Archaeology and the application of artificial intelligence : case-studies on use-wear analysis of prehistoric flint tools

Dries, M.H. van den

Citation

Dries, M. H. van den. (1998, January 21). *Archaeology and the application of artificial intelligence : case-studies on use-wear analysis of prehistoric flint tools*. Retrieved from https://hdl.handle.net/1887/13148

5 An expert system application for use-wear analysis: WAVES

5.1 Introduction

Chapters 3 and 4 introduced the fundamentals and functionality of artificial intelligence techniques and of the knowledge domain of use-wear analysis. These methods are brought together in this chapter by means of a case study: an expert system application for analysing wear traces on flint artefacts.

The aim of this case study has been twofold. The primary aim was to employ artificial intelligence techniques, especially the expert system approach, for the purpose of computer-assisted instruction. Since teaching use-wear analysis is a very time-consuming task for the expert and at Leiden University each year several students want to learn and practice it, a knowledge based application was believed to be a helpful tool. From a research point of view, this was also an interesting attempt as it enlarged the objective from capturing, analysing and reproducing knowledge, into the simulation of an entire process of analysis. In this way it was tried to turn know how into practical assistance.

An additional aim of the development of a knowledge based system was to formalize and model the expert knowledge that is involved in wear trace analysis and to subsequently design a standardized approach of the method. One of the most important reasons for formalization lies in the expectation that it will improve the academic acceptance of the method. Standardization means that it yields less subjective results and improves the comparability of the interpretations of different analysts. Moreover, once knowledge is formalized it can be evaluated in order to trace deficiencies and to refine methodical aspects.

While collecting and analysing the expert knowledge that would constitute the application, it became clear that the education of apprentices involves two levels: a basic and a more advanced level. At the basic level attention focuses on teaching the methodical principles and on guiding the process of analysis. Students who have arrived at the second level are becoming more experienced and are able to reach an interpretation autonomously. Therefore they are more interested in a means that verifies their interpretations rather than one that supports the interpretation process. Moreover, a hypothesis verification approach may also be useful for experienced analysts. The subjective nature of the method of analysis and the limited diagnostic value of many wear traces makes that they frequently would appreciate a second opinion on their interpretations.

The two-level approach automatically implied that the application had to be divided into two independently operating components, *i.e.* an analysing part and a hypothesis validating part. Whereas the analysing part is meant to be used by students in order to learn and practise the analysing process, the hypothesis validating procedure is to be used by more experienced students or analysts for the purpose of evaluating their interpretations. This evaluation means that the user presents the supposed contact material to the application which, in its turn, will ask the user to verify the presence of the wear characteristics that are expected on a tool that has been used on that contact material. If their presence cannot be acknowledged the user is advised to reconsider his hypothesis.

This chapter has been divided into two parts. In the first part the building process of the application, which is called WAVES (Wear Analysing and Visualizing Expert System), will be described. The second part will address to the functional aspects of the operational application. In outline this means that in paragraph 5.2 a description is given of how the knowledge has been elicited, analyzed and modelled. Paragraph 5.3 focuses on the phase of design and implementation. All important aspects and the subsequent decisions that were made will be discussed, such as the way uncertainty within the knowledge has been managed. This will be followed by a justification of these decisions (paragraph 5.4). Despite the fact that the development phases are handled separately, which might lead to the impression that they are strictly separated, in practice they are closely related. The boundaries between the different development stages are often fuzzy. In each phase the work from the previous stage is modified and refined and during the entire development process all steps ahead are kept in mind. That is why at certain points in this chapter decisions may be discussed which formally may not belong to that stage of the process.

The second part of this chapter shows WAVES in action. First it will be discussed how both procedures of the application are composed and how they work: in paragraph 5.5 a description is given of the composition and the knowledge representation features of the analysis procedure, in paragraph 5.6 the same has been described in reference to the hypothesis validation procedure. Subsequently the functionality of the application as a whole is looked at (paragraph 5.7) and evaluated in paragraph 5.8. This has been done on the basis of the demands that are posed by archaeologists (as mentioned in chapter 2) and by use-wear analysts in particular (as mentioned in chapter 4). Both the requirements that were met and the unsolved difficulties will be discussed. Finally, WAVES will be compared with an expert system application which was developed for use-wear analysis by Roger Grace (Grace 1989). For an evaluation of the actual performance of WAVES, the user is referred to chapter 7, in which the results from blind tests are given that were carried out by students as well as experienced analysts.

The reader will notice that this chapter pays considerably more attention to the construction of the analysis procedure than to that of the hypothesis validation procedure. This disparity, which reflects the division of the attention during the development process, has two main reasons. First of all the educational task of the analysis procedure required several additional facilities and abilities which the other procedure did not need. These were requirements, such as clarifications of the reasoning process, that demanded much attention. The other reason is the simple fact that the analysis procedure was built first. As a consequence, several aspects of the development process, such as the knowledge formalization approach and the solutions that were found for all kinds of difficulties, are only discussed in the context of this procedure. Many of these aspects could be employed in the hypothesis validation procedure as well, but have not elaborately been discussed again.

part one: WAVES under construction

5.2 The knowledge acquisition

5.2.1 INTRODUCTION

In order to create a knowledge-based application, one must go through a trajectory of knowledge elicitation, analysis, and modelling, followed by the design of the application and the implementation of the knowledge. In chapter 3 (section 4.3) these stages of a development trajectory were discussed in rough outline. In this paragraph they will be discussed in relation to the construction of WAVES.

Normally, the first step of any application building process is a viability study. It must be investigated whether a project has an actual chance of succeeding. In the case of WAVES it seemed that most of the demands that were posed on the knowledge and on the context in which an application was going to be used (see paragraph 3.4.2), could indeed be met. The domain expert was willing to share her knowledge and to support the development process, students in use-wear analysis did not oppose to computer aided learning, and all means for obtaining hardware and software were available. This was considered to be an acceptable basis to start an investigation on the adequacy of artificial intelligence techniques for computer-assisted instruction of use-wear analysis.

When the actual development process starts, the next step is to collect the required knowledge (paragraph 5.2.2). The result of this elicitation phase is believed to be "*…critical since the power and utility of the resulting expert system depend on the quality of the underlying representation of expert knowledge.*" (Kidd 1987: 1). Once the knowledge is withdrawn from the experts, it must be analyzed (paragraph 5.2.3). It has often been stressed (*cf.* Breuker & Wielinga 1989: 273) that it is of major importance to make a full analysis of the problem domain and of the knowledge involved, because it is only possible to develop a system that fits a task well if that task has been analyzed in detail.

When the analyses are finished, a model of the knowledge must be made (5.2.4.) in order to create a basis for the subsequent design of the application.

5.2.2 THE ELICITATION OF THE EXPERT KNOWLEDGE

The main goal of the elicitation phase is to find out what knowledge the domain expert applies and what decisions he or she makes during the process of wear-trace analysis and for what purpose. The aim is to gain as much 'raw data' as possible in order to get insight into the principles of the method.

