Perception, 1983, volume 12, pages 233-238

Editorial

Forty years on: Kenneth Craik's *The Nature of Explanation* (1943)

Kenneth John William Craik (1914-1945) was an outstanding visual physiologist, especially on mechanisms of dark adaptation, and he was more than a scene-setter for computer-based psychological accounts of perception, learning, and skill. His early training in philosophy at Edinburgh was useful, as his only complete book, *The Nature of Explanation* (1943), demonstrates. It shows the power of explicit analogies from technological principles, as its central concepts are drawn, with brilliant imagination and cogent arguments, from the then newly invented analog predictor mechanisms of the Second World War. As the first Director of the Medical Research Council's Applied Psychology Unit (the APU), in Sir Frederic Bartlett's Department at Cambridge, Craik was concerned with visual problems such as glare and adaptation, and also with matching new kinds of machines to men and comparing their performances. The Craik Laboratory at Cambridge is, of course, named after him. Tragically, he was knocked off his bicycle in Cambridge and so died, at the age of 31. Craik's ideas were hardly known in the United States until several years after his death, and therefore they were not directly influential in the American development of Cybernetics; but Warren McCulloch learned of Craik's work, and arranged for his papers to be collected by Steven Sherwood, and published under the title *The Nature of Psychology* (1966).

Here I shall be concerned only with Kenneth Craik's *The Nature of Explanation* (1943), which introduces his notion of cognitive brain function in terms of physical 'internal models'. It was written just before the impact of digital computers, so Craik necessarily based his ideas on analog devices. I may touch on how conceptual models are often—perhaps always—based on technologies and here on the special significance of electronic programmable digital computers compared with analog computing for models of brain function. Craik described the brain's internal models thus (page 57):

> "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".

But first he asked why explanations of any kind are sought—describing an explanation as "a kind of distance-receptor in time, which enables organisms to adapt themselves to situations which are about to arise" (page 7). He broadens this notion, to extricate us from the biological straightjacket, by suggesting that "apart from this utilitarian value it is likely that our thought processes are frustrated by the unique, the unexplained and the contradictory and that we have an impulse to resolve intellectual frustrations, whether or not there is a practical problem that needs a solution".

This plea for physical explanations of cognition does not, however, mean that

> "it is useless or incorrect to give apparently non-physical clinical explanations of psychological phenomena—for instance, to say that an unpleasant experience or shock may *cause* amnesia or suppression. This is a correct statement of the phenomena as far as it goes; but we are entitled to go further if we can. If we then find a more ultimate physical and physiological train of events to be invoked 'in between' the shock and the suppression, we should regard this as a more ultimate part of the mechanism, just as it is correct to say that the pressure of one's finger causes current to flow in the windings of the starting motor and still more fundamental to give an account of the flow of current and torque exerted by the motor in terms of electronic and electromagnetic theory".

Here we see clearly what he is getting at, and why he disagreed with Sir Frederic Bartlett for suggesting that a mechanistic physical explanation particularly in psychology may merely be more complicated. For Craik a mechanistic explanation "covers the most facts by the fewest numbers of postulates and leaves the fewest anomalies outstanding". He adds the interesting thought:

> "This adequacy of and freedom from anomalies is also the credential of the particles as being ultimate, remote as they may seem from everyday life. For, once rigid causality is admitted, whether it be thought to be mechanistic or not, anomalies must be regarded as cases where explanatory concepts are wrong ...".

So he begins to separate internal models, causal though they are, from the external physical world. This leads to the body of the book (chapter v): "The Hypothesis on the Nature of Thought".

Reminding us of the importance for survival of predicting events, Craik considers the use of physical models (which for him can serve as symbols) and of words and numbers for prediction and reasoning—with three essential processes:
(i) 'Translation' of external process into words, numbers or other symbols.
(ii) Arrival at other symbols by a process of 'reasoning', deductive inference, etc.
(iii) 'Retranslation' of these symbols into external processes (as in building a bridge design) or at least recognition of the correspondence between these symbols and external events (as in realising that a prediction is fulfilled).

Then comes the critical step (page 51): "... this process of reasoning has produced a final result similar to that which might have been reached by causing the actual physical process to occur (eg building the bridge haphazard and 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 chemical". Then:

> "... 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 which it is desired to predict is *imitated* by some mechanical device or model which ever is cheaper, or quicker, or more convenient in operation".

This, Craik says, is very similar to the three stages of reasoning, as the external processes are 'translated' into positions of gears etc in the model; the arrival at other positions of gears, etc by mechanical processes; and "finally the translation of these into physical processes of the original type".

Here we reach the nub of Craik's account of how internal models are mindful:

> "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' I do not mean some obscure non-physical entity which attends the model, but the fact that it is a working physical model which works in the same way as the process it parallels, in the aspects under consideration at the moment".

The model does not have to resemble the object or situation pictorially: "since the physical object is 'translated' into a working model which gives a prediction ... we cannot say that the model invariably precedes or succeeds the external object it models. In the case of the nervous system, models permit trial of alternatives ...".

