M. van den Dries

Artificially intelligent archaeologists: fundamentals, facts and fictions

Artificial computing techniques can be very useful tools for archaeological research and education, but they are not much being applied yet. In this paper the abilities of two of these techniques, i.e. expert systems and neural networks, are described and compared. With both techniques an application has been built for the analysis of use-wear traces on flint implements.

1. Introduction

Computers started to be used by archaeologists in the early 1960s. At first they were only used as facilities for the storage and statistical analysis of large data sets (*e.g.* Kendall 1963). In the 1970s the 'New Archaeology' clearly affected the way in which mathematics and computers were used. Gradually their use became more differentiated. They evolved from data-describing aids to process-modelling, hypothesis-generation and data-explaining aids (Doran/Hodson 1975). With this shift of attention came a need for more advanced problem-solving techniques that could utilize specialist knowledge, *i.e.* artificial intelligence techniques. Doran even expected knowledge utilization to become fundamental in archaeological data analysis (Doran 1974, 70). However, it was not until the 1980s that artificial intelligence technology reached a sufficiently high level to enable the first archaeologists to build their own expert system applications. This development was part of a process in which computing techniques were integrated in all kinds of archaeological research — a process which was made possible by the introduction of the personal computer and the subsequent explosive growth of the amount of software available. The archaeological world swiftly adopted many of the new computing techniques, for instance geographical information systems, but the same does not hold for artificial intelligence techniques. Despite the facts that several successful expert systems have been developed (*cf.* Bishop/Thomas 1984; Brough/Parfitt 1984; Francfort 1991; Grace 1989; Lagrange/Vitali 1992; Patel/Stutt 1989) and that several researchers have pointed out their potential value for archaeological research (*cf.* Baker 1987; Doran 1987, 1988, 1990; Voorrips 1990; Wilcock 1986), they have been neither readily accepted nor developed and applied on a large scale. Their usefulness is still being discussed.

This lack of popularity of expert systems in archaeology is a rather strange phenomenon. Expert systems offer means for modelling and formalising subjective and heuristic expert knowledge and for making that knowledge accessible to and usable for non-experts. As most of our knowledge is subjective, there are certainly abundant potential applications in archaeology. The lack of popularity in archaeology is also strange in view of the great amount of attention that is being paid to this technology in many other scientific research disciplines, and also commercially; numerous applications are operational in all kinds of fields.

According to Gibson this lack of popularity is 'due perhaps to the limited potential of expert systems in host disciplines' (1992, 263). But, in my opinion, their lack of popularity is due mainly to their supposed limited potential rather than to their actual limited potential. When archaeologists are asked about their opinion on these techniques, they are very interested. Ignorance and threshold fear, however, keep them from exploring their abilities. They either do not think of these techniques as potential aids for their research because they simply don't know the abilities of these techniques, or they hesitate to use them because of some fictions they have heard or because of prejudice. In other words, as long as they are not repeatedly confronted with useful applications and good results they will not start using or developing them.

The aim of this paper therefore is to show how these techniques can be useful for, for instance, educational and research purposes. Only expert systems and neural networks will be discussed below, because they are the two artificial intelligence techniques that simulate human reasoning for the purpose of problem solving. First, some fundamentals of both techniques will be given to demonstrate their specific abilities. Secondly, the functionalities of these techniques will be compared by means of two applications that have been developed for the analysis of use-wear traces on flint artefacts. The results of a test-case in which both applications were involved will be presented. Finally, an attempt will be made to separate some facts from fictions regarding these techniques and artificial intelligence in general.

2. Artificial intelligence applications

Expert systems and neural networks were developed by the research discipline called Artificial Intelligence. This discipline is concerned with the development of computer techniques that enable the simulation of human intelligence. The artificial intelligence research has yielded several techniques, each specialised in imitating aspects of human behaviour, speech or reasoning. Examples of these are robotics, (visual) pattern recognition, natural language processing, speech recognition, expert systems and neural networks.

Both expert systems and neural networks are computer programs that simulate human reasoning processes. If they are provided with human knowledge and problem-solving methods, they can solve highly specialised problems or execute complex reasoning tasks.[1] The aim of these techniques is to offer an opportunity to organise human expert knowledge into a form in which it can be used by non-experts.

Apart from their background and aim, expert systems and neural networks have nothing in common. They have different architectures and use specific knowledge storing and processing methods. Furthermore, they work with different data formats. Expert systems can process non-numerical (symbolic) knowledge, whereas neural networks are based on numerical data. The most important difference, however, concerns the type of knowledge they can handle. Whereas expert systems require explicit knowledge (*e.g.* decision rules), neural networks can work with examples which contain knowledge implicitly. Those differences imply that the two techniques should be used for different purposes. Expert systems are successful in simulating heuristic methods and techniques. Neural networks, on the other hand, are capable of detecting (hidden) relationships between the properties that describe patterns within large and complex data sets. They can therefore be employed in analyzing problems of which the relations between the variables are unknown.

2.1 Expert system fundamentals

An expert system consists of three components, *i.e.* a knowledge base, an inference mechanism and a user interface (fig. 1). In a way, a knowledge base can be compared with a data base in that both are storage facilities. A knowledge base, however, consists of knowledge instead of raw data. The inference mechanism can be seen as the central nervous system. It applies the knowledge and controls the reasoning process. The latter means that it makes sure that the appropriate knowledge is applied at the appropriate moment. The communication between the system and its users is handled by the user interface. It serves as an intermediate between the two sides by receiving and translating their respective messages.

For an expert system to be able to simulate expert reasoning it must be provided with specific expert knowledge. Usually, this is heuristic knowledge, which has been extracted from the expert himself. Such knowledge is based on formal facts and theories (gained through education) and on subjective rules-of-thumb and intuition (gained through experience). Before the extracted knowledge can be stored in a computer, it must be modelled, formalised and translated. The extraction, formalising and modelling of knowledge is called the acquisition phase. It is the most delicate phase of the expert system development process. Due to the subjective nature of the knowledge involved, it is for instance difficult to retrieve the expert's underlying reasoning processes. Moreover, it may be hard to have this kind of knowledge represented by means of explicit representation methods such as (IF-THEN) decision rules.[2] Eventually, the acquisition yields a formal knowledge model that consists of all the knowledge and procedures necessary for solving a specific problem. The knowledge model can be translated by means of a computer language and subsequently implemented into a computer. Nowadays, the actual development of the expert system can be facilitated by using an expert system shell. A shell is a program which provides all the facilities of an operational expert system but leaves the knowledge base empty.

An expert system uses its knowledge either to interpret information or to validate hypotheses. Such systems are called data-oriented (forward-reasoning) and goal-oriented (backward-reasoning), respectively. A data-oriented system 'reasons' in a forward direction, which means that it has no predefined goal. Instead, the reasoning process reacts to information that the system receives from the external world.[3] The system will try to interpret this information by searching in its knowledge base for conclusions that can be drawn from it. For instance, if a user indicates that his data are 'red, round, and small', and the system has a rule saying 'IF red and round and small THEN it is a cherry', the system will conclude that the data represent 'a cherry'.

A goal-oriented system does the opposite, it 'reasons' in a backward direction in order to confirm a predefined goal. This means that such a system tries to retrieve the information that is required to confirm that goal. For instance, if a user indicates that his hypothesis is that 'it is a cherry', and the system has a rule saying 'IF it is a cherry THEN (it must be) red and round and small', the system will ask whether the characteristics 'red and round and small' are indeed present. Such a goal-oriented system cannot only retrieve the required information by questioning its user but also by consulting its knowledge base or another external data source, such as a data base.