Theoretically, it should be possible to extract the required domain knowledge from human experts directly and from their publications. In practice, however, this may be difficult to accomplish because various problems may be encountered. During the elicitation for WAVES, a number of such problems indeed appeared. A first difficulty that knowledge engineers often experience is that experts tend to give a somewhat idealized view of the work they are engaged in (*e.g.* Payne & McArthur 1990). Unintentionally they emphasize ideal and clear situations and tend to forget problematic ones. As a result, knowledge which is withdrawn from experts does not have to be fully representative for the real-world situation.

This was one of the difficulties which was indeed encountered during the knowledge elicitation for WAVES. At first, it was tried to extract the required knowledge from human analysts[1] directly and from research reports. But unfortunately, the information that was withdrawn in this way was unbalanced and insufficiently detailed. It consisted predominantly of data on wear patterns that are diagnostic, while deviations from these 'ideal' patterns were underrepresented. Consequently, it did not cover enough of the variability of the wear traces that may occur both in an experimental environment and archaeologically. Especially in papers this kind of information is missing because it is considered not to contribute much to the archaeological discourse. From a scientific and methodical point of view, this is of course understandable and acceptable. From the point of view of a knowledge engineer, however, the emphasis on diagnostic patterns is undesirable, in particular when one aims to develop an educational application. Students have to learn to deal with both straightforward wear-patterns and complex situations, because a large part of the wear patterns they will encounter in practice belongs to the latter category. Therefore, an expert system application should not only be able to interpret diagnostic patterns, but ought to have as much knowledge on use-wear variance as possible. WAVES would not be a useful product if it were merely applicable for a small number of cases.

The process of elicitation can also be complicated due to the involvement of various experts which all cover part of the domain. The difficulty of knowledge that is collected from various people or publications is that it is often merely a collection of fragments which can hardly be structured into a coherent model. During the elicitation of knowledge on use-wear analysis this problem occurred as well. It was experienced that there is not enough similarity between the descriptions of the observations given by different use-wear analysts to allow for a meaningful comparison of interpretations. The reason for this is twofold. First of all it has to do with the subjective nature of the method: different analysts may observe or describe a particular wear feature differently. Secondly, there is no absolute consensus within the discipline as to nomenclature: each expert develops his or her personal method of analysis and reasoning. This may be explained by the fact that on a world wide scale there are only few analysts. Since they are geographically isolated they are more or less forced to operate in a manner they developed single-handed. Again, from a scientific point of view deviating approaches may have a stimulating effect. Each newly developed manner may consist of useful methodically innovations. But, students and knowledge engineers who want to understand the discipline beyond the knowledge of a single expert, will experience these differences as confusing.

A third potential problem that may be encountered during the knowledge acquisition phase lies in the fact that expert knowledge not only consists of formal facts and theories that are gained by education, but in particularly of subjective rules-of-thumb that are gained by experience. Because of its subjective nature, it is exactly the latter kind of knowledge which experts feel hard to describe explicitly (Kidd 1987: 3). Experts, as well as other reasoning humans, simply do not reason in the same formal way as an expert *system* does. This lack of clearly defined knowledge turned out to be the most difficult problem to overcome in the elicitation for WAVES. In order to understand and to reproduce the deductions and interpretations of an analyst, a knowledge engineer — and a student — needs systematical information of the format "if you observe wear attributes A, B and C then this is an indication (with certainty degree X) that this tool was used in activity D or E, and not in activity F, because...". Although all analysts surely reason in this way when they analyze a tool, they not always put this in writing in their final reports, at least not in every detail. This has two reasons. Firstly, this kind of knowledge is very basic for experts and they do not need to discuss this on paper anymore. Secondly, use-wear 'types' cannot be clearly circumscribed. They tend to shade off into one another. Therefore, the interpretation of wear traces does not solely involve formal rules but aspects of association as well. This makes that analysts can arrive at an interpretation through almost an infinite number of reasoning paths. Unfortunately, they can impossible put all of these in writing, even if they would wish to.

In short, the dissimilar character of the available knowledge and the fact that it is very difficult to discover how *exactly* an analyst arrives at a particular interpretation, made the first elicitation attempt not very successful. It turned out to be a hazardous task to gather sufficient and suitable data for building a coherent knowledge model from papers and interviews with analysts. This is certainly not meant as a critical note towards use-wear analysts, because their knowledge domain hardly allows them to do differently. Nevertheless, it has been a useful experience. It made clear which educational aspects ought to be incorporated in WAVES and which methodical aspects required further research. Moreover, this may be important information for teachers. Students probably encounter the same difficulties as I did in trying to get a comprehensive understanding of the domain.

In order to cope with most of the above mentioned difficulties, it was decided to use an alternative source of knowledge as a starting point: a reference collection of wear-traces obtained by experiments. Fortunately, such a reference collection was already available. Since Annelou van Gijn had started to employ use-wear analysis at Leiden University in 1984, she and her students have been building a reference collection of experiments and they systematically documented the wear traces that they obtained. She kindly has put her data at the project's disposal for the purpose of constructing a knowledge base.

Important advantages of experimentally obtained data are that they render a large amount of information on the variety of the traces that may occur under more or less controlled circumstances; that the obtained results become comparable and calculable because they are systematically and uniformly

Observed wear pattern: regular shape use retouched; domed polished surface with few striations, etc.

Variables:	**Attributes that describe wear features:**
use-retouch distribution:	close, clumped, overlapping, etc.
polish topography:	cratered, pitted, domed, etc.
number of striations:	none, few, many, etc.
edge shape:	straight, concave, convex, etc.
etc.	

Fig. 11. Explanation of nomenclature: a wear pattern consists of all observed features. Variables and attributes are used to describe the observed wear features and characteristics of an implement.

described; that the subsequently deduced knowledge is less subjective compared to knowledge elicited from various experts; and that the origin of the knowledge is known, traceable and, therefore, assessable.

5.2.3 THE DATA ANALYZED

The experimental program from which the knowledge for WAVES has been derived, consists of 301 experiments with replicated flint artefacts that were altogether used in five types of movements and on twenty contact materials (Van Gijn 1989: 168-174). This program consisted of a broad and varied sample of experiments that were executed with various types of implements of various raw materials and for various periods of time. For the description of the experimentally obtained data 50 variables and 353 attributes had been used; 32 variables (and 253 attributes) for the description of the morphological aspects and condition of an implement prior to its use and 18 (100 attributes) for the obtained wear traces (see Van Gijn 1989: 163-168) (see figure 11 for explanation of nomenclature).

Although these data showed a more divergent occurrence of wear patterns than the elicited expert knowledge did, it was still impossible to translate it directly into lines of reasoning that relate wear features to contact materials or motions. The problem was that very few wear features were exclusively linked to a particular activity or motion. Different activities caused similar wear patterns, while similar or identical activities resulted in various wear patterns. It was also clear that there is a restricted variation in wear features, which does not coincide with the endless variation of contact materials and motions. These were no new discoveries, however, but had been stressed before (*cf*. Unrath *et al*. 1986).

