"Conclusions of Part II, Computer simulation and the Reverse Engineering Method" in R.Franck (Ed.) 2002, *The Explanatory Power of Models,* Dordrecht/Boston/London : Kluwer Academic Publishers (Methodos Series vol.1), 141-146

CONCLUSIONS OF PART II

Robert Franck

Reverse engineering

"The aim of an engineer studying a system by *reverse engineering*, de Callatay writes, is to find machines *functionally equivalent* to this system." A machine is functionally equivalent to another machine or to a natural system when it possesses the same properties[1] as the original, at a given level of description. This machine can differ as much in terms of the materials used as in terms of its machinery. De Callatay mentions the example of an artificial heart. "If an artificial heart is grafted to a patient, this heart is an engineering model of the natural heart. This heart uses different pumping principles. It does not use the energy molecules produced by the body. (...) But it still performs the heart's essential function." In this example two materially different structures perform identical functions. In a similar manner, it is unnecessary for a computer to resemble the human brain in order to mirror some of the properties of the brain. All the computer has to do is to perform some of the same functions as the brain.[2]

An engineer who has designed a machine which is functionally equivalent to another system has in this way modelled the functional structure of the other system. In manufacturing the machine he tests this functional architecture: does the machine actually possess the properties of the original? As he adjusts the machine and as it approximates more and more closely the properties of the system, the engineer discovers with greater and greater precision the functional architecture which must be carried out.

Again, in an analogous way, a model of functions within a socio-economic system, created through a computer simulation, can be adjusted for better and better fit of its results with actual properties of the real system. This was shown by de Callatay in Chapter 5. Thus, the functional design of the real system can be brought up to date as long as we know its properties. Reverse engineering makes it possible to reconstruct the functional architecture of a system without needing to know its material architecture.

---

1 We have defined properties as the specific outputs of a system, which vary with given specific inputs. In other words, properties are various specific effects that may be produced by the system in interaction with a definite environment.

2 The word 'function' is used here in the classic sense, as in physiology (*usus partium* following Galien) and later in sociology (Durkheim, Malinowski, Merton). This sense is also its everyday meaning: the function of education may be fulfilled by schooling, or by the family, or in other ways.

Denett's 'artifact hermeneutics'

Daniel Dennett has devoted many pages to the subject of reverse engineering, and he was probably the first to emphasise that reverse engineering is indeed currently practised throughout the field of biology. But a clarification is needed. The way in which Dennett (1996) conceives of reverse engineering is far distant from that which has been argued here. For Dennett, reverse engineering is "artifact hermeneutics". The 'reverse' engineer's task, he argues, is to interpret the intention of the manufacturer: "Why did GE make these wires so heavy?" Or, for the biologist, the task is to interpret the intention of Mother Nature: "we try to figure out what reason, if any, 'Mother Nature' - the process of evolution by natural selection itself - 'discerned' or 'discriminated' for doing things one way rather than another. (...)you can't do reverse engineering without asking what reasons there are for whatever it is you are studying. You have to ask 'why' questions."

Following the conception of reverse engineering described earlier, as opposed to that of Dennett, the engineer is freed from the need to guess the intentions of the manufacturer. What is sufficient is to study the properties of the product. Hermeneutics has no place here. True, reverse engineering involves an effort to interpret the properties of a system: why are they the way they are? But to speak of hermeneutics suggests that we can approach a machine, or nature, in the same way as we read a book, seeking to understand what the author is saying (whether the author be God, General Electric, Mother Nature or the process of evolution). But the 'reasons' which guide the evolutionary processes, for instance, are not comparable with the thoughts of an author. The processes of evolution, as theorised by Darwin, are guided by their functional architecture, i.e. the combination of variation and selection. Dennett's point of view leads us to forget that Darwin's success was based on observation of the properties of the processes of evolution. Note that Darwin's success was not based on observation of the processes themselves. Darwin applied the reverse engineering method.

The explanatory power of ANN's

"Most ANN models," Michel Verleysen has emphasised, "are far from the biological reality; modelling brain neurons and synapses, and developing computational tools able to perform efficiently in perceptive tasks are different businesses, despite a common inspiration. (...) This is clearly a specific assumption: our interest is to build something that works and which can be *used*, whatever is the way to reach the goal." We recognise here something which we learned just now about reverse engineering, illustrated by the artificial heart. This is why ANNs are described as 'blind', or as black boxes, and why such descriptions are accurate: the example of the Multi-Layer Perceptron from Chapter 6 shows this clearly. Some have concluded from this that ANNs are not explanatory. If this is the case, the same charge could be levelled against reverse engineering. Is the claim correct? De Callatay asserts strongly that it is not, as we have seen: "an engineer understands a real system when he can design a (virtual) machine that is functionally equivalent to this system." Let us try to settle the question.