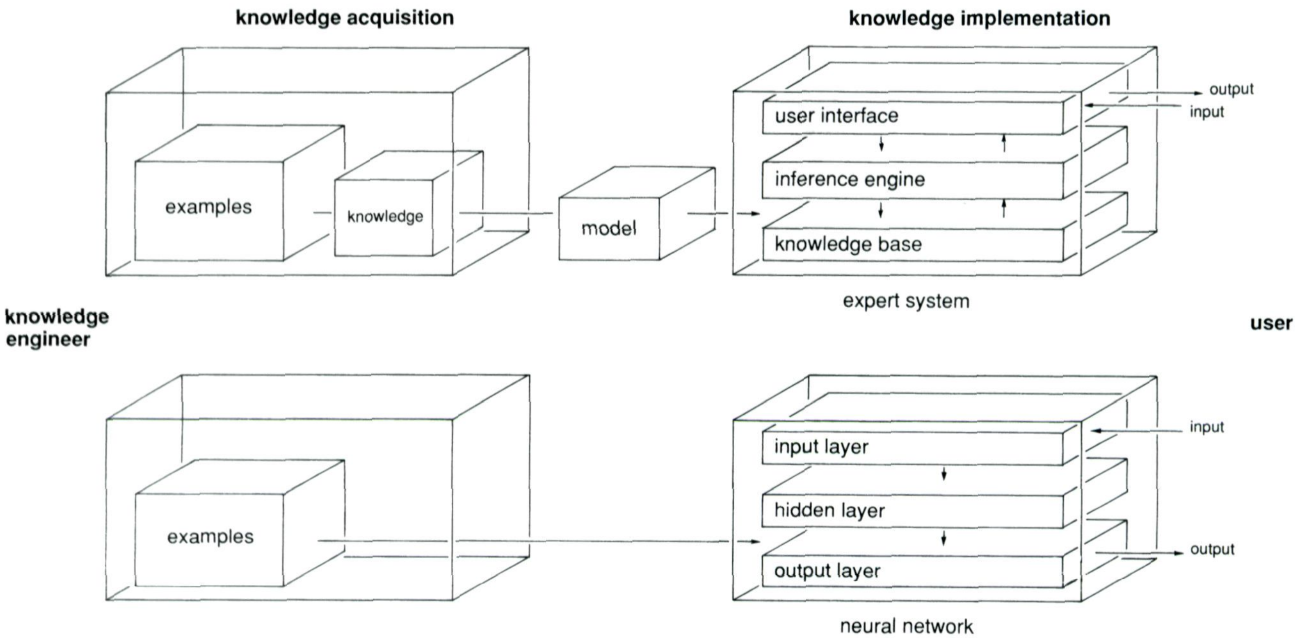

Figure 1. An expert system versus a neural network regarding their architecture and the composition of their knowledge.

As data-oriented systems can be used to interpret data or to react to (changes in) incoming information, they are most suitable for applications with analytical and educational purposes, especially for those that require an immediate reaction of a 'master'. A goal-oriented system can be applied to situations in which a user wants his hypothesis verified or wants to obtain a second opinion due to uncertainties.

Used in either of these two ways, an expert system can only be applied to tasks that can be clearly defined and that are of a restricted extent and complexity. As far as the complexity is concerned, a directive may be that a human expert must be able to solve a particular task or problem within a couple of hours. Of even greater importance is that the required knowledge can be represented by means of explicit methods. However, some expert knowledge can hardly be formalised because of its subjective character; experts may have difficulties in describing their knowledge explicitly and in explaining the underlying reasoning processes which they apply (Kidd 1987, 3). Moreover, some problems are so complex that it is not possible to determine the relations between their variables. This means that such tasks cannot be simulated by means of the expert system technique. For those cases a neural network application may be a suitable alternative approach.

2.2 Neural network fundamentals

The architecture of a neural network is deduced from the biological structure of the human brain. The brain is a very complex organ that is made up of ten to one-hundred billion cells, called neurons. Neurons are special cells that are capable of receiving, storing and sending information. Each neuron is connected to approximately ten thousand other neurons and together they form a complex network.

Via the connections (dendrites) they send electrical and chemical signals through which they communicate with each other. Since these signals can be transferred simultaneously, thousands of impulses can pass the neurons per second (Carling 1992). That way the network structure enables a massive neuron activity. This implies that the brain can process enormous amounts of information and, thus, adequately respond to the situations it is confronted with.

An artificial neural network is a computer program that tries to simulate the principle of a biological neural network by means of a mathematical model. Like the human brain, an artificial network consists of neurons, *i.e.* small processing elements that can receive, process and send signals. These neurons are arranged in at least three layers: an input layer, one (or several) hidden layer(s), and an output layer. The neurons of the input layer represent the variables describing a problem. The output neurons represent the solution that is associated with it. The neurons of the hidden layer act as an intermediate between the input neurons and the output neurons. They are invisible to both the application developer and the user for they have no direct connection to the outside world.

As with the human brain, the neurons of a neural network are connected to one another, although not physically. They can pass information to one another through programmed instructions. Each neuron in one layer is connected to those of another layer (fig. 2). A network's knowledge is distributed among these connections. It is not stored within the neurons.

The principle of a neural network is that it can be instructed with examples of a problem to enable it to solve similar problems. It searches for relationships between the variables of a reference data set in order to apply these to similar data. The network translates these relationships as connections between the neurons. Each connection has a specific weight, a numeric value. The collection of weights is stored as a data matrix.

The configuration of the connections is based on the results of calculations and has been established through a process of repetition of examples. Creating the right weights is a complex and time-consuming process, which can take a long period of training. Training means that a network is provided with a large set of examples which consist of input and output patterns, which the network tries to simulate in order to construct the connections and their weights between the various input and output neurons. For instance, the input for a network that has to recognise fruit would be 'red, round, small' and the output 'cherry' or 'yellow, round, large' and 'melon'. Both the connections between the input and output and the weights are entirely programmed by the network software. They are constructed by means of predefined mathematical functions and the network developer has hardly any influence on this process. The network starts its training by giving an arbitrary output to a specific input of an example. It compares this output with the expected output of the example. The arbitrary outputs are of course predominantly wrong. However, the network evaluates these mistakes and subsequently adjusts the connections or the weights of the connections until the network is able to generate all the outputs correctly. In other words, it 'learns' by experience.

Once this training is finished, the application building is finished. The network can then be employed for the interpretation of 'new' situations or problems that resemble those it has learned. With a neural network this means that a user only has to select the input variables that represent the properties of the problem. With reference to the prior example, the neurons 'yellow', 'round', and 'small' could for instance be selected. These activated neurons subsequently send a signal to the hidden layer. As each hidden neuron is connected to each input neuron, it will always receive several signals. These incoming signals are of various weights, because they come from various connections. The hidden neuron calculates its activation strength by summing up the incoming signals. Only if a certain activation level is reached, does a hidden neuron pass a signal on to the neurons of the output layer. Again, the strength of these signals and the weight of the connections are responsible for the degree in which the neurons of the output layer are activated. Finally, the combination of the activated output neurons represents the network's interpretation of the information that was presented by the user (Lawrence 1991).

From a developer's point of view, the development of a neural network is far less time-consuming than that of an expert system. Although it probably seems quite complicated, building a neural network is fairly simple. The connections and the weights matrix are programmed by the network software on the basis of predefined mathematical functions.[4] The knowledge acquisition phase consists only of selecting the input and output variables and collecting sufficient examples (fig. 1). As for the implementation, the developer's task is twofold. First, he makes the network structure, *i.e.* he defines the input and output neurons and the number of hidden layers. Then he provides this network with a large set of examples. So, unlike with an expert system, the knowledge does not have to be made explicit before it can be implemented into a network. The developer has to build decision rules (IF-THEN) for an expert system, whereas the neural network deduces the relations from the examples and builds the decision 'rules' itself. The only requirement is that the examples must be representative descriptions of the various situations (and the associated solutions) that have been experienced within the problem domain; they must cover the range of variability exhibited by the real world.