In order to deal with these difficulties and to be able to deduce a knowledge model from these data, it had to be examined whether relations exist between the observed wear-features and the contact materials and motions that were used in the experiments. The main aim of the analysis was to determine the diagnostic value of the variables and attributes and to subsequently select those that are relevant for the application.

For this purpose, the data were analyzed by means of cross-tabulations. These made it possible to count the frequency of occurrence of all attributes of each variable in relation to the various contact materials (see for an example table 1) and motions.[2] In this manner it could be studied whether the presence of the observed wear-features are diagnostic for working a particular contact material or for a specific motion. It turned out that few features were exclusively associated with one particular contact material or motion, but some seemed to relate to the hardness of the worked material. With respect to the use-retouch attributes, for instance, it was found that the step termination of scars predominantly occurs with hard animal material like bone and antler, while feather shaped scars often relate to soft or medium hard materials. Hinge terminations were mainly found on implements used on medium hard vegetal and animal materials, like hard wood, soaked antler and fish scales.

The hardness of the worked material seemed to be of significance for the length of the retouch too; removals larger than 500 were not caused by soft materials and removals larger than 5.0 mm by very resistant materials only. No relation was found between the length of scars and the applied motion. The opposite was true, however, with the distribution of the scars along the edge of the implement. Regular distributions were found to be related to a transverse motion, whereas irregular and overlapping removals seem to be predominantly caused by longitudinal motions. Concerning the development of edge rounding it was concluded that this remains difficult to understand. It surely depends on the applied motion as well as on the hardness of the worked tissue and on the duration of the activity. In a few cases, however, it was noticed that edges hardly lost sharpness after being used for one hour or more on sturdy matter like wood and antler. Some of these did not show edge retouch either.

Table 1. Relationship between retouch terminations and contact materials. All applied motions are included, because no relationship was discovered between the type of termination and the involved movement.

	step		hinge		feather	
hide	0/5	(0%)	2/5	(40%)	3/5	(60%)
wood	1/24	(4%)	15/24	(63%)	8/24	(33%)
bone	9/35	(26%)	23/35	(66%)	3/35	(8%)
antler	0/6	(0%)	3/6	(50%)	3/6	(50%)
siliceous plant	0/5	(0%)	0/5	(0%)	5/5	(100%)
non-siliceous plant	0/4	(0%)	2/4	(50%)	2/4	(50%)
cereals	0/6	(0%)	0/6	(0%)	6/6	(100%)
meat	0/2	(0%)	1/2	(50%)	1/2	(50%)
butchering	1/4	(25%)	1/4	(25%)	2/4	(50%)
fish	0/19	(0%)	8/19	(42%)	11/19	(58%)
pottery	0/2	(0%)	1/2	(50%)	1/2	(50%)
shell	2/5	(40%)	2/5	(40%)	1/5	(20%)
tooth	1/1	(100%)	0/1	(0%)	0/1	(0%)
limestone	0/1	(0%)	0/1	(0%)	1/1	(100%)
Total	14/119	(12%)	58/119	(49%)	47/119	(39%)

As it was clear that the majority of the single attributes were not sufficiently diagnostic, it was also investigated whether combinations of characteristics were. This revealed that some tendencies were indeed present in the data of the reference collection. Especially the development of edge rounding seems to be related to the occurrence of particular types of use retouch. If the implement showed no edge rounding but abundant step-terminated scars, this was an indication for use on the hardest animal materials. Medium rounding with feather-terminated scars only occurred with tools used on medium to soft animal or vegetal materials. By combining these wear characteristics, six material groups could be identified, although some overlap still remained (fig. 12).

With respect to the polish attributes a similar analysis was made. It was noted that, in contrast to the distribution of the use retouch, the distribution of the polish was seldomly related to the applied motion but clearly to the hardness of the worked material. Isolated spots of polish, for instance, were only caused by hard materials, whereas more widely dispersed types of polish were only seen with medium and soft materials. The texture of the polishes was found to be related to the type of contact material too. Smooth and matt polishes indicated wood and antler working, whereas rough and greasy looking polishes seemed to be caused by hide working predominantly. Like the length of edge removals, the width of the polish appeared to be exclusively related to the resistivity of the worked material. The applied motion neither seemed to affect the topography of the polish. The attributes of the latter, however, were found to be highly diagnostic for particular contact materials. A bevelled topography, for instance, was only caused by bone working and a cratered-like topography seemed to be characteristic for hide working. Furthermore, the analysis indicated that the direction of linear features within a polish strongly relates to the applied motion, whereas it had little relation with the material worked.

Like with the retouch, many polish features appeared to be hardly confined to particular contact materials. Therefore, the approach to look for the occurrence of particular combinations of attributes of several variables was applied again. Once more, it was noticed that a few non-diagnostic attributes gained an increased diagnostic capacity by combining them. In figure 13 an example of such a combination (polish texture and brightness) is given. It shows that an artefact which displays a particular combination of features can only have been used on a few types of contact materials. For instance a very bright polish with a smooth texture and with a domed topography can, according to the results of the experiments, only be caused by working hard or soft wood, silicious or non-silicious plants, cereals or soaked antler. Consequently, in this approach the other materials can be excluded, while on the basis of the diagnostic capacity of the single attributes other materials could have been responsible for these traces as well. For example, a very bright polish can also be caused by hide, leather, soil, dry clay, etc., a smooth texture also by soaked bone, and a domed topography also by dry antler and stone.

termination	rounding		
	absent	medium	heavy
step	HARD ANIMAL	–	–
hinge	HARD ANIMAL HARD VEGETAL medium hard animal	HARD VEGETAL MEDIUM HARD ANIMAL medium hard vegetal	–
feather	MEDIUM HARD ANIMAL SOFT ANIMAL soft vegetal	MEDIUM HARD VEGETAL medium hard animal soft animal SOFT VEGETAL	MEDIUM HARD VEGETAL

Fig. 12. The presence or absence of edge rounding and the kind of termination of the edge removals in relation to different categories of materials. Capitals indicate the main constituent of a group. The minor constituent is set in lower case.

brightness	texture		
	smooth and matt	rough and matt	rough and greasy
very bright	cereals silic. plants soft wood hard wood antler bone polish ‘23’	clay soil antler bone meat and fish hide leather soft wood	–
bright	hard wood soft wood cereals silic. plants non-silic. plants bone antler meat and fish stone	tooth shell pottery clay bone meat and fish non-silic. plants	hide leather fish scales meat and fish antler
dull	–	fish scales leather polish ‘23’ meat and fish bone hide antler wood	leather hide meat and fish

Fig. 13. The relation between the combination of the texture and brightness of the polish and the various contact materials

By means of this simple method of making cross-tabulations, it could be made clear which of the characteristics were diagnostic and to what degree. It must, however, be stressed that even though some wear patterns can — to a certain degree — be related to contact with specific materials, some overlap remains. It was already known that especially not very well-developed use wear can be produced in several tasks and that it remains difficult to interpret (see also Vaughan 1985 and paragraph 4.4). Subsequently, these results allowed the disposal of the experimentally obtained data of noise: wear characteristics that were associated with all materials or motions were considered to be of no diagnostic value.

