Vehicle 10

Getting Ideas

The time has come to sit back and consider the strange variety of vehicles that populate our laboratory. They all go about their business according to certain rules, some of which we understand, because we invented them ourselves, and some we don't, because they emerged from a sort of Darwinian evolutionary process. The objects of their interest are defined by simple properties such as smell and color, or by more abstract properties, such as the periodicity of their coloring or the symmetry of their outline. Formal properties may stand for even more abstract definitions, as we have seen in the case of bilateral symmetry signifying the situation of "somebody having me in mind."

Some of our vehicles seem to move around smoothly, as if attracted and repelled by the sources of various fields of force superimposed on one another. Others appear to make sudden decisions, rousing themselves from a rather phlegmatic condition to take off on isolated ventures, after which they resume their state of rest. The vehicles seem to know their environment rather well, so much so that they are able to reach some objects with closed eyes, so to speak, apparently on the basis of some internalized map on which the object's location is recorded. On the whole, these vehicles

are surprisingly smart, especially considering the limited amount of intelligence that we, their creators, have invested in them.

But do they think? I must frankly admit that if anybody suggested that they think, I would object. My main argument would be the following: No matter how long I watched them, I never saw one of them produce a solution to a problem that struck me as new, which I would gladly incorporate in my own mental instrumentation. And when they came up with solutions I already knew, theirs never reminded me of thinking that I myself had done in the past. I require some originality in thinking. If it is lacking, I call the performance at best reasonable behavior. Even if I do observe a vehicle displaying a solution to a problem that would not have occurred to me, I do not conclude that the vehicle is thinking; I would rather suppose that a smart co-creator of vehicles had built the trick into the vehicle. I would have to see the vehicle's smartness arising out of nothing, or rather, out of not-so-smart premises, before I concluded that the vehicle had done some thinking.

But this does not mean that we cannot create vehicles that would satisfy this condition. We shall do this gradually, starting with the problem of *having ideas*. Let us take one of the vehicles of type 7, the ones with the Mnemotrix connections that introduce the effects of experience into the brain. This vehicle has been around for some time and has absorbed a great deal of knowledge about the world. This knowledge takes the form of statistical correlations between elementary events in the vehicle's sensory spaces or statistical correlations between more complex events represented in some threshold devices of its interior (or between elementary events and complex events).

Suppose the vehicle has learned that certain objects, A, B, C, D, are situated near the rim of the table top on which it lives: a broken-down vehicle, a light, a battery, a hill, a supply of screws. It has learned to associate these objects with the concepts "margin of the universe" and "dangerous cliff." On its occasional excursions to-

ward the margin of the universe, it will also have noticed the neighborhood relations among pairs of these objects: the screws are next to the hill, the light next to the battery, and so forth. One day, after enough excursions, the vehicle will suddenly realize that all these paired associations (A next to B, B next to C . . . , Z next to A) make sense if the whole situation is seen as a closed chain. The vehicle now has the idea of a finite bounded universe, with objects A to Z marking the marginal closed line. Once this "image" or "idea" is generated out of individual items of knowledge, it is there to stay. It may, in fact, be immediately recorded on the maps, whose use we have discussed. If so, we will observe that forever after the vehicle moves around much more expertly.

We must be careful, however, not to let the process of acquiring new ideas interfere with the detailed knowledge that our vehicle has assiduously collected and carefully stored in many associative connections during its lifetime. We know that this may happen in humans who are overly dedicated to the development of ideas. They tend to connect many individual cases into general categories and then use the categories as if they were things, losing the potential for categorizing in other ways by remembering each instance.

In the example of the discovery of the margin of the universe, I can see this danger. The idea of a closed chain of objects may be so strong that it keeps the images of these objects permanently active in the vehicle's brain. The consequence is that associations will develop between every object on the margin—and every other object on the margin. The serial order that led to the original idea will thereby be lost or at least submerged in a system of much stronger, massive associations. The way out in our case would be to let the excitation circulate in the closed chain associations. This would strengthen the associations representing the serial order of the objects and would not allow cross-association to develop.

Here are some more examples of ideas that may arise in vehicles. There are coins lying around on the floor in the universe of the

vehicles. Some of the coins are decorated with a picture of a human head and others are decorated with a number. One of our vehicles has already learned to recognize and to distinguish the two types of coins. That is, there are distinct patterns of activity, say two different threshold devices becoming active when one or the other kind of coin is seen by the eye of the vehicle. Now one of the coins showing a human head is flipped around by the vehicle—and suddenly it shows the number. This happens again and again until, by the learning process that we have already incorporated in our vehicles, an association is formed: "head, flipping, number." Of course, the association also works the other way around. Once the association is acquired, the vehicle knows that, after adding the action of flipping to the sighting of "number," the picture of the head will be seen. It may also be reinforced by the contrary experience, when the flipping of coins showing the number reveals the head.

We may call the whole complex of head-flipping-number and number-flipping-head the idea of a coin with two faces. It arises in the vehicle although the two faces of a single coin are never seen together. The idea of a coin with two faces can arise even if there are some coins around with human heads on both sides, as long as these coins escape the vehicle during the phase of "getting the idea."

Here's another example. Moving through a garden, a vehicle finds out that flower number one of a row is a source of food, flowers 2 to 7 are not (they are poisonous), flower number 8 is again a source of food and so are flowers 15, 22, and so forth. After a while it may happen that in the brain of the vehicle only one of 7 threshold devices (connected in a circular fashion) always becomes active in temporal coincidence with the finding of a source of food in a flower. This is again "getting the idea": that particular threshold device will be associated with the food finding system—with the consequent advantage of being able to predict sources of food without having to invest much energy in the process of sniffing

around. We must suppose, of course, that the time it takes for one threshold device to become active after another has been activated is exactly the same as the time it takes to get from one flower to the next or, better still, that the advancement of activity by one step in the ring of threshold devices is triggered by each flower.

All of this is not complicated in principle but boring to carry out in detail. We rely on the process of Darwinian selection that, starting with the vehicles of type 6, has introduced a great variety of different patterns of connections into the vehicles without our even noticing it (although we do recognize the vehicles' greatly increased complexity of behavior). We can well imagine that the vehicle could get the idea "edible flower" even if the only flowers that were edible were those whose ordinal numbers were square or whose ordinal numbers were prime. There is, however, a complication in the cases of squares and of prime numbers. If these numbers get too large, the vehicle has to perform a long and intricate dance between one flower and the next in order to find out whether the flower's number is square or prime, leaving marks on the earth and retracing them according to complicated rules. We have seen this before, at the end of the chapter on Vehicle 5, which also had its limitations. No such difficulty arises if the vehicle has to find out whether a number is even or odd, or whether a number is a multiple of six or of eleven, as long as the vehicle can count to eleven.

In this chapter we were only interested in the general idea of "getting ideas." Readers who want to know exactly what kind of network of threshold devices is necessary to calculate numbers that are square, or prime or whatever, must read the textbooks of automata theory.