3. Two applications[5]

In 1990, a project was launched at the Institute for Prehistory of Leiden University to develop an expert system application for the analysis of use-wear traces on flint artefacts. This application was intended to support both students and experienced use-wear analysts in the analysis of use-wear traces and in the evaluation of interpretations, respectively.

The reason for launching this project was the desire to study the possibility of formalising use-wear analysis for the purpose of computer-assisted instruction. Since the use-wear expert at Leiden University[6] spends much time on the training of students, it was decided to develop an expert system that would be able to provide support in this task;

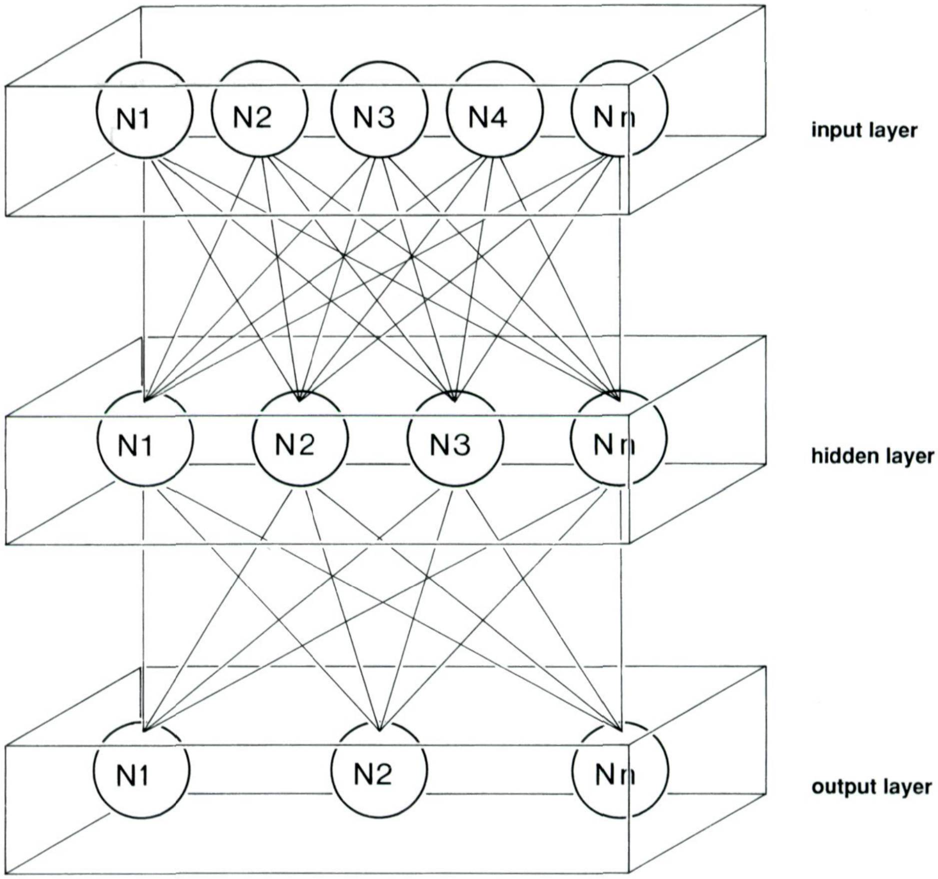

Figure 2. Neuron and layer structure of a neural network. Each neuron (N) in one layer is connected to each neuron of another layer.

if parts of the expert's task were to be taken over by an educational system, the expert would be able to spend more time on research. Moreover, unlike a human expert such an artificial expert is available at any moment and is easily duplicated. That means that both the research and the educational capacity of the human expert can be increased. Using an expert system can be advantageous for students as well. They can control the pace and the direction of their learning and they have the opportunity to work with advanced computing technology.

Another advantage of developing an expert system was that it would offer an opportunity to formalise and standardise the method of analysis. Formalising a method improves its scientific acceptance while standardised procedures yield less subjective results. Moreover, once the knowledge is in a formalised format, it can be evaluated in order to trace deficiencies.

The project has resulted in a system called WAVES.[7] At present this system is partly operational. The working parts have been tested and one of these tests will be discussed in the next section. While this application was being developed, the neural network technology was introduced into the archaeological world. Neural networks were launched as 'a superior alternative' (Gibson 1992, 263) to expert systems, whose major functional disabilities they overcame. In order to verify these statements and to make a comparison with the achievements of the expert system, it was decided to develop such a system for use-wear analysis as well. A neural network application[8] could be realized with relatively little effort by using the knowledge source used for the expert system. Moreover, if both systems were to be based on the same knowledge source, it would be possible to compare their analyzing qualities.

3.1 The expert system application

As WAVES is meant to be used by two different groups of users, it consists of two independent parts. One part executes the analysis method as a step-by-step process. In this way students can be trained to execute the required procedures. While supervising the process of analysis, the program also tries to arrive at an interpretation. This means that it reacts to the information given by the user. It is data-oriented.

The other part evaluates interpretations in order to support the experienced analyst who is already familiar with the procedures but not very confident of his or her interpretations. This part is goal-oriented; it tries to confirm the interpretation by questioning the user. The following description and test-case refer to the analyzing part only.

As the analyzing part of the WAVES application not only steers the direction of the process of analysis, but also interprets the observed wear traces, it needs detailed information on the characteristics of the wear. This is obtained via a question-and-answer game. It means that the system will ask the user for a description of the analyzed implement and of the observed wear traces. In order to make sure that the system receives the appropriate information, the answers to the questions are of multiple-choice format. This means that for each answer (variable) a lists of possible answers (values) is given (fig. 3). The user chooses a value that corresponds best to the traces observed. Whenever the user is uncertain about his answer, he can ask for background information on the meaning of certain questions or of variables and values. This information consists of descriptions, photos or schematic drawings that explain the differences between the values.

The system's method of analysis resembles that of the expert. It starts by verifying whether an implement is analyzable and whether the observed wear traces can be ascribed to use.[9] This is followed by the actual analysis of the use-wear traces. The analysis is divided into two parts: one for the interpretation of use retouch and the other for polish, edge rounding and striations. These parts work independently of one another and can be employed separately. The reason for this division is that the various wear categories are either not always simultaneously present or not equally diagnostic. In such cases it should be possible to arrive at an interpretation that is based on either one of the categories. An advantageous side-effect of the division is that two independent interpretations can be obtained if all wear categories are present. The system's final questions relate to the morphological aspects of the implement. This information can be used to verify whether the result of the analysis is in accordance with the implement's morphology.

On the basis of the received information, the system subsequently attempts to identify the nature of the materials and the motions that may have caused the observed traces. As different contact materials may sometimes have caused similar wear attributes and as similar materials may have resulted in different wear patterns (*e.g.* Unrath *et al.* 1986), it is often difficult to identify the exact contact material and motion. Moreover, a tool may have been used on several materials. Hence, it would be wrong to focus on the identification of one contact material only. Therefore, the result of the system's interpretation consists of a list of all the materials and motions that may have caused the wear and the probabilities of those materials and motions actually being responsible for the observed traces (fig. 4). These probabilities are expressed by means of scores, which are the results of the calculation of the indications suggesting a specific contact material.[10] They also give an indication of the value of the interpretation; traces that are not diagnostic, such as a generic weak polish, will not yield interpretations with high scores. If, on the other hand, the wear characteristics are indicative of only one contact material then this material will have a higher score than the others. Furthermore, the system's interpretation is accompanied by pictures showing the wear traces that the system associates with the result of the analysis. With the aid of these pictures, the user can verify whether the system's interpretation corresponds to his observations.