With respect to the second aim of the analysis, the selection of relevant variables and attributes, the cross-tabulations yielded useful information as well. Especially a large number of the variables by which the morphological aspects and condition of the implement are described did not seem to have directly influenced the occurrence of the wear patterns. Most patterns were present on various kinds of tools and edges. Only 3 variables out of 32 were considered to be relevant and were selected for the application: grain size, edge shape and edge angle. Together these 3 variables comprise 9 attributes. Of the 18 variables that were used for the description of the experimentally obtained wear traces, 14 were selected for the application. Furthermore, some attributes of these variables turned out to have no diagnostic value and were left out as well: 80 out of 100 are included.[3]

5.2.4 KNOWLEDGE MODELLING AND UNCERTAINTY HANDLING

The modelling phase aims to develop a more abstract but comprehensive model of the knowledge and procedures that a task entails. It comprises an inventory of these procedures, of their relations, of the conceptual knowledge that they need, etc. Before such a model can be constructed, however, decisions must be made on the range of the tasks of the application first. For instance in the case of use-wear analysis, the expert's role involves various tasks, like method development, data analysis, hypothesis validation, student tutoring, etc. An expert system application does not necessarily have to simulate all of them. In any case it must be clear which functions will be incorporated and, if applicable, their order of importance. With respect to WAVES, it was decided to focus on three tasks: foremost student education, secondly data analysis, and thirdly hypothesis validation. Consequently, the two main aspects of the knowledge modelling for WAVES were the preparation of the knowledge for the data analysis and hypothesis validation task and to find a means to handle the uncertainties that are involved in the interpretation process.

Firstly, the knowledge which had been unraveled during the analysis phase was organized and structured in order to get a grip on the tasks of the application and to discover hiatuses in the knowledge. It implied that the *conceptual* (or declarative) *knowledge* was separated from the *procedural aspects*. The conceptual knowledge consists of *facts* and *rules*, *i.e.* everything which concerns the content, while the procedural knowledge organizes the reasoning process. It specifies at what point and in what manner which pieces of knowledge will be employed. An example of a procedural knowledge-rule is: 'if a tool shows scars, then it is important to find information on the average length of these scars — if this is unknown, then ask the user'. Such procedural knowledge can be organized as a chain of actions which activate each other. This is knowledge about the handling of the knowledge and may be called 'meta knowledge'. An example, on the other hand, of a rule containing conceptual knowledge is: 'if the average length of scars is less than 1 mm, they are probably caused by contact with a soft material'.

Next an inventory was made of the knowledge that is involved in the wear-trace analysis and in the validation of hypotheses. It was realized that these two approaches were so much distinct that they could best be performed by two independently operating procedures. They not only use a different reasoning strategy but a differently structured knowledge base too. The process of use-wear analysis requires an approach by which a standard trajectory of tasks is followed step by step. Each analysis session must be pursued according to a given method: the observations are being described and on the basis of that description the application must be able to present an interpretation. The hypothesis validating method, on the other hand, cannot work according to a standard procedure, because the input of each validation differs. Therefore, this method needs a flexible reasoning strategy. Moreover, it needs a reversed usage of the conceptual knowledge because by validating a hypothesis, the above described conceptual knowledge is applied in the opposite direction.

One of the difficulties that were encountered in structuring the conceptual knowledge, especially in relation to the analysis procedure, was the modelling of uncertainties. The analysis of use-wear traces is subject to uncertainties on at least three levels: that of the observation, the description and the interpretation. At the level of the observation, the analyst may be uncertain about the observation of the features. It may, for instance, be questioned whether particular wear features can be attributed to human use or whether they are a reflection of the incident light of the microscope or an intrinsic characteristic of the core material. Subsequently, uncertainties may arise at the level of the description. Often it is difficult to describe an observation by means of predefined attributes: the observed feature may not entirely resemble one of the attributes or it may not be clear what is meant with a particular attribute. Finally, at the interpretation level, the limited diagnostic value of wear traces and/or the overlap in wear patterns may complicate the process of deduction (paragraph 5.2.3).

Most of these uncertainties are inherent to the method. Even the most experienced expert may sometimes doubt an observation, may have second thoughts about his description and may not be convinced of his interpretation. By formalizing the method and its knowledge these difficulties become visible, but they cannot be overcome easily. An automated approach does not automatically provide solutions for all problems. For example, an expert system application will not be able to direct the eyes and mind of the human observer or to check whether a description exactly fits the characteristics of the features on the implement. It can only be helpful to direct the descriptions, for instance by providing adequate explanations and illustrations of the attributes and their meaning.

An expert system may also be helpful at the third level of uncertainty, *i.e.* that of the interpretation. First of all, a standardization of the interpretation process makes sure that the analysis of identical patterns always leads to identical interpretations. Secondly, it is technically possible to equip deduction rules with probabilities. These probabilities give an indication of the certainty of an interpretation. They can be based on the experience of an expert or calculated on the basis of a large database. In the first case they are subjective, in the second they can be quite reliable if the database is sufficiently large. Although the experience of the expert may yield good estimations of probabilities, they only ought to be applied when no quantitative data is available.

During the knowledge modelling process for WAVES, it was realized that it was desirable to handle the uncertainty at the interpretation level. Since the basic knowledge was derived from a reference collection a quantitative approach seemed logical. It was thought that on the basis of the experimentally obtained data, it might be possible to calculate the probability that a wear pattern relates to a contact material or motion. However, this expectation turned out to be too optimistic. The collection has not been put together for the purpose of statistical analyses. Therefore, it consists of insufficient similar experiments with the various contact materials. Moreover, it was already known (see paragraph 5.2.3) that similar experiments did not yield corresponding patterns. Therefore, the chances of the occurrence of most patterns were very little. Consequently, all deductions of WAVES could only reach low probabilities. This was not considered to be very meaningful for the students that would apply the analysis process of WAVES.

Nevertheless, it was still considered worthwhile to incorporate in the deduction process of WAVES the quantitative information that the reference collection offered. If it could not yield true probabilities it could at least yield impressions of the frequency in which certain wear features occur in relation to particular contact materials or motions. These frequencies could then be used to associate a wear pattern with the *diagnostic value* that is has in relation to each contact material and motion. This would provide students with a means to assess the interpretation of WAVES and would be a starting point in anticipation of true probabilities.

It is illustrated in figure 14 how the diagnostic values of the wear patterns have been composed. The first step was to deduce the frequency of occurrence of a particular wear attribute in relation to a particular contact material. For instance, if an attribute (A1) was observed four times (3+1) on the experiments that were performed with a particular contact material, its frequency of occurrence is 40 percent. Subsequently the diagnostic value of a complete wear pattern was calculated by simply summing the frequencies by which the individual attributes occurred in the reference collection (step 2).