He does not commit himself to any physiological account of how brain models are realized physically; the concern is only with underlying similarities of function, though there may be very large surface differences. So one could not see how they function from anatomy, however detailed. Like any other analogy, the brain model's use is "bound somewhere to break down ...". Hence a rich bunch of thinking errors and perceptual illusions, which are endemic to any cognition.

The account is in terms of the analog devices of his time which did not have the flexible symbol-handling powers of the electronic digital computers that soon followed. So it is interesting to see how Craik treats numbers and words with his analog account of brain function. (Imagine an analog word processor!) Craik denies that numbers have 'real existence', and he puts much more weight than we would now on restraints of the mechanics of the device. Thus, Craik does not think that the great range of applications, without leading to inconsistencies, is a proof of 'real existence' of numbers; but that it may, rather, suggest that the neural model or the machine may have extreme flexibility. (This he attributes tentatively to there being only a small number of functional units—like the range and lawfulness of objects composed of only a few kinds of atomic particles.) In any case, rather than asking "what kind of thing a number is, he considers instead, "what kind of mechanisms could represent so many physically possible or impossible, and yet selfconsistent, processes as number does". Implication is also described in these terms, as being "a kind of artificial causation in which symbols connected by rules represent events connected by causal interaction ..." (page 63).

It was almost impossible to see at that time how a machine could generalize. Craik points out that recognizing objects from different positions, or from stimulation of different retinal regions, is easy for an organism but difficult for a machine; and that this may indicate a fundamental difference between "recognition in its physical and psychological senses; on the other hand, they may only show that a different form of mechanism is involved in psychological recognition". Craik reminds us here of Lashley's warning that lack of discrimination in the response of an animal may give its behaviour a false appearance of generality and abstraction; but Craik urges that properties of objects can be "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". He is, though, clearly worried as to how generalizing power could be achieved by a conceivable machine, for example to recognize objects over various distances and orientations—for here he is tempted to invoke consciousness as a supermechanical causal principle for perception (page 68):

> "This may be one of the functions of consciousness—to permit greater 'elasticity' and flexibility and unity of response than the known properties of co-ordinations of mechanisms will accomplish".

He tries to resist this postulate of active powers of consciousness taking over where mechanisms seem inadequate, by looking at characteristics of some selected mechanisms (page 71): "Sometimes a simple mechanical device will show this power in high degree. Thus, an inclined plane will 'recognize' spheres, of whatever size, material and colour, and 'distinguish' them from cubes, since the former will roll down it while the latter will not". He concludes though that recognition in man and animals is too flexible and adaptable to be accounted for by "a few specific and special mechanisms of this kind".

So consciousness as an active principle is not entirely rejected. But however this may be, Craik sees that the internal-model brain mechanisms must work causally in physical terms though they *represent* reality rather being linked in the usual causal way to the events of the external world. Thus Craik says (page 81):

> "Language must evolve rules of implication governing the use of words, in such a way that a line of thought can run parallel to, and predict, causally determined events in the external world. The ability of a particular 'line of thought' to do this is the test of its correctness as an explanation".

Then he raises a likely criticism—or doubt that internal models (or cognition) are needed:

> "Some may object that this reduces thought to a mere 'copy' of reality and that we ought not to want such an internal 'copy'; are not electrons, causally interacting, good enough? Why do we want our minds to play the same sort of game, with laws of implication instead of causal laws to determine the next step?"

Craik's reply is surely an essential part of the deep answer to why we have cognitive processes. His answer is in two parts:
(i) That an internal model can predict events that have not yet occurred.
(ii) That the interactions of electrons do not tell us what we want to know—for, though the 'machine' (Nature) works equally well whether we are ignorant of steps in it or not, scientific thought does not work equally well if there is a gap in the chain.

There is uncertainty, though, over the status of symbols and how mechanisms can represent. He explains the so-to-say unNatural precision of geometry, which is far greater than any perception or measurement allows, by analogy with administrative law—that for example conscription age is fixed by a precise date of birth, rather than by physique, or whatever. So the brain's internal models are seen to legislate and impose their legislation upon how we see—as well as free us from causal nature though they function by physically causal mechanisms. We can hardly accept Craik's account of geometry, though, for why should such 'legislation' turn out to be causally predictive, for what objects turn out to be like on independent grounds? The account does not look right at this point.

Undoubtedly these ideas were triggered largely by the new technology of the Second World War, in which Craik's own contribution was considerable, and his understanding profound. He was writing when gun predictors acting from minimal received signals were new and had striking conceptual significance. Craik was surely right to accept the principles by which they worked not merely as analogies, but as *examples* of what he conceived for the brain. The same is so for the concept of servocontrol for limb movements—there are actual biological servosystems, not mere analogies of what are realized in man-made machines. Likewise, Craik supposes that hundreds of millions of years ago brains incorporated technologies we have invented over the last few centuries.