Since ANNs do not mirror reality, it seems obvious that ANNs do not provide any insight into reality. However Verleysen points out that the interpretation of the parameters in a neural network is not impossible, and is in fact a major concern of ANN research. He cites examples: "many models of the human visual system are valid at the material level (they do resemble - in some way - the human system)." To resolve the difficulty, we must reiterate the distinction between the system's functional and material architecture. Is it the functional architecture of the system that is modelled by an ANN, or is it its material architecture? An engineer who is ignorant regarding the material structure of the system may yet discover, by way of reverse engineering, the functional structure of this system. The question of the explanatory power of ANNs may then be formulated in alternative terms. Such power does not depend on the capacity of the model to represent a system's material architecture correctly. It is the accuracy of representation of the functional architecture that is central. The question then becomes: if they correctly represent the functional architecture of a system, can ANNs explain, or contribute to an explanation of, its properties?

An example

As an aid to answering this question, we might consider the famous model of McCulloch and Pitts (1943), the ancestor of the ANNs. This model represents just one neuron. Yet what was represented? Not the soma, nor the axon, nor the dendrites, nor the gene nuclei, nor the membrane, nor the shape of the neuron, nor the way that the various parts of the neuron work together!

Figure 1 : Mc Culloch and Pitts' model

Starting from the observation of some main properties of the neuron, McCulloch and Pitts tried to represent the functional architecture - not the material architecture - without which these properties could not come about. Five functions were identified: the reception of stimuli (x1,x2...xN); the weighting of stimuli (synaptic coefficients w1,w2...wN); the calculation of the sum of weighted stimuli received (p); the fixing of a threshold of stimulation ( ) above which transmission would occur; and the stimulus exit (s). The model does not represent the combination of factors that fulfil these functions; it represents the functions themselves. And these functions are ranged in a specific order: the weighting of stimuli must precede the calculation of the weighted sum, and so on. Notice that such a model represents nothing observable, but is wholly conceptual. It represents nothing of the material structure interior to the neuronal cell by which we might explain the neuronal response to stimulation. Rather, it represents the functional architecture of the cell. Can knowledge of this functional architecture contribute to the explanation of the properties of a neuron? Yes, if it is true that without this functional architecture the observed properties of the neuron could not be produced. This was the ambition of McCulloch and Pitts. The functional architecture of the system that generates the properties of the neuron also provides, up to a point, an explanation of the material structure of the neuron in so far as this last structure can be described.

After all, the material structure serves to fulfil the functions specified. The explanation of the neuron's material structure by way of its functional architecture gives us information regarding the purposes which the different parts of the neuron, and the influence which they mutually exert, might serve. It is in this dual sense that McCulloch and Pitts' neuronal model is explanatory: this functional model contributes to the explanation of the system's properties, and of its material structure.

The foundation of reverse engineering

The example shows us that ANNs can contribute to the explanation of the properties of a system, and even to that of its material architecture, if they correctly represent the functional architecture of this system. But how do we know if a functional model correctly represents the functional architecture of a system? Different ANNs might lead to the same outputs. Among these, is there one which represents the true functional structure of a given system?

We affirmed that reverse engineering, by succeeding in reproducing the properties of a system, makes possible the discovery of the system's true functional structure. This rests on one key assumption: it is supposed that two systems with the same properties have the same functional architecture, whatever their material differences. In other words, we presume that the same properties require the same functional architecture, or that there can be only one functional architecture which is compatible with a given set of properties. This presupposition does not seem tenable, unless we subject it to a quite specific condition. This special condition is fulfilled in the McCulloch and Pitts model.

The functions of reception and weighting of stimuli, of summation, of threshold-fixing and of exit - plus their ordering - are the combination of functions *without which it would apparently be impossible* for the neuron to have the observed effects. The model of McCulloch and Pitts is grounded in this manner: it represents the functional architecture *necessary* to produce all the observed properties of the neuron. This is what renders valid the assertion that a (virtual) machine and a natural system or another machine which share similar properties must also share the same functional architecture: if this architecture is *necessary*, if it is *that without which* the properties would be impossible, then both systems could have by definition no other functional architecture. In cases where different systems have the same properties, i.e. produce the same outputs for the same inputs, they must share some functional configuration without which these effects could not be produced. When different functional configurations are compatible with the same set of observed properties, it's only because they have not been subjected to the criterion of necessity. But we could try restricting the model so that it only retains those features without which it would be impossible to generate the observed properties. McCulloch and Pitts' model is of that type. There are examples in social science of such models. The several chapters of Part III will furnish examples.