The wear patterns that are caused by a particular contact material (number of experiments is 10):

3 times A1 / B2 / C1
4 times A2 / B1 / C1
1 time A1 / B1 / C3
2 times A3 / B2 / C2

Step 1: deducing the frequency of occurrence of single attributes (= total occurrence of an attribute divided by the number of experiments):

A1	40%	B1	50%	C1	70%
A2	40%	B2	50%	C2	20%
A3	20%			C3	10%

Step 2: calculating the diagnostic value of the observed patterns by summing the frequencies by which the attributes occur (highest possible value is 160):

A1 / B2 / C1 = 40 + 50 + 70 = 160
A2 / B1 / C1 = 40 + 50 + 70 = 160
A1 / B1 / C3 = 40 + 50 + 10 = 100
A3 / B2 / C2 = 20 + 50 + 20 = 90

Step 3: calculating the diagnostic value of the remaining (non-observed) combinations of attributes:

A1 / B2 / C2 = 40 + 50 + 20 = 110
A1 / B2 / C3 = 40 + 50 + 10 = 100
A2 / B1 / C2 = 40 + 50 + 20 = 110
A1 / B1 / C1 = 40 + 50 + 70 = 160
etc.

Fig. 14. The composition of diagnostic values of wear patterns in relation to a particular contact material. Capitals (A, B, C) represent variables, numbers (1, 2, 3) represent attributes.

However, to distinguish only the diagnostic value of the wear patterns that are present in the experiments was considered insufficient. It had already been experienced that new experiments could yield new combinations of wear characteristics, and it was therefore to be expected that the application would be confronted with such new patterns sooner or later. The application would not be very useful if it would not be flexible enough to deal with such situations. For this reason it was decided to calculate the diagnostic value of unknown patterns as well (step 3).

Similar calculations were made for each contact material and motion. Finally this resulted in lists with diagnostic values that all wear patterns have for each contact material and motion (appendix I). It can be noticed from these lists that the values vary considerably. This not only indicates that some materials are more likely to cause certain wear-patterns than others, but also that particular attributes and patterns are more diagnostic for material X then material Y.

5.3 Design and implementation

5.3.1 INTRODUCTION

To start with, it must be stressed that in this paragraph it has not been tried to include all technical and functional aspects that are involved in the design phase. This would have led to a highly detailed and tedious description. Moreover, the technical aspects of a design are predominantly determined by the abilities of the implementation tool rather than by the design. It was therefore decided to discuss in this paragraph only the main aspects of the design. The emphasis lies on the requirements that were posed on the knowledge handling facilities (section 5.3.2) and the interaction with the user (5.3.3), and subsequently on the consequences that these had on the hard- and software that would be employed (5.3.4).
For each of these main aspects of the design, again only the most important requirements will be discussed. With reference to the knowledge handling abilities these are *representational adequacy*, *correctness*, *modularity*, and *simplicity*; concerning the user interface these are *graphical possibilities*, *user friendliness*, and *explanatory facilities*. Obviously, the requirements on the hardware and software are that they allow the implementation of what is decided on the aspects of knowledge handling and interaction. The explanation of the requirements is followed by a discussion on the decisions and their subsequent implementation in the application. The main criteria for all design decisions were *flexibility*, *transparency* and *maintenance*.

5.3.2 KNOWLEDGE HANDLING

5.3.2.1 Requirements

In chapter 3 several methods for handling knowledge have been discussed. Although it is known that the combination of the knowledge representational and the reasoning method is of crucial importance for the performance abilities of any application, there are no regulations as to what knowledge should be managed by which methods. A system developer is free to use the method of his personal preference. It is, therefore, recommendable to formulate the knowledge handling requirements that the selected method should meet in order to create some guidelines that may ease the selection process and may provide a framework for the evaluation of the final application. The main demands that were posed for the knowledge handling methods used for WAVES are: representational adequacy, modularity, simplicity, and correctness.
The *representational adequacy* is a prerequisite of any combination of representation and inference method. It means that the applied methods enable the incorporation of all the types of knowledge that the application needs, *i.e.* the procedural aspects as well as the conceptual knowledge.
A second demand on the knowledge representation method is that the resulting knowledge base is easy to maintain. This can be obtained by programming an application orderly and modular. A *modular approach* comprises the division of knowledge into small, coherent portions and the handling of separate procedural tasks by separate bits of program. This makes it possible to adapt each procedural aspect or bit of knowledge without having to change the entire application. One of the reasons for this is that some parts need more maintenance than others. Especially conceptual knowledge is highly dependent on regular maintenance because it tends to fossilize quickly. Therefore, the procedural and the conceptual knowledge must be recognizable as such and be kept separate. But, in general, it is important that as many bits of knowledge as possible can be revised without influencing the functionality of other bits. A clear and transparent structure together with detailed descriptions (of the technical functions) of its tasks, maximizes the application's maintenance possibilities and minimizes the required efforts and financial investments.
Thirdly, a representation method should be as simply as possible in order to create a knowledge base that is easy to understand. A plain method not only reduces the chance of programming mistakes, but optimizes maintenance possibilities as well. In my opinion, this is the most important reason for requiring *simplicity*. The use of a simple representation method may enable that others besides the knowledge engineer understand and adjust a knowledge base. This may be helpful to invite experts to take care of their application's maintenance once the engineer has delivered the product.
A fourth demand of the knowledge handling method is that it allows for all the reasoning facilities that are needed to achieve correct interpretations. This *correctness* demand seems plausible, but in some cases absolute correctness is very difficult to achieve. It not only implies a perfect tuning of the input, but also means that an application does not give a wrong interpretation even if it is confronted with false or incomplete input. The difficulty, however, is that the application has no full control over the input that is given by its user. Input can of course be screened on its format, its existence, its reasonableness and its consistency, but it is hardly possible to control its validity. It cannot be checked whether a description of a user fits his or her observation. The application can be equipped to prevent as many

unreliable descriptions as possible, but it has to assume that a user does his utmost to give adequate descriptions.

5.3.2.2. Decisions

In order to meet the first requirement, that of the representational adequacy, it was decided to represent the reasoning processes of both the analysis procedure and the hypothesis validation procedure of WAVES by means of IF-THEN decision rules rather than by means of proposition logic, semantic nets or frames (see chapter 3, paragraph 3). The use of decision rules has several advantages. First of all, they match the deduction processes applied in use-wear analysis best. It was found that the structure of a decision rule corresponds with the way the expert knowledge was modelled: "if characteristic A is observed, then it can be deduced from this that it may be caused by activity X".

A second advantage of IF-THEN rules is that they can be just as easily applied in a forward direction as in a backward direction. Especially the possibility to use rules for both forward and backward reasoning was very convenient, because this could serve both the analysis approach and the hypothesis validation approach of WAVES. In order to create a validation approach, decision rules can simply be used in the opposite way as in a deduction process: "if I want to relate these traces to activity X, then the presence of characteristic A has to be acknowledged".