The knowledge on which WAVES is based was derived from the results of an experimental programme consisting of 301 experiments with replicated flint artefacts (Van Gijn 1989, 168-174). It was believed that these results could serve as a reference collection from which knowledge rules could be deduced. A detailed analysis of these data revealed what combinations of the wear attributes are diagnostic of specific contact materials and motions (Van den Dries/Van Gijn in press).

The main reason for using a reference collection as the basic knowledge source was that it is very difficult to build a knowledge model with knowledge extracted from human experts directly and from research reports (Van den Dries 1994). Such expert knowledge often covers predominantly the category of the diagnostic wear traces, whereas the category of the exceptions and uncertainties is underrepresented. A knowledge model used for educational purposes should cover this variability. It was therefore believed that a reference collection would include parts of both categories and that expert knowledge could be used to supplement the knowledge derived from such a collection.

3.2 THE NEURAL NETWORK APPLICATION

The reference collection that formed the basis for the knowledge of the expert system was also used to train a neural network application. The difference between WAVES and the network is that the knowledge was translated into a different format: into decision rules for the expert system and into presence/absence scores for the network. Furthermore, the decision rules were deduced from the raw data, whereas the network was trained with the aid of unaltered data.

Up till now, this application has only been trained to interpret polish. As only 161 examples of the reference collection showed traces of polish, this was a relatively

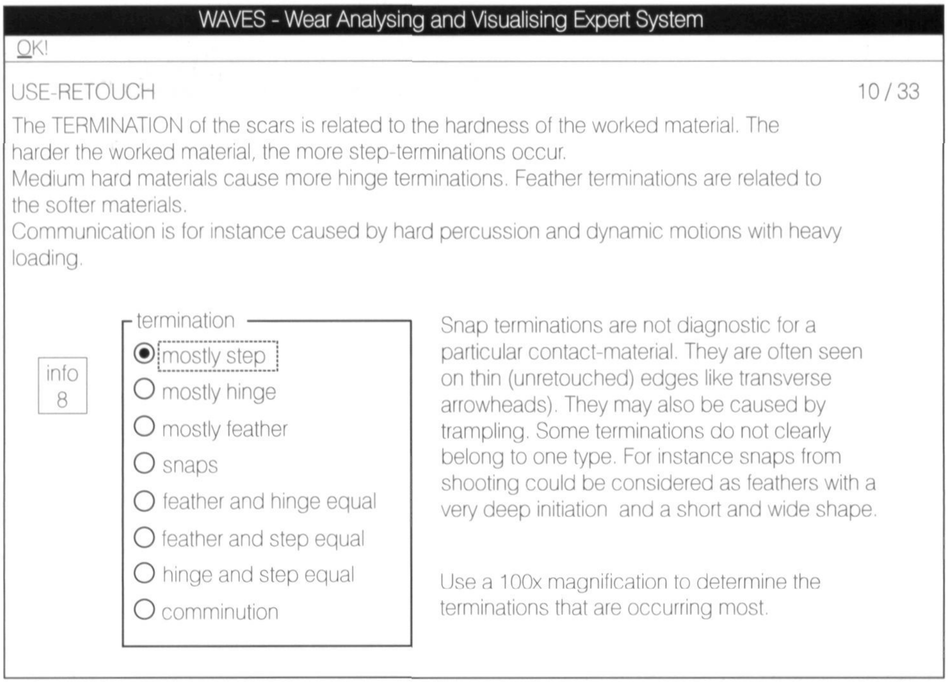

Figure 3. Example of a screen of the expert system application. It enables the user to choose the wear-attribute that corresponds best with the observed traces.

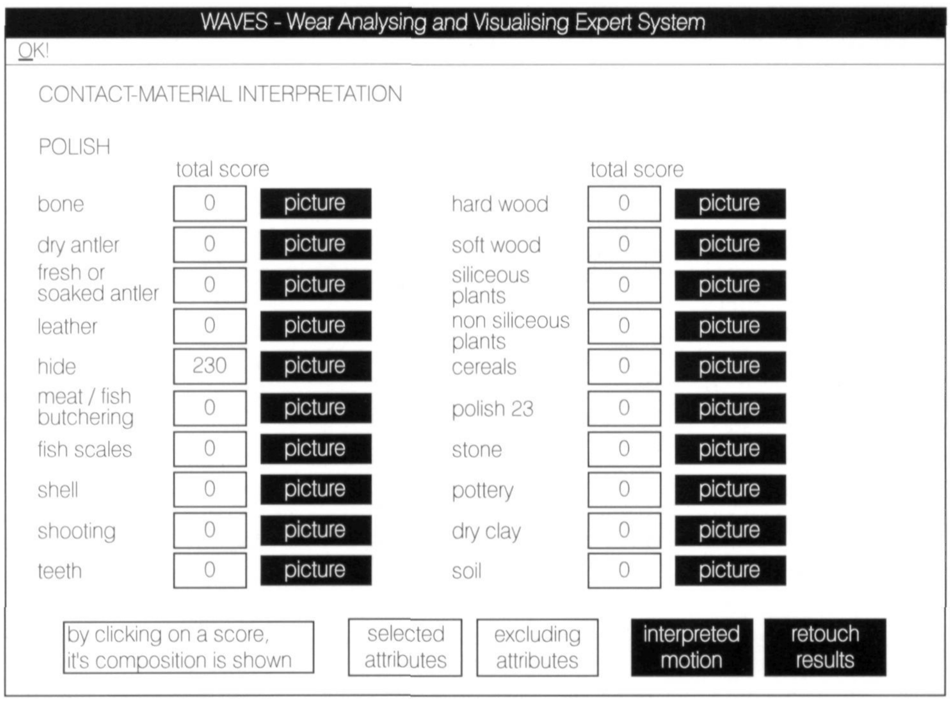

Figure 4. Screen of the expert system application showing the results of the polish analysis. By clicking on a "picture"-button photographs of the associated use-wear are being displayed. The other buttons give access to the results of the retouch analysis and of the motion analysis.

small training set. It took the network approximately one hour (650 runs through all examples) to 'learn' the input and output patterns of the examples.

The network's input layer consists of 31 neurons and the output layer of 15 neurons (fig. 5). The input neurons represent the attributes of the five variables that describe the characteristics of the polish. The neurons of the output layer represent the contact materials that may have caused the polish. As with the output of the expert system, several of the output neurons may be active at the same time if the traces analyzed indicate several contact materials. However, the degree of activity of these neurons may differ. This

