The Nature of Explanation - Kenneth Craik

First Published by Cambridge University Press 1943.

Chapter 5 - Hypothesis on the nature of thought

From this point onwards we are advancing a hypothesis and shall take the existence of the external world and of causation for granted.

One of the most fundamental properties of thought is its power of predicting events. This gives it immense adaptive and constructive significance as noted by Dewey and other pragmatists. It enables us, for. instance, to design bridges with a sufficient factor of safety instead of building them haphazard and waiting to see whether they collapse, and to predict consequences of recondite physical or chemical processes whose value may often be more theoretical than practical. In all these cases the process of thought, reduced to its simplest terms, is as follows: a man observes some external event or process and arrives at some 'conclusion' or 'prediction' expressed in words or numbers that 'mean' or refer to or describe some external event or process which comes to pass if the man's reasoning was correct. During the process of reasoning, he may also have availed himself of words or numbers. Here there are three essential processes:

(1) 'Translation' of external process into words, numbers or other symbols,

(2) Arrival at other symbols by a process of reasoning, deduction, inference, etc., and

(3) 'Retranslation' of these symbols into external processes (as in building a bridge to a design) or at least recognition of the correspondence between these symbols and external events (as in realising that a prediction is fulfilled).

One other point is clear; this process of reasoning has produced a final result similar to that which might have been reached by causing the actual physical processes to occur (e.g. building the bridge haphazard mid measuring its strength or compounding certain chemicals and seeing what happened); but it is also clear that this is not what has happened; the man's mind does not contain a material bridge or the required chemicals. Surely, however, this process of prediction is not unique to minds, though no doubt it is hard to imitate the flexibility and versatility of mental prediction. A calculating machine, an anti-aircraft 'predictor', and Kelvin's tidal predictor all show the same ability. In all these latter cases, the physical process which it is desired to predict is imitated by some mechanical device or model which is cheaper, or quicker, or more convenient in operation. Here we have a very close parallel to our three stages of reasoning-the 'translation' of the external processes into their representatives (positions of gears, etc.) in the model; the arrival at other positions of gears, etc., by mechanical processes in the instrument; and finally, the retranslation of these into physical processes of the original type.

By a model we thus mean any physical or chemical system which has a similar relation-structure to that of the process it imitates. By relation-structure l do not mean some obscure non-physical entity which attends the model, but the fact that it is a physical working model which works in the same way as the process it parallels, in the aspects under consideration at any moment. Thus, the model need not resemble the real object pictorially; Kelvin's tide predictor, which consists of a number of pulleys on levers, does not resemble a ride in appearance, but it works in the same way in certain essential respects-it combines oscillations of various frequencies so as to produce an oscillation which closely resembles in amplitude at each moment the variation in tide level at any place.

Again, since the physical object is 'translated' into a working model which gives a prediction which is retranslated into terms of the original object, we cannot say that the model invariably either precedes or succeeds the external object it models. The only logical distinction is on the ground of cheapness, speed, and convenience. The Queen Mary is designed with the aid of a model in a tank because of the greater cheapness and convenience of the latter; we do not design toy boats by trying out the different plans on boats the size of Atlantic liners. In the ' same way, in the particular case of our own nervous systems. The reason why 1 regard them as modelling the real process is that they permit trial of alternatives, in, e.g. bridge design, to proceed on a cheaper and smaller scale than if each bridge in turn were built and tried by sending a train over it, to see whether it was sufficiently strong.

Many mechanistic views of life and behaviour have been advanced, e.g. those of Hartley and Cabanis. But on the one hand there has been a tendency to assert a mechanistic theory rather than to regard it as a hypothesis which should, if followed out, indicate exactly how and where it breaks down; and on the other hand, there has been little attempt to formulate a definite plan of a mechanism which would fulfil the requirements. Hull has, however, made some models which show response to an altered Stimulus, or conditioning. 1 have not committed myself to a definite picture of the mechanisms of synaptic resistance, facilitation, etc.; but 1 have tried, in the succeeding pages, to indicate what 1 suspect to be the fundamental feature of neural machinery-its power to parallel or model external events and have emphasised the fundamental role of this process of paralleling in calculating machines. Thus, it is perhaps better to start with a definite idea as to the kind of tasks mechanism can accomplish in calculation, and the tasks it would have to accomplish in order to play a part in thought, rather than to draw analogies between the nervous system and some specific mechanism such as a telephone exchange and leave the matter there. A telephone exchange may resemble. The nervous system in just the sense I think important; but the essential point is the principle underlying the similarity.