A third reason favouring rules relates directly to the requirements of modularity and simplicity. Rules can meet the requirement of modularity fairly easy as they intrinsically provide the opportunity to structure knowledge into small, task-dedicated modules. This facilitates the maintenance of the knowledge since each rule is a small piece of independent knowledge which can be edited or deleted on an individual base. Moreover, a knowledge base that consists of rules can easily be supplemented with new rules. Especially for WAVES, it was significant to take the aspect of maintenance into consideration, because some parts of its knowledge, such as the diagnostic values, are susceptible to changes. Rules also meet the requirement of simplicity due to the fact that their IF-THEN syntax resembles ordinary human language. This makes them easy to learn and work with and, therefore, more *developer-friendly* than, for instance, a representation method like proposition logic.

In contrast with the former three requirements, the answer to the fourth, *i.e.* that of correctness, is not determined by the representation method. Correct interpretations can only be achieved on the basis of high quality input and of adequate internal reasoning processes. Correctness first of all implies that well-described observations must lead to correct interpretations. This is such a fundamental aspect that it has been a continuous point of attention during the design and implementation of the application. Moreover correctness is an aspect that can and must be verified. Therefore a validation test was designed for WAVES. It had to be carried out once the implementation would be finished. The application would be tested on a carefully compounded set of cases and its interpretations had to be in close harmony with those of the expert before it would be released for practical use. The results of this evaluation will be discussed in chapter 7.

Furthermore, correctness implies that non-interpretable input should not lead to false interpretations. That is why the analysis procedure of WAVES only connects a wear pattern with a material or motion if the observed features match the features which are expected with that particular material or motion exactly. If a feature is observed which has never been seen in relation to that particular contact material or motion, this material or motion is excluded from the interpretation. Moreover, when the application receives incomplete descriptions or encounters a combination of characteristics which has never been seen with any contact material before, the application should not give an interpretation. This may seem a rigorous approach, but I consider it to be better to have no interpretation at all rather than one which is false. Besides, if WAVES cannot interpret a pattern the conclusion is justified that this pattern is really unusual, because it is able to interpret all kinds of non-diagnostic patterns.

A requirement that relates to correctness is the reliability of input. Since this is entirely dependent of the user, this demand is the most difficult to meet. In WAVES the format, existence and reasonableness of the input is controlled by using pre-programmed multiple choice options from which the user can select his answers. The validity of the input, however, is hardly verifiable. In some cases the consistency between the description of two variables can be checked (in case there is a dependency), but in most cases the input cannot be validated. The reason lies in the subjectivity of the observations of the wear traces. It has been tried to employ as many quantitative variables as possible (like the width of a polish or the number of striations) in order to obtain more objective descriptions, but most variables concern qualitative observations (like the texture of the polish or the degree of edge rounding) and cannot be replaced by a figure.

Concerning the requirement of correctness it can therefore be concluded that, despite the attempts to control the incoming data, it cannot be avoided that in some cases wrong interpretations will be given. It is, for instance, possible that a description which is procedurally correct and therefore allowed by the system, but which does not match the observed traces, leads to an interpretation. In that case the interpretation is procedurally correct but it has nothing to do with that particular implement. This leads to the conclusion that as long as a user is not capable of providing reliable input, the output will not be reliable either. Examples of this will be shown in chapter 7.

This, of course, is not a problem that exclusively pertains to expert systems. Any computer program needs good input. Even human experts cannot give good interpretations if their observations are inaccurate. Perhaps in future versions of WAVES the demand of correctness may be further met by improving the verification of the input. One possibility would be to warn the user when, for instance, his description deviates from known and interpretable wear patterns. Such adaptions, however, will have to be subjected to serious discussions because they may infringe upon the underlying learning processes. For instance, the warning proposed in the above may be instructive from an educational point of view. It may, on the other hand, also invite the user to adapt his description for the sake of obtaining a conclusion.

It is to be expected that an improved guidance of the input yields an increased number of interpretations rather than more reliable interpretations. The latter might be obtained by improving the supervision of the input by means of the computerization of the observation and subsequent description of wear traces. Attempts that have hitherto been made in this direction, such as *pattern recognition*, were discussed in chapter 4 and it is clear that there are no truly satisfactory methods yet. Again, it may be questioned whether such an approach, when successful, would indeed be desirable. It would surely reduce the educational value of the application because it would deprive students of the process of learning to observe.

5.3.3 USER INTERFACE

5.3.3.1 Requirements

A first requirement concerning the application's user interface was the possibility to apply graphics. Since use-wear analysis is a highly visual method (see chapter 4), it is important that the visual aspects are incorporated in the application as well. It ought to include a reference collection of photos of wear traces in order to give users the opportunity to compare their observations with those the application has knowledge of. Obviously, this requirement had serious consequences for the graphical capabilities of the software and hardware that would be employed, because displaying photographic material demands sophisticated graphic facilities.

Another reason for posing high demands on the user interface is the requirement that WAVES has to be highly interactive. Since the application is entirely dependent on the observations of the user, a smooth communication with the user is crucial. This requires a user friendly approach, consisting of nice and inviting looking screens which are easy to understand and work with. Moreover, it would be necessary to include facilities which guide the information input and give additional user support concerning use-wear analysis.

For the credibility and subsequent acceptance of an educational system, it was considered to be of vital importance that the application would also give information on its internal procedures. By giving insight into the way the system applies its knowledge, the user may gain understanding of and confidence in the reasoning processes, which he may eventually start to simulate. This requirement meant that, so-called how- and why-facilities had to be incorporated. On request, a user must be informed on how and why a conclusion is drawn, or why not. Usually, why-facilities are features which keep track of the reasoning process during a session, until a conclusion has been reached. They allow the user to query the system about the reason why certain questions are being asked or about the logic behind a conclusion. This means, for example, that the rules are shown which are active at a particular stage of a reasoning process and which will be activated next if a particular answer is given. A how-facility is a complementary feature which allows the user to ask for the reasoning strategy that has been followed to reach an interpretation. Often this means that an outline is given of all the rules that have been activated during a session.

5.3.3.2 Decisions

In order to meet the first requirement, the incorporation of the visual aspect of the use-wear analysis, it was decided to include a large collection of photos of use-wear traces and other illustrations. Especially the analysis procedure was believed to require many illustrations. In this procedure, they first of all support the description of the wear traces. If a user is uncertain, for instance about the difference between the attributes of a variable, the images may be helpful. Secondly, the illustrations support the interpretations that are obtained in the analysis procedure. The interpretations are accompanied by photos showing the relevant wear patterns of experimental artefacts. In this way, the user can compare these patterns with those of the analyzed implement.

The second requirement, *i.e.* user friendliness, has played an important role during the entire development process of WAVES, because the performance of the system depends on the quality of the interaction with its user. In order to achieve user friendliness three principles were applied.

