Vehicle 11

Rules and Regularities

Most of you will not yet be convinced that the process of getting ideas as it was described in the previous chapter has anything to do with thinking. It is not surprising, you will say, that occasionally something clicks in the workings of a fairly complicated brain and from then on that brain is able to perform a trick (an algorithm, as some people say) that can be used to generate complicated sequences of numbers or of other images. It is also not surprising that these may occasionally match sequences of events or things in the world of the vehicle.

I will show you that this is just one step in the direction of creating behavior akin to thinking. In the following chapters we will introduce more elements of the thought process, making new vehicles to show new tricks, new types of performance. In the end our vehicles will surprise us by doing some real thinking.

We want to equip Vehicle 11 with a brain about which it can be said—in a more radical way than it could be said about previous editions—that it is a model of the world. We already introduced partial aspects of this model idea, when we talked about the usefulness of internal maps representing external spaces (Vehicle 8), and when we described a learning process (Vehicle 7) that discovers things in the environment and establishes their internal models (called concepts). But this is not enough. These things move around, bang against each other, associate and dissociate, grow and break. We have altogether missed these dynamic aspects up to now.

We will introduce these dynamics by improving on the system of Mnemotrix connections already introduced in the type 7 vehicle. You will remember that these connections between elements in the vehicle's brain were of different strengths and could be made more effective when the elements they connected were often activated together. This turned out to be very convenient, because so many of the facts about the world that are interesting and important to us (and to the vehicles) may be expressed as things or events that tend to occur together. For this reason it is unlikely that we will give up the trick of associative learning in any further development of more refined vehicles.

But we soon discover that there are important pieces of knowledge about the world expressed in a different form: events that do not present themselves at the same time but in succession—pairs of events, of which one is always the first and the other the second, like lightning and thunder, swinging a hammer and hitting the nail, or, in the world of vehicles, meeting a source of food and tasting the food. When we discover a pair of such events, we tend to think that one is the cause of the other, whatever that means. But this may lead to wrong interpretations, for instance when both events are produced by a third hidden event, only with different delays. Most of the time, however, when two events regularly occur in succession, it is no accident. And it certainly is useful for a vehicle to know what to expect when events occur that have important, possibly dangerous, consequences.

We could use our old supply of Mnemotrix wire together with a little electronics to incorporate into the vehicles' brains all those delayed coincidences of events we have been describing. What we want to achieve is a connection between the two internal representatives of an event A and an event B such that, when the representative A is activated by the input, the representative B is activated by the connection, but not vice versa. The connection would then represent the fact that "B often follows A" or, if you wish, the causal tie between A and B. This would force us to do a rather complicated wiring for every such connection. In order not to burden our constructive imagination too much, we prefer to buy a different sort of wire, called Ergotrix, which conducts in one direction only and has an increased conductance when it is interposed between elements that are active in succession within a brief time. We must be careful, of course, to install the wire in the right direction, conducting from the element that tends to be active first to the one that tends to be active second.

Once again we will see to it that all of this happens automatically. Plenty of Ergotrix wire will be installed between as many elements as possible so that whatever sequences occur can be recorded in the system. Of course there will be no lack of opportunity for learning. With all the movement in the world around the vehicle, with all the natural laws operating, and with all the other vehicles displaying fairly regular behavior on the basis of all the tricks that we (or the processes of evolution) have built into them, many sequences of events will repeat themselves and they will be worth learning.

You may ask why we did not use Ergotrix wire in the first place (Vehicle 7) when we first gave our vehicles the capacity to learn, starting with those complexes of properties that frequently occur together because they belong to one "thing." We used the Mnemotrix wire, which is ideal for associations, because it couples elements in a symmetrical fashion; once coupled, each of the properties can recall the other in quite the same way. For each Mnemotrix connection we could have used two Ergotrix wires (one for each direction) to obtain almost the same result. But there are two reasons to leave things as they are.

First of all, we don't want to go back in evolution and change things that have already proved to be convenient, since we might lose some advantage that we have not even realized. (Remember the law of uphill analysis and downhill synthesis: we run the risk of not understanding any longer what we previously put together.) Second, it is probably a good idea to keep the two processes conceptually separated—the associations of elementary properties into things or concepts on one hand and the sequencing of concepts on the other hand, one controlled by the Mnemotrix, the other by the Ergotrix system. The two kinds of learning produce two different kinds of knowledge, like geography and history, or systematic zoology and animal behavior, referring to what kinds of things exist and to how they develop and interact.