Is the functional architecture that we regard as necessary - for example McCulloch and Pitts' model - really necessary? We might be mistaken. Further observation of the properties of neurons, and further theoretical reflection, could lead to correction or to replacement of the initial model. But there can be only one true model of the functional architecture, such that everything which is necessary for the production of the observed effects is included in the model, and all that is not necessary is excluded from the model, even if it could contribute to the production of these effects.

The advantages of dissociating two types of models

The distinction made between a system's material and functional architecture suggests that we can profit by dissociating two types of model. Instead of representing, in the same model, both the network of material components of a system (factors, variables, geographical spaces etc.), and the functions which we attribute to them, it would be preferable to model the functional structure of the system separately. I already put forward this suggestion in the concluding part of Part I. The advantage appears evident: the model of the functions of the social mechanism can guide the statistical modelling of the social factors which generate the observed phenomena. We have already seen, here in Part II**,** new advantages coming from the separate modelling of the functional architecture and the material architecture of a social system. I would like briefly to underline these new advantages once more.

We have seen the immediate advantage, for the engineer, of this separation: if he does not know the material structure of the system, he can nevertheless, thanks to the modelling of its functions, reproduce and explain, at least up to a point, its properties. This is not only useful for engineers. For example, reproducing by other means the success of a road safety campaign, or efficiency in the medical service, is no longer a matter of luck when those in charge make use of reverse engineering, i.e. when they concern themselves with studying and reproducing the functional architecture through which, alone, the new process can produce the same results.

Another advantage from separating the modelling of functional and material structure is that functional models, when independent of a particular system's material structure, are transferable to other systems. This offers new perspectives with regard to the possibility of generalizing research findings in social science. The singularity of social situations seems to forbid any claim to general applicability of advanced explanations. But the general applicability of an explanatory functional structure appears practicable. Still, the functional model's capacity to be generalized is not absolute. It can be transfered to different empirical mechanisms only insofar as these mechanisms possess the same properties.

When we retain in a functional model only that which is necessary for the explanation of observed properties, the impossibility of conceiving of another general functional model can promote consensus on a general program of research among a larger number of researchers. Naturally it is possible that the appropriate model will not agreed on immediately, and that corrections will be in order. It is also possible that more than one functional model will seem viable at a given moment, in the judgment of others. However, these models will never be numerous, since the models proposed must be judged necessary in order to explain the phenomenon under study, and not simply plausible candidates.

What is theory?

At the end of Part I, it appeared that the possible contribution of theory to statistical modelling consists in the modelling of the functions of the social mechanism - such was our first approximation. Now at the end of Part II, we can identify the nature of a theory more precisely. We have learned that theory - taken as the modelling of a system's functions - can be built up on an empirical base, beginning with systematic observations about the properties of the social system under examination - micro or macro. Thus does theory cease to be taken for a creation of imagination which must be corroborated. Certainly, much inventiveness may be generated in the process of the construction of a functional model. But this labor is rooted in observation of the properties of the given system[3]. If we succeed in linking theoretical consideration to observation with regard to properties of a system, we can measurably reduce the gap between theoretical and empirical research.

We have also learned that theories can be necessary, or have the character of necessity. For this to be concluded, a theory must be judged to contain nothing which is not necessary in order to make conceivable the presence of those properties of the system which have been observed. When a theory contains *that without which* those properties would be impossible, it becomes the foundation of the social system under study. We can equally say that it is a principle. The chapters of Part III will allow us to deepen our understanding of the character of necessity which social theories may attain.

REFERENCES

DENETT D., 1996, *Darwin's Dangerous Idea : evolution and the meanings of life*, London, Allen Lane The Penguin Press

McCULLOCH W. S., PITTS W., 1943, "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biophysics* 5 : 115-133. Reproduced in : ANDERSON J., ROSENFELD E.,1989, *Neurocomputing, Foundations of Research*, Cambridge (Mass.), The MIT Press, Fourth Printing.

3 In Part IV it will be shown that work which consist in starting from the properties of a system in order to construct the required conceptual model corresponds to an operation of *induction* such as this was conceived in the 17th Century.