Now it may be that a mind does not function only in this way; but as this is one way that 'works', in fact the only way with which we are familiar in the physical sciences, and as there is abundant evidence of the great mechanical possibilities of the nervous system, it does not seem overbold to consider whether the brain does not work in this way that it imitates or models external processes. The three processes of translation: inference, and retranslation then become the translation of external events into some kind of neural patterns by stimulation of the sense-organs, the interaction and stimulation of other neural patterns as in 'association', mid the excitation by these of effectors or motor organs.

Without enquiring into the relation between such neural patterns and the unitary symbols of thought-words, numbers, etc.- we can study to some extent the scope and limits of this modelling or imitative process, by studying the scope and limits of the two great classes of symbols-words and numbers.

Any kind of working model of a process is, in a sense, an analogy. Being different it is bound somewhere to break down by showing properties not found in the process it imitates or by not possessing properties possessed by the process it imitates. Perhaps the extraordinary pervasiveness of number, and the multiplicity of operations which can be performed on number without leading to inconsistency, is not a proof of the 'real existence' of numbers as such but a proof of the extreme flexibility of the neural model or calculating machine. This flexibility renders a far greater number of operations possible for it than for any other single process or model.

Of course we have still to face the question why these analogies between different mechanisms-these similarities of relation-structure-should exist. To see common principles and simple rules running through such complexity is at first perplexing though intriguing. When, however, we find that the apparently complex objects around us are combinations of a few almost indestructible units, such as electrons, it becomes less perplexing. For it is inevitable that processes corresponding to arithmetical addition of these elementary units-electrons, protons, etc. should manifest themselves in many instances. That is to say, if all pieces of pure iron consist of similar groupings of similar units, it is very likely that two pieces of iron placed end to end will add in length according to some simple law, and that pieces of other substances will do the same. The emergence of common principles and similarities is, then, not so surprising if it is shown that all substances are composed of similar ultimate units, for the appearance of uniformity and similarity is then the reappearance of a uniformity and similarity which were in fact ever present. We are still faced with the more ultimate question, why diverse materials should consist of combinations of a very few types of ultimate particles. The short life of some articles, such as positrons, suggests that in the vicinity of other particles, such as electrons, they are not stable; if it could be shown that in any such encounter the electron is more stable and the positron less, some kind of explanation would have been given as to why electrons are more frequent. It would still be conceivable that innumerable entirely different types of ultimate particle could have existed; if there is only one type in existence at a time there is nothing for it to be inconsistent with (apart from such factors as the mutual repulsion of similar particles and their consequent inability to form combinations). If, however, we conceive the world as made of a number of different types of ultimate unit, it is possible that they would prove to be mutually unstable and that all particles must acquire the same properties in order to exist, much as water in different tubes oil a common arm always finds a constant level.

This, however, is very speculative; the point of interest for our present enquiry is that physical reality is built up, apparently, from a few fundamental types of units whose properties determine many of tile properties of the most complicated phenomena, and this seems to afford a sufficient explanation of the emergence of analogies between mechanisms and similarities of relation-structure among these combinations without the necessity of any theory of objective universals.

We have now to enquire how the neural mechanism, m producing numerical measurement and calculation, has managed to function in a way so much more universal and flexible than any other. Our question, to emphasize it once again, is not to ask what kind of thing a number is, but to think what kind of mechanism could represent so many physically possible or impossible, and yet self-consistent, processes as number does.

