7.

Plotinus. Enneads. Book 4.

Poe, E. A. (1836). Maelzel's Chess Player. Southern Literary Messenger. April.

Poisson, N. (1670). Commentaire ou remarques sur la méthode de René Descartes. Vandosme. S. Hyp. Part 5. observation 3.

Powell, T. (1661). Human Industry. London.

Power, H. (1664). Experimental philosophy in three books containing new experiments microscopical, mercurial, magnetical. London.

Pratt, F. (1939). Secret and Urgent. London. Hale.

Price, D. J. de Solla. (1964). Automata and the origins of Mechanism and Mechanistic Philosophy. Technology and Culture 4. pp. 9-23.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain. A critique of mental imagery. Psychological Bulletin 80 1-24.

Pylyshyn, Z. W. (1974-5). Minds, Machines and Phenomenology. Cognition 3 (1) pp. 57-77.

Pylyshyn, Z. W. (1975). Representations of knowledge. in Schank and Nash Webber (eds.).

Pylyshyn, Z. W. (1976). Complexity and the Study of Artificial and Human Intelligence. in Kaneff, S. & O'Callaghan, J. F. (eds.). Objectives and Methodologies in A.I. Academic Press. 1976.

Quine, W. V. (1969). Linguistics and Philosophy. in S. Hook (ed.) Language and Philosophy. New York University Press. p. 96.

Ramus, P. (1578). Scholae rhetoricae. Bale.

Randell, B. (1976). The Colossus. University of Newcastle upon Tyne Technical Report Series No. 90. June.

Raphael, B. (1976). The Thinking Computer. W. H. Freeman & Co. San Francisco.

Reed, E. (1978). How is scientific psychology possible? Manuscript. Edin. Uni.

Reed, E. (1978b). Are Organisms Machines? Unpublished manuscript. Boston University Philosophy Dept.

Reid, T. (1785). Essays on The Intellectual Powers of Man.

Reymond, E. du Bois (1848-9). Untersuchungen über thierische Elektricität. Berlin. Reimer.

Riese, W. (1959). A History of Neurology. M.D. Publications N.Y.

Robinson, B. (1734). A treatise on the animal oeconomy. Dublin.

Romanes, G. J. (1896). Mind and Motion. Longman's, Green and Co. London.

Rosen, D. A. (1975). An argument for the logical notion of a memory trace. Philosophy of Science 42. 1-10.

Russell, B. The Best of Bertrand Russell. (Introduction).

Sarton, G. (1931). Introduction to the History of Science. Vol.III.London.

Sayre, K. M. (1968). Intelligence, bodies and digital computers. Review of Metaphysics 21. pp. 714-723.

Schank, R. and Nash-Webber, B. (1975). Theoretical Issues in Natural Language Processing.

Scharff, R. (1975). The How and Why Wonder Book of Robots and Electronic Brains. Transworld Publishers. London.

Schmalz, E. (1847). Wochenschrift für die Gesammte Heil kunde. No. 49. 3rd December. ed. Casper. pp. 785-789.

Schwenter, D. (1651). Delitae Mathematicae et Physicae der Mathematischen und Philosophischen Erquickstunden. Lesson XIV. Vol. II. Nuremburg.

Semon, R. (1923) Mnemic Psychology. London. Allen and Unwin.

Seymour, P. (1973). A model of reading, naming and comparison. British Journal of Psychology 64. 35-49.

Shannon, C. E. (1949). Communication theory of secrecy systems. Bell System Technical Journal 28. 656-715.

Shannon, C. E. (1950). A Chess Playing Machine. Scientific American. Feb. Vol. 182. No. 2.

Shaw, R. E. (1971). Cognition, Simulation and the Problem of Complexity. J. Structural Learning 4. pp. 31-44.

Shelly, C. E. & Stenhouse, E. (1911). Life and Health. Macmillan & Co. London.

Singer, B. R. (1976). Robert Hooke on memory, association and time perception. Notes and Records of the Royal Society of London 31. 115-31.

Sloman, A. (1975). Afterthoughts on Analogical Representation. in Schank and Nash-Webber.

Sorabji, R. (1972). Aristotle on Memory. London. Duckworth.

Sperling, G. (1960). The information available in brief visual presentations. Psychological Monographs 74. Whole No. 498.