If we let our imaginations go and try to work out in detail what kinds of things the Mnemotrix system will discover in a real world, and what kinds of dynamic laws will be incorporated in the Ergotrix system, we soon discover that the two kinds of knowledge are perhaps related more than we had assumed initially for reasons of conceptual convenience. First of all, it would seem that the process of abstracting things from the environment—concept formation at the most elementary level—must occur prior to the process of discovering the dynamic properties of these things. For the laws of successions of events refer to the development and to the combination of things rather than their elementary properties. This is familiar from our own human experience: listening to a new language we want to learn, we must first discover individual words, or roots of words (something like the morphemes in linguistic terminology), before we can even hope to discover the rules that govern their use. Also, in the development of a science it is often apparent how the discovery and denomination of phenomena precedes the definition of the laws of their transformation. Chemistry had to go through a descriptive phase before the physics underlying the variety of substances could be understood. Zoology had to be taxonomic before it was organized by the theory of evolution.

On the other hand, purely descriptive classification is not only boring, it is also potentially misleading. It may lead to the wrong categories when it is not guided by at least the intuition of a theory of the underlying processes. A century of microscopic anatomy has filled the libraries with thousands of beautifully illustrated volumes that are now very rarely consulted because the descriptive categories of the old histology have been largely superseded by the new concepts of biochemical cytology. The example from linguistics that we have just mentioned may well serve to prove the contrary point, with word roots–morphemes–words as the segments of speech that must be learned. While it is true that these chunks of meaning in some languages (largely in English) coincide with acoustically well-defined episodes (the syllables, which the naive listener can recognize), it is certainly true that a better, more general definition of morphemes or words is derived from grammar. Words (I use this term loosely) are the segments of speech that we discover as the ultimate particles of grammar. If we had no idea or no experience of grammar, we might never discover that these are the pieces that are shuffled around to form sentences. We might propose a different, incorrect segmentation of speech, for example, a segmentation into syllables in a language with polysyllabic words. Words become meaningful insofar as they are used in a grammatical system.

In other words, abstracting meaningful chunks from the environment (things, events) and discovering the rules of their behavior are two processes that condition each other and are necessarily interlaced, like the learning of the vocabulary and the learning of grammar in a language course.

Coming back to Vehicle 11, it seems like a good idea to let the discoveries of the Ergotrix system influence the learning process in the Mnemotrix system, on whose initial abstractions it in turn depends. I don't want to work this out in detail, but something like the following scheme would clearly be possible. We have already described the conditions for the strengthening of an Ergotrix wire. These conditions are fulfilled when an element, say a threshold device, at one end of the wire becomes active shortly before another element becomes active at the other end. We have also seen that it is mostly groups of such elements, strongly interconnected and representing "things," that become active in succession. Now let's introduce the rule that whenever the Ergotrix wires become strengthened, the Mnemotrix wires within each of these groups will also become strengthened.

Thus concepts are established in the vehicle especially when they appear in regular sequences. How would this look to us? We would notice, observing the apparently erratic behavior of a vehicle in its world, that the vehicle displays particularly well defined reactions to events that are known to have consequences. Take, for example, a vehicle approaching an obstacle at high speed. We would not be surprised to see the vehicle promptly react to its perception of the danger of a collision. Similarly, Vehicle 11 will quickly remember which of its own behavior patterns regularly and quickly elicit a reaction from other vehicles. We observe that after an initial learning period Vehicle 11 will either produce these behavior patterns frequently or pointedly avoid them. It will use them as signals. It will also learn those signals that regularly precede certain behavior patterns of other vehicles. After a while Vehicle 11 will react to these premonitory signals just as it reacted, before the learning, to the behavior that regularly followed the signals.

But it would take prolonged observation to notice this particular aspect of learning in the vehicles. As a matter of fact, we might not have suspected it if we had not introduced a piece of our own philosophy into the construction of these vehicles. As our brain children become more efficient, we notice that the "law of uphill analysis and downhill synthesis" becomes more and more compelling. For the time being, take the message in this form: since you were not satisfied with the first meager showing of intelligence in our vehicles, we started adding a few more tricks, hoping that they would convince you a little more. The first trick we tried was the coding of the environment in those terms that yield a maximum of correlations and logical structure, in other words, in the most meaningful terms.