The key may possibly lie in the following fact: in causal chains and physical or chemical combinations, the possibility of a given combination tends to be limited by other factors than the mere self-consistency of the combination. If you try to determine whether the series of integers can be extended to infinity by piling bricks on top of one another, you find that after a time the bricks fall down, or you cannot reach to pile any more up, or you run short of bricks or die; all these are extraneous difficulties. More subtle are the difficulties of adding nine oranges to nine apples, or of trying to produce a physical four-dimensional object. In all these cases we have not been satisfied with simply finding whether a given combination can exist along with other combinations; we have chosen a combination of combinations (i.e. a number of objects) which of course limits the number of possible self consistent combinations, just as in a game of rolling balls into grooves under a glass lid the number of times all are simultaneously in their grooves decreases as the number of balls is increased. In a mechanism such as a telephone exchange or a nervous system, where one is not trying to produce new objects but merely combinations of active or excited elements, the possible

combinations are at a maximum, limited only by remoteness of excited elements (vide failure of association) or decrease of excitation with time (vide forgetting). Even these difficulties can be to some extent overcome by further use of written and spoken symbols to act as a kind of reinforcing or relay system.

This greatly extended power is not unique to a mind; it could be illustrated by calculating machines. A machine working on a graphical principle might try to represent squaring and cubing by pointers moving along the x, y and z axes; it would inevitably come to a standstill or repeat itself when the volume of the cube equalled its own volume. On the other hand, a machine working on the principle of picking up gear-teeth by a repeated-multiplication process could go on raising any number to any power however large if it had sufficient dials on it.

It is likely then that the nervous system is in a fortunate position, as far as modelling physical processes is concerned, in that it has only to produce combinations of excited arcs, not physical objects; its 'answer' need only be a combination of consistent patterns of excitation-not a new object that is physically and chemically stable.

We have now to enquire what meaning causality, mean in implication, consistency and so forth can have when applied to such a mechanism. Again, our question is not 'What kind of thing is implication or causality?' but 'What structure and processes are required in a mechanical system to enable it to imitate correctly and to predict external processes or create new things?'

In examining this question, we can divide the process of thinking or reasoning into the same steps as before representation by symbols, calculation, and retranslation into events.

The diversity of calculating machines, languages and words for numbers shows that a relation can be represented in several symbolic ways. Unique determination is the main principle; a symbol, a setting of a machine, or a neural pattern is liable to be misleading if it represents two distinct types of physical things or events.

Causality in the external world would be represented by some (causal) process of interaction between excited elements m our own brains. As a result of such interactive or associative processes we might have, for example, A=B, B=C, A not= C where A, B and C are neural patterns claiming to represent external things or processes. These patterns clearly cannot all remain simultaneously excited; inconsistency means a clash in the interaction of patterns.

My hypothesis then is that thought models, or parallels, reality-that its essential feature is not 'the mind', 'the self', ‘sense-data’ nor propositions but symbolism, and that this symbolism is largely of the same kind as. that which is familiar to us in mechanical devices which aid thought and calculation.

I hope no one will be deterred by the idea that such a theory regards thought as an inactive halo round mechanical brain processes; for though my hypothesis assumes that thought processes and consciousness are dependent on mechanical processes, it, tries to discover what function consciousness does perform, by seeing where a purely mechanical process fads to meet the facts. As will be discussed ill Chapter VI, if it is true, it would be a hylozoistic rather than a materialistic scheme; it would attribute consciousness and conscious organisation to matter when it is physically organised in certain ways. However, these are remote speculations; the important point is to propound the theory and to consider ways of testing it.

We shall not consider purely speculative consequences of it, but only inferences which have some possibility of being experimentally verified, though we cannot claim that they provide critical tests of it. There remains the vitalist possibility -that life and mind, something different and aloof from physical matter, enters, and that we are misguided in our attempts to explain any aspects of conscious processes in terms of their material basis. But if so, the failure ought to show itself somewhere, if we proceed with due caution in the proposal and testing of hypotheses.