So, forty years ago Kenneth Craik transformed the technology of his day into the image of man; as, long before, Homer's Hephaestus created Pandora from fire, and Pygmalion breathed life into the ivory statue of Galatea. But the technology of Craik's time, though so recent, was not the same as ours. There is indeed a smaller jump from Homer in 700 BC (who would have known the abacus) to the computing of the 1940s, than from the computing devices familiar to Craik to ours today.

It has only recently become clear that computers can be extremely flexible and adaptable, and the difficulties Craik raised for generalization have now been largely solved by AI algorithms. So we are less tempted to fall back on powers of consciousness to explain perceptual generalization, and abilities to accept and create different perceptual scales. (One might almost say that the scales have fallen from our eyes!) Paradoxically, though, this makes consciousness even *more* mysterious—for what does it do?

Many people now see mind as patterns and powers of symbols handled by software digital brain processing: so now Craik's account of symbols in relation to mechanisms without software hardly looks right. Software confers infinitely flexible restraints; but Craik had to rely on hardware characteristics which would not at all necessarily be appropriate for the relation-structures of his internal models. How would appropriate relation-structures be selected according to need? We see digital

programming as conferring the necessary flexibility and appropriateness as the machine changes according to needs while at the same time it serfes the symbols—in a unique symbiosis.

Is this how it will end? Or will we come to see as a quaint idea digital brains going through the precise tortuous *man-made* steps of elaborate mathematics—even to waggle or see a finger? Will something less analytically digital and perhaps more crudely analog return for accounts of brain function? Or will some new technology arise so that we come to see perception in a new light? Will the brain turn out to be millions of interacting special-purpose analog computing units functioning without the analytical steps of software, much as Craik supposed?

However this may be, in just over a hundred pages, Kenneth Craik distilled a great deal of thought which is still fermenting and inspiring now—forty years on.

Richard L Gregory

References

Bartlett F C, 1946 "Obituary notice: K J W Craik" *British Journal of Psychology* **36** 109-116
Craik K J W, 1943 *The Nature of Explanation* (Cambridge: Cambridge University Press)
Craik K J W, 1966 *The Nature of Psychology: A Collection of Papers and Other Writings by the Late Kenneth J W Craik* edited by Stephen Sherwood (Cambridge, Cambridge University Press)
Zangwill O L, 1980 "Kenneth Craik: the man and his work" *British Journal of Psychology* **211** 1-6

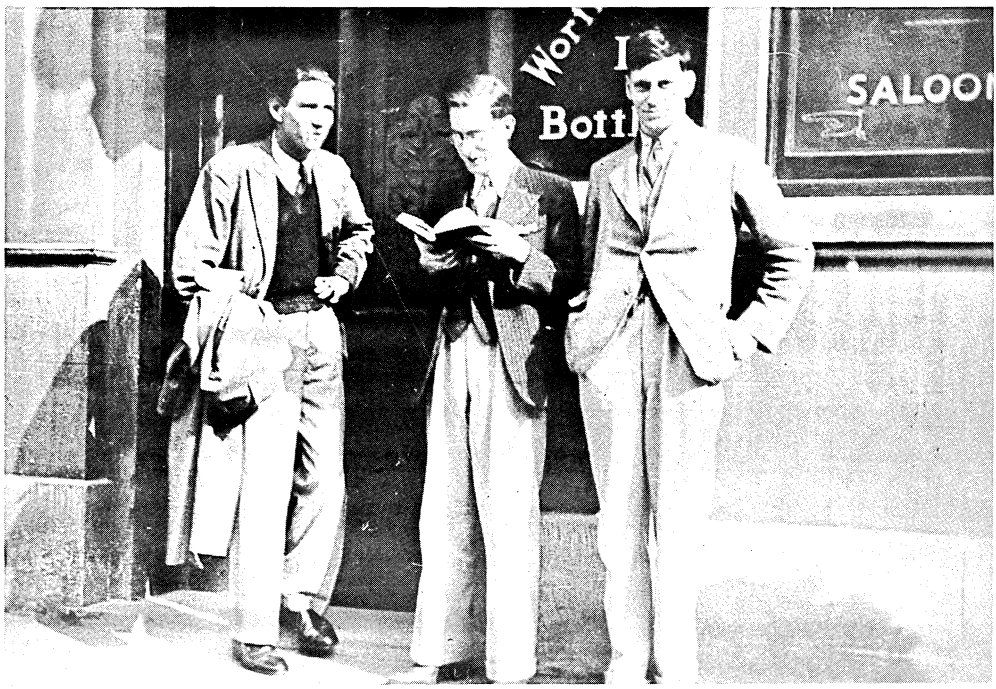

Figure 1. Oliver Zangwill, FRS, until recently Professor of Psychology at Cambridge. The late Carolus Oldfield, Professor of Psychology at Oxford. Kenneth Craik is on the right. They are standing outside the "Bun Shop" in Cambridge. This famous Pub is, alas, nor more. The photograph was taken in about 1937, or perhaps a little later.