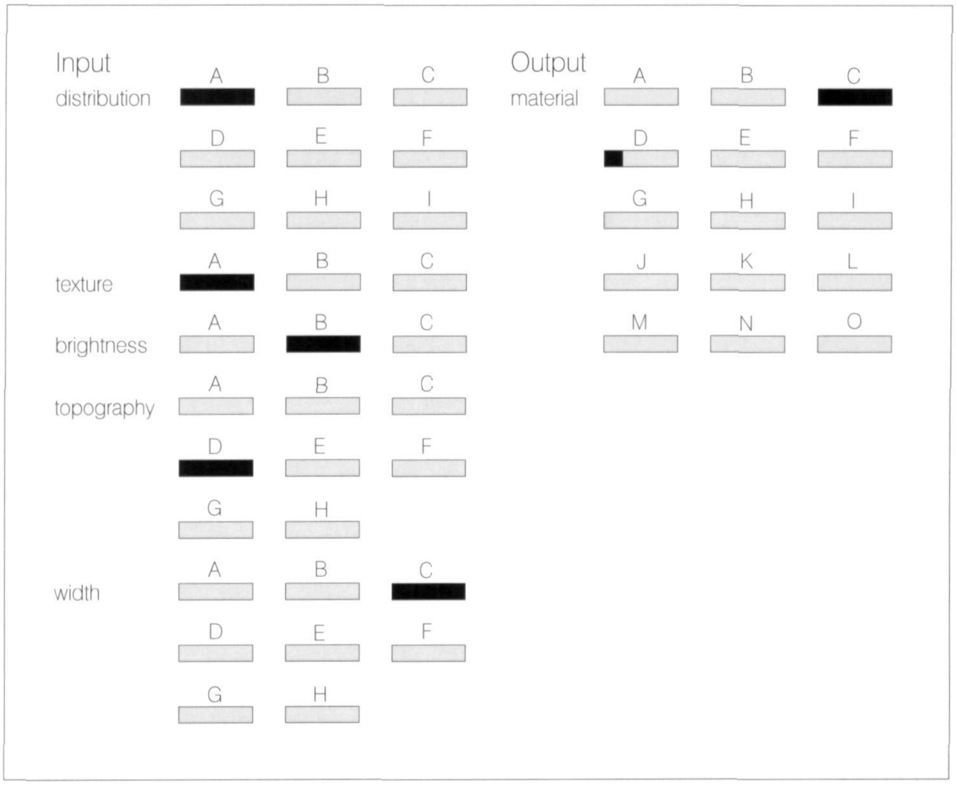

Figure 5. The only screen of the neural network application. The left hand side shows the input neurons (the polish distribution, texture, brightness, topography, and width), the other side the output neurons (the associated contact-material). The user chooses the five input neurons that correspond to the observed wear-traces and the system shows the contact-materials *that may have caused that specific* wear pattern.

degree of activity is an indication of the certainty of the interpretation.

Whenever a user consults the network, he only has to select five of the input neurons which correspond to the characteristics of the use wear traces observed. The network compares this information with its reference matrix and calculates the degree of activity of the output neurons. Because a network consists of compiled data matrices only, a processing session takes less than a second.

Since the application was intended for studying the potentials of a neural network, it was not provided with a user-friendly interface; it is not suitable for computer-assisted instruction. However, it can certainly be used for its achievements comparing to those of an expert system. It can also be a useful tool for a professional use-wear analyst who is interested in studying the diagnostic value of specific attributes. It may be interesting to observe the consequences for the output of manipulation of the input. If the preliminary results are promising, this application may be trained to interpret other use-wear categories as well.

4. A test-case

The expert system and neural network application described above have been applied to a test-case to enable comparison of their achievements. This comparison referred to the interpretation of polishes only, because the network had been trained for this purpose only.

The test-case consisted of the analysis of 16 replicated flint artefacts used for experimental purposes and of 10 archaeological artefacts from the Dutch Linearbandkeramik site of Elsloo, Limburg, the Netherlands. The traces on these implements were entirely new to both systems as none of them had been used to compose their knowledge.

For the purpose of studying the reaction of both systems, the test set was composed of implements displaying polishes that were difficult to interpret. For instance, three experimental tools (346, 378, 385) were included that had been used on materials of which it was known beforehand that neither system would be able to identify them. Many of the other experimental tools were selected because they displayed slightly different polishes although they had been in contact with similar materials and because they showed less diagnostic wear patterns.

Because the contact materials of the tools that had been used in the experiments were known, the interpretations for these tools could be evaluated as a 'blind test'. The interpretations for the prehistoric polishes were more difficult to evaluate because the worked material was, of course, unknown. Therefore, these results were compared with those given by a professional human use-wear analyst.[11] The assumption was that in the case of dissimilar interpretations the human expert would be right.

Table 1. The actually worked materials versus the interpretation of the expert system and the neural network.
* Both systems have not been provided with knowledge about these materials.

tool nr.	worked material	expert system interpretation	neural network interpretation
344	soaked antler	-	dry antler/ fresh bone
345	medium hard wood*	hard wood/ soft wood	soft wood
346	shell*	-	soft plants
350	soft wood	soft wood	soft wood
351	soaked antler	-	hard wood
352	soft wood	soft wood	dry hide
360	soft wood	-	fresh bone
363	soft wood	-	fresh bone
367	fresh hide	fresh hide	fresh hide
370	fresh hide	fresh hide	fresh hide
371	fresh hide	fresh hide	fresh hide
378	hide with ochre*	-	soft wood/ dry antler
383	soft wood	soft wood	soft wood
385	dry clay*	soaked antler	soaked antler/ soft wood
386	fresh hide	fresh hide	fresh hide
388	dry bone	butchering	fresh bone/ dry antler

The test procedure consisted of two steps. First, the characteristics of the wear traces were described by an experienced analyst. Then these descriptions were presented to the two systems. The reason for this was that the test was intended to validate the knowledge. An experienced analyst was used to exclude the possibility of a system's poor achievements being ascribed to a user's lack of experience — something which otherwise may very well have happened because with both systems the actions of the user were not yet perfectly controlled.

4.1 The expert system's achievements

The results obtained for the experimentally produced polishes (tab. 1)[12] show that the expert system could not identify the traces of 6 tools (344, 346, 351, 360, 363, 378). This is a rather large part of the test set. There are two possible explanations for this. First of all, the traces analyzed may have differed from the traces on which the system has knowledge. This is due to the fact that this knowledge was derived from experimentally produced traces. An experimental programme cannot cover the entire range of possible traces. Experience has shown that some traces occurring abundantly on archaeological tools cannot be replicated on experimental tools. An example of such traces is what has been termed polish '23' (Van Gijn 1989, 85). This type of polish (bright, plant-like on one side, hide-like on the other) has been observed by several analysts (Cahen *et al.* 1986; Van Gijn 1989; Juel Jensen 1989; Keeley 1977), but its origin has not yet been experimentally determined. Such problems reveal one of the limitations of expert systems. If a situation or problem differs too much from those from which the knowledge was derived, a system may be unable to deal with it. For this application it is therefore very important that the knowledge base is supplemented with expert knowledge. Only the human expert has knowledge about the variability of the traces exhibited by the archaeological record.

The second possible explanation has to do with the subjective nature of the variables used to describe the wear traces. Most of the descriptions are based on relative 'measurements'. It is, for instance, difficult to decide whether a polish looks 'bright' or 'very bright'. This implies that the descriptions of the wear characteristics given by the analyst need not necessarily match those given by the expert, on which the system is based. This may cause discrepancies between the descriptions, yielding information the system cannot interpret correctly. Presumably, this problem will be eliminated when the system will be expanded with more photographs and schematic images giving examples of the attributes. As pictures give a much better impression of what is meant, they will facilitate the selection of attributes.

In one case (tool 385) the system gave an incorrect interpretation. However, this implement had been used for an entirely new experiment (carving dried clay), of which the system had no knowledge. The fact that the system did come up with an interpretation means that, according to the system, the observed traces bore some resemblance to those produced in working soaked antler. For a use-wear analyst this may be a strange misinterpretation. It can however be explained by the fact that the observed traces were non-diagnostic and happened to look like other non-diagnostic traces (produced in cutting soaked antler) which were included in the knowledge base.

This example of a misinterpretation illustrates the problem of identifying non-diagnostic wear patterns. Even though some similar problems may be avoided by expanding the application with expert knowledge and by enlarging the experimental reference collection, it is likely that no system will ever have sufficient knowledge to exclude all such misinterpretations. Non-diagnostic wear and especially generic weak polish is very difficult to interpret, even for the best professional analysts.

In two other cases (tools 345, 388), the system's interpretation was acceptable because it approached the right answer sufficiently. In some cases this may be justified because different activities may cause similar traces.[13]

As already mentioned before, the results of the analysis of the prehistoric polishes (tab. 2) are less easy to validate. Although the expert system could not analyze all the artefacts, the results obtained for those that it could are in accordance with those given by the human analyst. This is promising and it is encouraging that the system can already do this while its knowledge is still based on the results of experiments only.