First, facilities were used to make the interaction as easy as possible. An example of such a facility is the possibility in both the analysis procedure and in the hypothesis validation procedure, to select answers from a predefined list of attributes. This not only allows the user to proceed swiftly through these procedures, since nothing needs to be typed, but it also eliminates the chance of semantically wrong input. The second principle is that all screens have a similar lay-out. For the sake of clarity and in order to get sufficient space for explanatory texts, in both procedures each question has its own screen. In most cases, such a screen contains a question, some explanation about the question, a selection menu to indicate the answer and a photo to illustrate the subject of the question. In the analysis procedure, these screens also contain a button which is connected with a screen that contains background information on that particular question. Throughout the application buttons with a similar function look identical and colours are used in the same, functional manner.

Thirdly, the user-friendly approach consisted of the incorporation of additional user-support. It was expected that the user might need background information on aspects like the recording of the input variables and on the argumentation underlying the interpretations. In order to determine which aspects this support should contain, a small inventory of the needs of the future users was made. For this purpose the expert, future users (students) and other analysts were questionend. Also my own experience, which was acquired during the knowledge elicitation and subsequent data analysis in an attempt to unravel the method's underlying reasoning processes, turned out to be useful. The difficulties that were encountered made clear which aspects of the learning process might need attention. On the basis of this inventory, a collection of background information screens was created.

With reference to the third requirement of the user interface, *i.e.* facilities to explain the procedural aspects of an interpretation, the decisions were also made on the basis of the inventory of the needs of the users. It was decided that the explanatory facilities would not provide the traditional why-facility which show the rules that are active during a session. These rules are in itself not very informative, because they are predominantly of procedural relevance. Especially regarding the educational purpose of the analysis procedure it was believed that laymen on use-wear analysis would be served better with specific information on the purpose of the reasoning process rather than with information on the internal structure of the expert system. Moreover, the individual rules are not as interesting or important as the combination of the rules, *i.e.* as the entire reasoning process. This is, however, not a feature of standard how- and why-facilities.

The reason for not being satisfied with the standard how-facilities is that the procedural aspects of WAVES are not of much interest for the user. These facilities only make sure that the analysis trajectory, which is the same every session, is pursued correctly. Besides, in the analysis procedure of WAVES it is redundant to explain why a certain question within that trajectory is asked at that particular point, because they are asked in a predefined sequence. In the hypothesis validation procedure the order within the sequence is important for the interpretation, but this is not the case with the analysis procedure. It applies a fixed sequence in order to train students to work systematically.

The inventory of the users' needs also made clear that the application ought to offer facilities that comply with the expectations and knowledge level of the users. The why-facility of WAVES therefore consists of background information on the meaning of the questions and on the consequences of the attributes for the interpretation. The how-facility is represented by the possibility to see how the application composed the diagnostic value of the wear pattern that it has analyzed. This not only offers a user the opportunity to validate the conclusions of the application, but also to gain knowledge on the relationship between the observed wear features and conclusions that can be deduced from them, *i.e.* the materials or motions that may have caused them.

5.3.4 HARDWARE AND SOFTWARE

5.3.4.1 Requirements

The requirements on the hard- and software that would be employed to implement WAVES, are primarily dictated by the above mentioned demands concerning the knowledge handling and user interface. First of all, the software had to allow for a combination of inference strategies. Both the backward reasoning procedure (the goal directed way of thinking) and the forward reasoning procedure (the data oriented method) had to be possible. It also had to offer the opportunity to apply and calculate with quantitative data, as the interpretation of the analysis procedure would be accompanied by a diagnostic value.

The above mentioned demands concerning the graphic abilities of the application influenced the selection of the software as well. They required a tool with advanced graphical facilities enabling the use of colours and grey scales together with a photographic resolution. Obviously, this had consequences for the hardware as well, as these facilities require a high quality monitor and graphics card. Other hardware requirements were that both the implementation tool and the final application would be running on a stand alone (IMB-compatible) personal computer (80486 processor) with an average configuration. The reason was that it ought to be possible to distribute the application among other analysts or universities and that it can be run on several computers simultaneously. At the time the project was started few institutes had more than one large computer available for other purposes than supporting Local Area Network facilities, whereas they all had ordinary personal computers.

5.3.4.2 Decisions

Because of the fact that several technical requirements concerning the software could hardly be determined in advance, it was decided to first build a small prototype. Prototyping is a useful means to gain insight into the complexity of representing the knowledge and inferencing processes and to validate design decisions. It could furthermore make clear what requirements should be posed on the final implementation tool. Another advantage of a prototype is that it can show the expert the abilities and restrictions of the application under construction, which gives the opportunity to make adaptions to the knowledge model or other aspects of the design.

The prototype was built by means of the programming language PROLOG, since this was considered to be "*an excellent tool for prototyping*" (Bratko 1989: 71). During the development of this prototype it soon became clear that it would be better to use a well-equipped expert system shell rather than a programming language for the final implementation of WAVES.[4] A shell would offer more possibilities to use advanced graphical features, and its built-in facilities to implement complicated reasoning processes would facilitate and accelerate the development process (see chapter 3.5).

After it was decided to employ a shell, it was not very easy to find one that could meet all demands. At the start of the WAVES-project especially the development towards the use of graphics in expert system building tools was only just beginning. Based on a careful comparison of the abilities of shells that were available at that time, eventually the expert system shell called *Level5 Object* (version 2.5) was selected.[5] It seemed that this shell would offer most of the representational and inferencing facilities that are commonly used for artificial intelligence applications (*cf*. Payne & McArthur 1990: 376-383). It is a user-friendly package with which one can relatively easy learn to work with because it offers all kinds of tools that support the application building process.

Another important aspect of this shell is that it is entirely Windows[6]-oriented, which enables the incorporation of graphics, including photos (.BMP format). Moreover, Level5 Object, and the applications that are built with it, runs on an ordinary stand alone personal computer with a 80386 or 80486 processor (it only requires two megabyte of internal memory and a minimum of four megabyte of hard disk space). [7,8,9] Given all its characteristics, this tool seemed to suit the purpose of this case study.

The only drawback of Level5 Object is that one has to buy a separate software environment that allows you to operate an executable file, *i.e*. the final application. This means that WAVES cannot run without this software package supporting it. Since for each copy of this package high runtime royalties are charged — at least for archaeological standards — it is financially less attractive to purchase WAVES. Consequently this does not stimulate the distribution of the application. This drawback of Level5 Object, however, was believed to be counterbalanced by its positive qualities.

As for the implementation of photographic material of use-wear traces, it was decided to use high quality photos that already existed of the wear patterns in the reference collection.[10] This was very convenient since these photos showed exactly the wear patterns on which the rules in the knowledge bases were based.