It is generally agreed that thought employs symbols such as written or spoken words or tokens; but it is not generally considered whether the whole of thought may not consist of a process of symbolism, nor is the nature of symbolism and its presence or absence in the inorganic world discussed. Further, it has been usual to restrict the word 'symbol' to words or tokens, which still leaves the processes of the relating of words to form sentences and the processes of inference and implication mysterious and unique. Let us consider whether these processes are not paralleled by familiar mechanisms.

First, we have seen that the possibility of verbal or other symbolism is the fundamental assumption of all philosophy communicated by anyone to anyone else. Without falling into the trap of attempting a precise definition, we may suggest a theory as to the general nature of symbolism, viz. that it is the ability of processes to parallel or imitate each other, or the fact that they can do so since there are recurrent patterns in reality. The concepts of abilities and patterns and formal identity in material diversity are all hard ones; but the point is that symbolism does occur, and that we wish to explore its possibilities. There are three main steps: first, is there any evidence of such symbolism in inorganic nature? Secondly, do we ourselves employ such symbolism in thought? And thirdly, is there any evidence that our thought processes themselves involve such symbolism, occurring within our brains and nervous systems?

There are plenty of instances in nature of processes which parallel each other-the emptying of pools and the discharge of a cat's fur which has become electrified, the transmission of sound and electromagnetic and ocean waves, and so forth. As mentioned above, human thought has a definite function; it provides a convenient small-scale model of a process so that we can, for instance, design a bridge in our minds and know that it will bear a train passing over it instead of having to conduct a number of full-scale experiments; and the thinking of animals represents on a more restricted scale the ability to represent, say, danger before it comes and leads to avoidance instead of repeated bitter experience. In inorganic nature, because of its simpler organisation, we should expect this function to be less fully exemplified. Indeed, there are very few examples at all. Perhaps the nearest approach is the fine trickle of water which first finds its way from a mountain spring down to the sea and smoothes a little channel for the greater volume of water which follows after it. But the material of symbolism the parallel mechanisms-seem to be there; it is only the sensitive 'receptors' on matter, and means of intercommunication or nervous system, which are lacking.

Again, there is no doubt that we do use external and mechanical symbolisation to assist our own thinking. Provided with a piece of paper we can perform long and complicated calculations which would be impossible in our heads; and the Busch differential analyser will solve problems which could not be tackled by any other method.

Finally, there is some, though scanty, evidence from anatomy and electrophysiology that our nervous systems do contain conducting sensory and motor paths and synapses in which there occur states of excitation and volleys of impulses which parallel the stimuli which occasioned them; so that, as far as experimental evidence goes, this symbolisation is found to occur in the central nervous system. But what produces and occasions it, on such a mechanistic theory? In any

mechanical system, the events which occur are those which result in the greatest possible equalisation of energy-roughly speaking, the reactions take the path of least resistance. If parts of an organisation are interconnected by a system of communication such as the nervous system, the reactions can be directed along the 'lines of least resistance' by the expenditure of a very little energy in the appropriate 'lines of least resistance' in the nervous system. The situation is enormously complicated by natural selection, which causes the survival of certain organisms-those for instance, in whom the passage of the 'monitoring' nerve impulse results in such activity of the whole organism as will tend to preserve it. In general, it is much more illuminating to regard the growth of symbolising power from this aspect of survival-value, rather than from the purely physical side of accordance with thermodynamics; but it does not seem that there is any inconsistency between the two.

Thus there are instances of symbolisation in nature; we use such instances as an aid to thinking; there is evidence of similar mechanisms at work in our own sensory and central nervous systems; and the function of such symbolisation is plain. If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilise the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it. Most of the greatest advances of modem technology have been instruments which extended the scope of our sense-organs, our brains or our limbs. Such are telescopes and microscopes, wireless, calculating machines, typewriters, motor cars, ships and aeroplanes. Is it not possible, therefore, that our brains themselves utilise comparable mechanisms to achieve the same ends and that these mechanisms can parallel phenomena in the external world as a calculating machine can parallel the development of strains in a bridge?