4.2 The neural network's achievements

A major difference between an expert system and a neural network is that the latter will always generate an answer, even if it is a very unsure one.[14] This explains why the network made more mistakes in interpreting the experimentally obtained polishes (tab. 1). Most of these mistakes concerned precisely those tools (344, 346, 351, 360, 363, 378) which the expert system could not identify either. The network tried anyhow and failed. It searched for the material that came closest. Unfortunately, in the case of a use-wear analysis the resultant answer is often misleading rather than helpful. But in some cases it may give a correct indication of the hardness category of the contact material. The problem, however, is that you never know when the answers are reliable; misinterpretations that are not due to a lack of knowledge cannot be explained. This is because the reasoning process of neural networks is invisible.

The network correctly interpreted the traces of six tools (350, 367, 370, 371, 383, 386). The interpretation of tool 345 was accepted as being correct, because the system has no output neuron for medium hard wood, only for hard wood and for soft wood. In two cases (tools 344, 388) the network's interpretation was almost correct. It turned out that the network had some difficulties distinguishing between materials showing comparable traces, like hard animal materials such as bone and antler. However, this is not surprising. Professional analysts may also have difficulties in such cases.

It is peculiar that, like the expert system, the network interpreted the implement which had been used for carving dried clay (385) as used on soaked antler. This means that the observed traces must indeed have been comparable with those produced in working soaked antler.

With no fewer than eight of the archaeological artefacts (3b, 5, 6, 10, 19, 20, 31, 34), the network's interpretation was similar to that of the human expert (tab. 2). In three of these cases (tools 6, 10, 31) the network's interpretation corresponded to that of the human analyst, whereas that of the expert system did not. The traces of only two implements were misinterpreted (tools 1, 3a).

Table 2. Interpretation of polish on 10 Linearbandkeramik artefacts, given by a human analyst, the expert system and the neural network.

tool nr.	analyst	expert system	neural network
1	dry hide	-	fresh hide
3a	dry hide	-	fresh hide
3b	bone	butchering	butchering
5	hide ?	fresh hide	fresh hide
6	bone	-	butchering
10	fresh hide	-	fresh hide
19	wood	hard wood/ soft wood	hard wood/ soft wood
20	fresh hide	fresh hide	fresh hide
31	hide	-	fresh hide
34	antler	soaked antler	soaked antler

4.3 Conclusion

From a comparison of the achievements (tab. 3) it can be concluded that as far as the experimental tools are concerned, the expert system performed slightly better than the neural network. With respect to the interpretation of the archaeological implements, however, the network yielded a better result. This difference may be due to the composition of the test set. The selection of the replicated tools displayed relatively more wear patterns that were not very diagnostic, whereas the archaeological tools contained more diagnostic patterns.[15] If provided with the appropriate knowledge, expert systems may be better at interpreting

Table 3. Final comparison of the test results.
* In this case true and false means that the answer is respectively equal and unequal to the answer of the human analyst.

interpretation of:	expert system	neural network
experimental replica's (N=16)		
true	9	7
false	1	9
none	6	-
archaeological artefacts (N=10)*		
true	5	8
false	0	2
none	5	-
total (N=26)		
true	14	15
false	1	11
none	11	-

exceptions than neural networks. When interpreting data, the latter focus on recognising similarities with the examples that they have learned. They try to relate new data — and thus also exceptions — to their generalised knowledge. Therefore, they can only interpret exceptions correctly if they have been provided with enough 'learn examples'. Unfortunately, the difficulty with exceptions is that there are only few examples. However, when it comes to real exceptions that have never before occurred, the expert system will not be able to give an interpretation. It will simply lack the appropriate knowledge. A neural network, on the other hand, might be able to give an interpretation that is in the right direction (for example the right hardness category).

From the results it can also be concluded that both systems, but especially the expert system, can be useful if a human analyst wants a second opinion on his interpretation. For example, the analyst was uncertain about the traces on tool number five, and both the expert system and the neural network confirmed the interpretation. It is, however, the expert system that is best suited to this purpose; no less than 93 percent (14 out of 15)[16] of its interpretations endorsed those of the professional analyst. Moreover, in contrast to those of the neural network, the expert system's performances for the replicated tools and those for the archaeological implements show no significant difference. Because of this the expert system is the most reliable of the two applications.

The final conclusion is that both applications already performed quite well. Especially in view of the fact that they were based on a rather small and unbalanced set of examples, their achievements were encouraging. The expert system interpreted 54 percent (14 out of 26 tools) correctly and the neural network 58 percent (15 out of 26 tools) — a difference in performance which seems small. However, if the wrong interpretations are also taken into consideration the difference is greater: 7 percent of the expert system versus 42 percent of the neural network.

Nevertheless, it cannot yet be concluded that one of the techniques is more suitable for this kind of analysis than the other, because the misinterpretations of both applications are still due to insufficient knowledge rather than to inadequacies of the applied techniques. More tests will have to be carried out to obtain a more clearly defined picture of their specific potentials for the analysis of use-wear traces. Such tests should incorporate a comparison of the performances of all three 'types' of analysts. This means that both applications should be employed in a real 'blind test' in which human experts also participate.

5. Some facts and fictions

As mentioned in the introduction, it is not only ignorance that may discourage people from employing artificial intelligence techniques, but also threshold fear. This threshold fear is mainly due to the endless discussions on the potentials, applicabilities and threats of specific techniques as well as of artificial intelligence in general. Unfortunately, most of these discussions are either based on fictions or cause new fictions. In the following I will therefore concentrate on three major points of discussion and will attempt to divorce some facts from the fictions.

A first point of discussion concerns the potentials of the techniques discussed above. It is said, for instance, that neural networks are superior to expert systems in terms of functional abilities and social acceptability (Gibson 1992, 265). But, first of all, as this is comparing apples and oranges, it is, strictly speaking, impossible for either one to be superior to the other. If this were possible it would imply that a person who is specialised in reasoning through association is superior to a person who is specialised in reasoning through deduction. Furthermore, neither one is superior to the other since both techniques have their advantages as well as their disadvantages. They are equipped with different knowledge representational and reasoning methods, which implies that they may only be more suitable for specific purposes. It is a fact that expert systems perform best in tasks involving explicit knowledge and deductive reasoning, while neural networks perform best in recognising complex non-linear patterns or tasks in which the relations between the variables are unknown. Therefore, if both techniques are used for the kind of tasks they have been designed for, they are equally useful.

The same line of reasoning can be used for the idea that neural networks can overcome the major functional disabilities of expert systems. The functional ability of any system depends on whether the applied technique suits the task it is employed in and on the composition of the knowledge that is utilized.

One of the arguments that is used to demonstrate the superiority of neural networks is that they work well with incomplete data and that their performance at the edge of knowledge is far better (Gibson 1992, 265). This suggests that neural networks are able to interpret exceptional situations, which deviate from those they have learned. However, if the functional ability of a neural network were superior to that of an expert system, the results of the above test case would have demonstrated this. It is true that the network reacted more flexibly to exceptions, in the sense that it generated answers in all the cases that the expert system could not. But we have seen that these answers were not very reliable; 8 out of 11 were wrong. It is a fact that a neural network's interpretation is always an estimation, which means that it may be correct but it may also be very wrong. In other words, when a network has to perform at the edge of its knowledge its achievements are poorer. And the question is whether in such cases an unreliable answer is preferable to no answer and whether the latter must, therefore, be classified as a functional disability.

Apart from this, the idea that neural networks can work well with incomplete or exceptional data is an example of wishful thinking. No artificial intelligence technique can work well with incomplete data as long as even humans have great difficulties interpreting situations they have hardly any knowledge of.

Another argument for the superiority of neural networks is that they have the capacity to formulate their own representations of the expert's reasoning processes, without a designer having to make a knowledge model. However, the question is whether this is always an advantage. It makes a neural network a 'black box' and the user has very little influence on the composition of the internal knowledge. As a consequence, the user may have difficulties finding out exactly what a network has 'learned'. It may have learned to distinguish the examples from one another on the basis of properties which are background noise and have nothing to do with solving an archaeological problem. Furthermore, if a network structure is not well designed, for instance if the hidden layer is composed of too many neurons, it may learn the examples by heart and may not be able to analyze any new problem (Lawrence 1991, 123).

