

A MATHEMATICAL EXPLANATION OF CREATIVITY

By E.C. Zeeman

To explain the mechanism underlying creativity, let us assume (1) that the brain is a dynamical system. More precisely the states of the brain are represented by points of a multidimensional space X . For example, if we take as coordinates the firing rates of the different neurons, then X has dimension 10,000,000,000. Alternatively we could go for more detail and take coordinates representing the various chemical concentrations in each part of each cell, and their rates of change. It does not matter which.

As time progresses the changing state is determined by a flow on X , representing memory. Next we assume (2) that this flow is stable, that is to say it preserves its quality under small perturbations. This assumption is justified because our memories are surprisingly stable. With only these two assumptions we can apply some remarkable new theorems of the Russian mathematician, Anosov, and the American mathematician, Smale. It appears that in a stable system, the state flows towards certain stable Anosov subsystems, called attractors. Each attractor is multidimensional, and packed with periodic and almost periodic orbits. The EEG evidence suggests that if the state of the brain flows round a periodic or an almost periodic orbit, then the mind is thinking of an idea, and so we make this our assumption (3). Therefore an attractor represents a body or context of ideas, and with only these three assumptions we can now predict the following behaviour.

The mind will seldom stay still and will tend to jump from idea to idea, but tend to remain within the same context of ideas. Also the mind can think of the whole context of ideas as a single idea within that context (rather like the Russell paradox of a set being an element of itself), which is a very significant ability if the mind is to develop any complicated architecture of thought.

Finally we assume (4) that two states can be superposed. The justification for this last assumption comes from the physiology: any state of the brain is ultimately caused by various chemical concentrations all over the place, and the superposition of two brain states is merely the result of combining the local chemical states. With just these 4 assumptions we can describe automatic mechanisms for the laying down, recall, destruction and simplification of memories, and for the creation of new ideas, as follows.

The superposition of states means that two attractors can be multiplied together, corresponding to the mind thinking of two thoughts at once and creating an associative memory (putting two and two together). However in the dynamical system the product of two attractors is unstable, and any arbitrary slight perturbation will cause the product to break up into new stable attractors (of lower dimension than the product). It is these new attractors that represent the new associative memory. In future, if either of the two original thoughts are presented to the mind, then the state of the brain is swept towards the new attractor, and so both thoughts are automatically brought to mind, explaining recall.

In the above explanation perhaps the most interesting words are those in italics, because they mean that the brain will work satisfactorily whatever tiny synaptic changes are taking place, be they facilitation or the opposite habituation or any mixture. Most brain theories assume some logic to the synaptic changes, but we assume none, and consequently the problem of explaining evolution and morphogenesis of the brain is much simpler.

Let us return to the phenomenon of the break-up of the product into new stable attractors because this is the essence of creativity. The shape of a new attractor is unpredictable, and there are an infinite number of new types possible. Therefore the brain is capable of automatically manufacturing ^{new} attractors which have never existed in a brain before. In other words the mathematics tells us

that the mind will be unpredictable, and capable of creating new contexts of ideas which have never been thought before. The newness concerns the whole context of ideas, rather than a single idea, (the whole attractor rather than a single periodic orbit). Each mind can only produce a finite number of new contexts, but these are drawn unpredictably from an infinite bank of possibilities.

What we have described above is the creativity associated with a "flash of insight". We now examine the creativity associated with the "growing conviction" or "mature reflection". A stable attractor is stable under small perturbations, but as the brain is being used all the time for all sorts of things, the dynamical system may be perturbed past the point of stability for this particular attractor. In which case a break-up again occurs, unpredictably, into lower dimensional attractors. Paradoxically these new attractors are more attractive, because fewer ideas are competing for the same basin of attraction. This explains the destruction and simplification of memory, why childhood memories become fewer, but those that are remembered stand out in sharper detail. Similarly if a mind dwells upon a large body of ideas and experiences, the dynamical system of the brain will build huge attractors, which in time will break up into fewer smaller very stable attractors. Therefore although the mind may start out with somewhat diffuse and perhaps confusing impressions, it will tend to harden into certain principles, prejudices and unshakable convictions.

For the development of this theory I am indebted to conversations with René Thom.

10th October 1968