Sturt, M. & Oakden, E. C. (1926). Modern Psychology and Education. Kegan Paul, Trench, Trubner & Co. Ltd. London.

Suppes, P. (1968). The desirability of formalisation in science. J. of Philosophy LXV. p. 663.

Suso, H. (1333-41). Horologium Sapientae. trans. French 1389.

Sussman, G. J. (1975). A computer model of skill acquisition. N.Y. American Elsevier.

Sutherland, N. S. (1974). Computer Simulation of Brain Function. in Brown, S. C. (ed.). 1974.

Swift, J. (1735). Travels into Several Remote Nations of the World. Dublin. Faulkner.

Temple, W. (1692). Essay upon ye Ancients and ye Moderns.

Tence, U. (1836). Portraits et Histoires des Hommes Utiles. Paris.

Thicknesse, P. (1784). The Speaking Figure and the Automaton Chess-Player Exposed and Detected. London.

Thomson, W. (1884). Lectures on Molecular Dynamics, and the Wave Theory of Light. Baltimore. John Hopkins Uni.

Thomson, W. H. (1907). Brain & Personality. N.Y. Dodd, Mead and Co.

Thorpe, W. H. (1963). Learning and Instinct in Animals.

Treisman, A. M. (1964). The effect of irrelevant material on the efficiency of selective listening. Am. J. Psych. 77. pp. 533-46.

Turing, A. (1950). Computing Machinery and Intelligence. Mind LIX. pp. 433-460.

Underwood, B. J. (1957). Interference and forgetting. Psych. Rev. 64. pp. 49-60.

Vallauri, M. (1960). See Encyclopaedia of World Art. Entry under Automata.

Vaucanson, J. de (1742). An Account of the Mechanism of an Automaton or Image playing on the German Flute. trans. Desaguliers, J. T. London. T. Parker.

Vinci, L. de. Codice Atlantico Folio 179. (Ambrosiana Library, Milan).

Vogt, K. (1845-9). Physiologische Briefe für Gebildete aller Stunde, p. 206.

Waller, R. (ed) (1705). The Posthumous Works of Robert Hooke. London. Smith and Walford.

Weizenbaum, J. (1966). Eliza. C.A.C.M. 9.

Weizenbaum, J. (1976). Computer Power and Human Reason. San Francisco. W. H. Freeman & Co.

Werner, H. & Kaplan, B. (1967). Symbol Formation. N.Y. Wiley.

Westlake, P. R. (1970). The possibility of neural holographic process within the brain. Kybernetik 7, 129-153.

Wilks, Y. (1976). Dreyfus's Disproofs. B. J. Phil. Science 27. No. 2.

Williams, B. (1973). How smart are computers? N.Y. Review of Books. Nov. 15. pp. 36-40.

Willis, J. (1621). The art of memory as it dependeth upon places and ideas. London. W. Iones.

Willis, R. (1821). An attempt to analyse etc. London. J. Booth.

Willis, T. (1672). De anima brutorum. Oxford.

Windisch, K. G. von. (1784). Inanimate Reason. London.

Winograd, T. (1972). Understanding Natural Language. Cognitive Psychology 3. No. 1.

Wittgenstein, L. (1921). Tractatus Logico-Philosophicus.

Wittgenstein, L. (1953). Philosophical Investigations.

Wood, D. and Shotter, J. (1973). A preliminary study of distinctive features in problem solving. Quarterly Journal of Experimental Psychology 25. 504-10.

Woodcroft. (1851). Edition of Hero of Alexandria's Pneumatics.

Yates, F. A. (1966). The Art of Memory. London. Routledge and Kegan Paul.

Zangwill, O. L. (1972). 'Remembering' revisited. Quarterly Journal of Experimental Psychology 24. 123-38.

Bennett, J. (1969). Rationality. London. Routledge and Kegan Paul.

Bower, T. G. R. (1978). The Perceptual World of the Infant. London. Fontana.

Burtt, E. A. (1932). The Metaphysical Foundations of Modern Physical Science, 2d ed. London. Routledge and Kegan Paul.

Premack, D. (1971). Language in chimpanzees? Science, 172: 808-22.

Young, R. M. (1971). Mind, Brain and Adaptation. Cambridge. C.U.P.