With expert systems, on the other hand, this is not a problem as they are 'transparent boxes'. Because their reasoning processes are based on a formal model, they are controlled by the designer and are visible to the user. The advantage of such a model is that it can be used to check the consistency of the knowledge, to localise performance deficiencies and to maintain a system. This means that if an expert system's knowledge needs to be maintained, its knowledge base can simply be adapted or extended. A network, on the other hand, must be trained all over again in such a case.

However, working with a predefined model means that the development of an expert system application is far more time-consuming than that of a neural network. As it is often difficult to build a knowledge model, it is easier for archaeologists to develop a network application than an expert system application.

This brings us to the point of the social acceptability. According to Gibson (1992, 264), this is problematic with expert systems as 'People tend to fear technology when it is professed to have qualities that humans have.' (Gibson 1992, 264). Indeed, the social acceptability of expert systems has been far from satisfactory. However, this is not an inadequacy of an expert system itself. It has to do with the readiness of the potential users and I therefore doubt whether this will not be a problem with neural networks, too. It may even be worse. A network is also professed to have qualities that humans have. And on top of this, neural networks allow less human interference. The fact that they make their own 'rules' for handling the knowledge means that their composition cannot be controlled by the developer nor recovered by the user. Moreover, their internal processing is not only invisible to the user but also more complicated to understand than that of an expert system.

One of the facts that may favour the social acceptability of neural networks is their processing speed. Since they have a neuron structure and a data matrix (representing the weights of the connections between the neurons), the analysis process amounts to nothing more than a calculation of the degree of neuron activity, which takes only a split second. Expert systems, on the other hand, involve lots of rules. During execution they must check all the rules, or at least those for retrieving the appropriate knowledge. That is why a neural network can be faster than an expert system.

To summarize this discussion, I do not think that the neural networks' performance at the edge of their knowledge and their independence in knowledge modelling give reason to believe that their functionality is superior, nor that their ease of development and their processing speed will be crucial for an improvement of their social acceptability.

A second major point of discussion concerns the suitability of archaeological knowledge for an artificial intelligence approach. It has for instance been said that archaeological knowledge does not lend itself well to representation by means of an expert system due to its subjective and intuitive nature (*cf.* Gibson 1992; Vitali/ Lagrange 1988; Wilcock 1986). Expert systems are claimed to be inadequate for representing this kind of knowledge because it can hardly be translated into explicit rules. In my opinion, such scepticism is based mainly on disappointment following overoptimistic expectations of abilities. No technique can provide answers to all questions or handle all problems. Moreover, these techniques are research and educational means instead of solution-generating tools. Their most important merit, however, is that they can play an important role in modelling knowledge (Doran 1990) and thus in understanding and improving it. And in that respect they can be applied to a vast range of archaeological research fields.

Moreover, part of the criticism is momentary and not as absolute as it sometimes seems to be. The knowledge representational abilities of expert systems are not confined to rules. Since these techniques are relatively new they are still evolving. Each new development creates new knowledge representational abilities or facilitates application building.

Finally, I would like to comment on the use of artificial intelligence techniques in general. It is often thought that they can only be used by mathematically grounded archaeologists or that they should not be used at all because they threaten the position of human experts. First of all, the idea that artificial intelligence techniques are difficult to learn is a relic from the early days of computing. Until a few years ago, applications could only be developed with the aid of complicated computer languages. However, this is changing rapidly. Artificial intelligence technologies are still evolving and are already designed to be applied by different groups of users: many of these techniques are commercially available as user-friendly packages.[17] The consequence is that application development no longer requires an awful lot of hardware or software knowledge.

In this respect expert system technology in particular has evolved much. So-called expert system shells have been developed to facilitate application building. Shells are software packages which offer all kinds of knowledge representational and reasoning facilities. With those shells it is possible to build applications without having to program complex procedures. That means that archaeologists who have no experience whatsoever with sophisticated computing techniques can now build an application.

Anxiety should not be an argument for not using artificial intelligence tools either. The fear of these tools becoming a substitute for human intelligence is probably due to the fact that the consequences of future developments cannot be known.[18] The risk of a computer application threatening the work or position of an archaeologist is however entirely fictional. An artificial intelligence application can serve as an assistant only. It can simulate predefined reasoning processes but can certainly not generate intelligence[19]. In other words, an artificially intelligent archaeologist is artificial, not intelligent; it cannot replace an expert. On the contrary, it may even consolidate the expert's position, because it enables the expert to expand his knowledge and to exploit exactly those human abilities that a computer cannot simulate.

6. Concluding remarks

Since papers on computing techniques are often intended to show the disabilities rather than the abilities of artificial intelligence techniques, the majority of the potential users of these techniques among archaeologists have definitely been negatively influenced by this scepticism. They think of them as techniques with a limited potential or as too complicated. In spite of the scepticism surrounding these techniques, there is a lot of work they could do in archaeology. They offer a means for formalising and modelling subjective expert knowledge and they can make this knowledge accessible to and usable for non-experts. In that respect they can be useful tools for educational as well as research purposes. I hope that this paper will help archaeologists to think of ways of using artificial intelligence techniques to approach archaeological tasks or solve problems. Only by building or using an application will archaeologists be able to discover the abilities of artificial intelligence techniques and realize that there is a broad range of potential applications.

However, both purposes and techniques must be selected with care. As argued above, neither one of the artificial intelligence techniques is more useful than the other. Each performs best when the principles of the applied technique suit the problem being tackled. There is, for instance, no sense in applying these techniques to problems having an algorithmic structure which leads to a clear solution. These problems are best handled by conventional programming techniques. But problems that can only be solved by searching for associations with similar situations are probably best tackled via a neural network application. Neural networks are good at recognizing patterns in non-linear problems and at revealing relationships between a problem's input and output variables. However, they cannot be used for problems that require (mathematical) precision. The output of a network is always an estimation which is based on generalisations. The output of expert systems is based on rules of deduction. Because of this neural networks may be less accurate than expert systems. As the deduction process of expert systems is based on a formal model it is also more controlled. If such a model has been developed well, the interpretation of the application will hence be accurate and reliable. Furthermore, an expert system may explain its reasoning process to its user, and thus show how its output was established. That is why expert systems can make better educational and hypothesis verifying applications than neural networks.

The emphasis in this paper has been on the differences between these artificial intelligence techniques and their specific abilities. However, this does not mean that they should only be used for different applications. In archaeology, many research fields involve several types of knowledge. The advantages of different techniques should be combined so as to make artificial intelligence a useful approach for these fields, too. Future discussions should therefore also concentrate on the possibilities of joining different techniques within single applications.

notes

1 Such tasks may be diagnosis, process control, instruction, prediction, classification, planning, *etc*.

2 Knowledge can be represented by means of several methods, not only IF-THEN decision rules, but also Object-Value-Attribute triplets, Inheritance frames, Logic rules, *etc*. For more detailed information the reader is referred to Lucas/Van der Gaag (1988) or Payne/McArthur (1990).

3 Information may consist of observations or facts. The external world may be, for instance, a human user, a database or an instrument.

4 Many different types of neural networks are commercially available. They all have different layer structures, learning algorithms, weight-processing and signal-transfer functions, *etc*. The interested reader is referred to Carling (1992) or Lawrence (1991) for detailed information.

5 Both applications run on a stand-alone computer with a 80386DX processor.

6 A.L. van Gijn

7 WAVES stands for Wear Analyzing and Visualising Expert System. An expert system shell called LEVEL5 Object (version 2.5) is used for the implementation of this system. LEVEL5 Object is a registered trademark of Information Builders Inc.

8 A software package called BrainMaker (version 2.3) is used for this application. Brainmaker is a registered trademark of California Scientific Software.

9 For instance, polish may have been obliterated by post-depositional processes and edge removals may have been caused intentionally.

10 For more information on the composition of the scores the reader is referred to Van den Dries (1994).

11 M.J. Schreurs.

12 The interpretations were evaluated exclusively on the basis of the materials that scored best.

13 Since traces of butchering include traces of bone, the system's identification of either one of the two was accepted.

14 An unsure interpretation was recongisable as an output showing several materials with low scores.

15 Because of the diversity of the experiments the replicated tools showed a large variety of traces, including generic weak polishes. The archaeological tools analyzed, on the other hand, had been carefully selected. Tools without diagnostic traces could often not be analyzed. The remaining archaeological tools showed more diagnostic traces than the experimental tools and may therefore have been easier to interpret.

16 This calculation included the interpretations of the implements used in the experiments because these interpretations were based on descriptions given by a human analyst.

17 These packages are available for all kinds of computer systems at varying prices.

18 Something similar happened when calculators were introduced: people were so afraid of loosing their jobs that they went on strike. It is natural for people to tend to look before they leap when innovations are introduced, especially when those innovations aim to unravel the key to humanity, intelligence.

19 It has been said that a neural network can beat the human brain as far as the speed of signal interchange is concerned (Carling 1992). The speed of the interchanges in the human brain is restricted by chemical processes whereas neural networks can exploit the advantages of using much faster electronical processes (Vuik 1993). Nevertheless, the processing capacity of an artificial network does not even approach that of a biological network. This is due to the fact that biological neurons can all be active at the same time whereas the activity of artificial neurons is restricted by the hardware, which is only capable of sequential processing. Furthermore, a neural network has only few connections compared with the brain; some ten thousands versus billions. Only if it should prove possible to equip future computers with such massive parallel processing capabilities (or human brains) could the speed of the information transfer of artificial intelligence applications possibly defeat that of human beings. But even then, those computers would only be extremely fast, not intelligent.

references

Baker, K.G.	1987	Towards an archaeological methodology for expert systems. In: C.L.N. Ruggles/ S.P.Q. Rahtz (eds), *Computer and quantitative methods in archaeology* 1987, 229-236.
Bishop, M.C. J. Thomas	1984	BEAKER: an expert system for the BCC micro, *Computer applications in archaeology*. Conference proceedings 1984, 56-62.
Brough, D.R. N. Parfitt	1984	An expert system for the ageing of a domestic animal, *Computer applications in archaeology*. Conference proceedings 1984, 49-55.
Cahen, D. J.P. Caspar M. Otte	1986	*Industries lithiques danubiennes de Belgique*. Liège (ERAUL, 21).
Carling, A.	1992	*Introducing neural networks*. Wilmslow.
Dries, M.H. van den	1994	*WAVES: an expert system for the analysis of use-wear on flint artefacts*. In: J. Johnson (ed.), *Methods in the mountains*, Proceedings of UISPP Commission IV Meeting Mount Victoria, Australia. Sydney University Archaeological Methods Series 2, 173-181.
Dries, M.H. van den A.L. van Gijn	in press	The representativity of experimental usewear traces. In: A. Ramos-Millán (ed.), *Siliceous rocks and culture*, Proceedings of the VI international flint symposium, Madrid.
Doran, J.E.	1974	What computers can't do for archaeologists, *Computer applications in archaeology*. Conference proceedings 1974, 69-70.
	1987	Anthropological archaeology, computational modelling and expert systems, In: M.S. Aldenderfer (ed.), *Quantitative research in archaeology: progress and prospects*. Newbury Park, 73-88.
	1988	Expert systems and archaeology: what lies ahead ?, In: C.L.N. Ruggles/S.P.Q. Rahtz (eds), *Computer and quantitative methods in archaeology 1987*, BAR International Series 393, 237-241.
	1990	Computer-based simulation and formal modelling in archaeology: a review, In: A. Voorrips (ed.), *Mathematics and information science in archaeology: a flexible framework*, Studies in Modern Archaeology 3, 1-6.
Doran, J.E. F.R. Hodson	1975	*Mathematics and computers in archaeology*. Edinburgh.
Francfort, H.P.	1991	Palamede: application of expert systems to the archaeology of prehistoric urban civilisations, In: K. Lockyear/S. Rahtz (eds), *Computer applications and quantitative methods in archaeology 1990*, BAR International Series 565, 211-214.
Gibson, P.M.	1992	The potentials of hybrid neural network models for archaeofaunal ageing and interpretation. In: J. Andresen/T. Madsen/I. Scollar (eds), *Computing the past: computer applications and quantitative methods in archaeology 1992*. Aarhus, 263-271.
Gijn, A.L. van	1989	The wear and tear of flint: principles of functional analysis applied to Dutch neolithic assemblages, *Analecta Praehistorica Leidensia* 22.
Grace, R.	1989	Interpreting the function of stone tools; the quantification and computerisation of micro-wear analysis, *BAR International Series* 474.

Juel Jensen, H. 1989 Plant harvesting and processing with flint tools in the Danish Stone Age. A view from the microscope, *Acta Archaeologica* 59, 131-142.

Keeley, L.H. 1977 Beobachtungen über Mikro-Abnutzungsspuren an 14 Klingen von Hienheim, *Analecta Praehistorica Leidensia* 19, 71-72.

Kendall, D.G. 1963 A statistical approach to Flinders Petries' sequence dating, *Bull. International Statistics Institute* 34, 657-680.

Kidd, A.L. 1987 Knowledge Acquisition: an introductory framework. In: A.L. Kidd (ed.), *Knowledge acquisition for expert systems: a practical handbook*. New York, 1-16.

Lagrange, M.-S.
V. Vitali 1992 VANDAL: an expert system dealing with the provenance of archaeological ceramics, based on chemical, mineralogical and data analysis information, In: P. Reilly/S. Rahtz (eds), *Archaeology and the information age*. London, 276-287.

Lawrence, J. 1991 *Introduction to neural networks: computer simulations of biological intelligence*. Grass Valley.

Lucas, P.J.F.
L.C. van der Gaag 1988 *Principes van expertsystemen*. Schoonhoven.

Patel, J.
A. Stutt 1989 Beyond classification: the use of artificial intelligence techniques for the interpretation of archaeological data. In: S. Rahtz/J. Richards (eds), *Computer applications and quantitative methods in archaeology 1989*, BAR International Series 548, 339-347.

Payne, E.C.
R.C. McArthur 1990 *Developing expert systems; a knowledge engineer's handbook for rules & objects*. New York.

Unrath, G.
L.R. Owen
A.L. van Gijn
E.H. Moss
H. Plisson
P. Vaughan 1986 An evaluation of use-wear studies: a multi-analyst approach. In: L.R. Owen/G. Unrath (eds), *Technical aspects of microwear studies on stone tools*. Tübingen, 117-176.

Vitali, V.
M. Lagrange 1988 VANDAL: an expert system for the provenance determination of archaeological ceramics based on INAA data. In: S.P.Q. Rahtz (ed.), *Computer and quantitative methods in archaeology 1988*, BAR International Series 446 (i), 369-375.

Voorrips, A. 1990 Expert systems and archaeologists, In: A. Voorrips/B.S. Ottaway (ed.), *New tools from mathematical archaeology*. Warsaw, 115-121.

Vuik, K. 1993 De heuglijke mens, *Personal Computer Magazine* 11 (1), 184-191.

Wilcock, J.D. 1986 A review of expert systems: their shortcomings and possible applications in archaeology, *Computer applications in archaeology* 1985, 139-144.

M. van den Dries
Instituut voor Prehistorie
P.O.Box 9515
NL 2300 RA Leiden