For the actual implementation of the knowledge into the application a combination of the traditional practice of *linear programming* and the *rapid prototyping* approach was used. The former means that an application is fully programmed on the basis of a detailed knowledge model before it is tested and the latter means that a trial-and-error method is applied to establish the required functionality. The implementation of WAVES was based on a rather detailed knowledge model, but all tasks were tested by means of a prototype before the application was validated as a whole. The first step was the implementation of the analysis procedure. Only when this was fully operational, the hypothesis validation procedure followed. This could be built much quicker due to the experience that was gained in building the first part. Moreover, the photographic material that had to be gathered for the analysis procedure could be used for the hypothesis validation procedure too.

The approach that was followed during the implementation was that the modules, into which the knowledge of the four wear categories was divided, were appended separately. First, the procedural aspects of the model were constructed in outline. Then the procedural handling of each wear category was tested with small bits of knowledge, evaluated and refined. Subsequently the connection between the modules was established and the knowledge base was filled with the remaining bulk of decision rules. Finally the functionality and performance of the whole application was tested and improved.

The advantage of this approach is that it is based on a worked-out knowledge model while the validation of the modules provides the opportunity to carry out small adjustments. The worked-out knowledge model enables the expert to discover hiatuses prior to its implementation and avoids that the engineer encounters large and unexpected problems during its implementation. Moreover, the systematic validation of the implemented modules avoids that fundamental mistakes or large biases are discovered during the evaluation of an application as a whole and that large parts or even the complete application needs to be revised radically.

Not just in theory, but also in practice the described approach turned out to be convenient and the prototypes were fairly quickly developed. It nevertheless required much effort to transform this prototype into an acceptable operational system, which performs on an expert level, is user friendly and 'fool' proof. Especially the optimisation of the user interface was very time-consuming.

5.4 Discussion

5.4.1 INTRODUCTION

Some decisions that were made during the development process of WAVES have clearly put a mark on its functionality as a whole. Especially decisions that concern its knowledge are fundamental. Of these, the use of experimentally obtained information as the basic source of knowledge and the handling of the interpretational uncertainties by means of diagnostic values are the most important and perhaps the most disputable. Although these decisions were already discussed briefly, they may need a more detailed review in order to understand why they were made, what consequences they have and what alternatives might be chosen in future.

5.4.2 ON USING KNOWLEDGE DERIVED FROM EXPERIMENTS

In paragraph 5.2.2 it was explained why a large part of the knowledge of WAVES has been derived from the results of an experimentally obtained reference collection rather than from human experts. It must be stressed, however, that apart from advantages this has limitations as well. The most important limitation is that it cannot compete with expert knowledge. Since the latter is a combination of different kinds of knowledge, like rules of thumb, experience, creativity and imagination, this is of a totally different quality. Therefore, the reference collection can only provide basic information.

A second limitation is that the scope of knowledge that experiments can cover is restricted. For instance, some wear patterns that experts know from archaeological contexts, have never been replicated. The application could not interpret such patterns if it would be based on purely objectively obtained knowledge. Moreover, the experimental program does not yield knowledge on traces that are caused or obliterated by burning, or by post-depositional processes, such as patination.

In short, by using experiments only, the added value of the human expert, *i.e.* his or her expertise, would not be used, while it is beyond doubt that the application cannot achieve optimal results if it has not been supplemented with expert knowledge. But, especially in relation to the analysis procedure, the adding of subjective knowledge raised some dilemmas. First of all, it was difficult to decide on the height of the subjectively appointed diagnostic values of the wear traces. Since there was no quantitative base, they had to be estimated by the expert. Furthermore, it was wondered whether in the interpretation these estimated values should contribute equal to the calculated values, more or less.

In order to include some expertise as well, WAVES was indeed equipped with *subjective decision rules*, although with some reservation. They are not abundant and have only been used for cases for which no experimental data are available and for procedural purposes. It was also decided to treat the diagnostic values that were established subjectively similar to the values that were established by calculation. The user is being informed, however, as to what part of the interpretation is based on what type of knowledge: the calculated values are called 'objective' and the estimated values 'subjective'.[11] This offers the user the opportunity to validate the interpretation and to adjust it according to what he or she thinks is decisive.

5.4.3 ON MANAGING UNCERTAINTY

In WAVES uncertainty could not be handled by means of probabilities. As an alternative it has been tried to give an impression of the diagnostic value of a wear pattern. This approach, however, has two implications that the user needs to be aware of. First of all, the fact that the diagnostic value of a wear pattern as a whole is determined by the frequencies of the individual wear features implies that a pattern that occurs most frequently in the experimental program does not have to get the highest diagnostic value. An example of this could be seen in figure 14. The first two combinations of wear attributes (A1/B2/C1 and A2/B1/C1) did not occur equally frequent (three and four times) but received an identical diagnostic value (160 and 160). It is even possible that a pattern that does not occur most frequently, is the most diagnostic. An example of this is shown in figure 15 (figures in bold): the attribute combination A2/B2 occurs four times, but is less diagnostic than the combination A2/B1 (100 versus 130), which occurs only three times. In other words, the most frequently observed pattern is not by definition the most characteristic for that particular material.

The second implication of this method concerns the possibility that it offers to interpret wear patterns that have not been established experimentally but which may occur archaeologically. In paragraph 5.2.4 it was discussed that this flexibility was one of the reasons for applying this approach. This possibility presupposes, however, that all experimentally non-observed combinations of attributes can indeed occur in the archaeological context. This may be unrealistic, but there is no way to demonstrate that particular combinations are absolute nonsense as there are simply too many potential combinations to check them all. Moreover this check could only be based on an expert's experience. In my opinion it is allowed to use this approach for as long as the presumption has not been invalidated. Another important aspect of the interpretation of the unknown patterns is that it is based on the same decision rules and diagnostic values as by which known patterns are interpreted. Consequently, the above shown consequences count for these cases as well. This implies that a wear pattern which has never been observed before, but which can be interpreted on the basis of its well-known features, may be more diagnostic than those that have been established experimentally. An example of this can be seen in figure 14 (step 3). The diagnostic value of the unknown wear pattern A1/B1/C1 (160) is higher than the value of some of the replicated patterns.

Observed wear patterns	number of occurrence
A1 / B1	1
A2 / B2	4
A2 / B1	**3**
A3 / B1	2

(number of experiments = 10)

Step 1. Deduction of the frequency of occurrence of single attributes:

A1 has been observed in 1 case = 10%
A2 has been observed in 7 cases (4+3) = 70%
A3 has been observed in 2 cases = 20%
B1 has been observed in 6 cases (1+3+2) = 60%
B2 has been observed in 4 cases = 40%

Step 2. Calculation of the diagnostic value of the observed patterns

A1 / B1 = 10 + 60 = 70
A2 / B2 = 70 + 40 = 110
A2 / B1 = 70 + 60 = 130
A3 / B1 = 20 + 60 = 80

Fig. 15. Consequences of the uncertainty handling approach in WAVES: when focusing on the occurrence of single attributes, it is possible that combinations of attributes (patterns) that occur less frequently (figures in bold) may have a higher diagnostic value than more frequently observed patterns.

Eventually, this approach may turn out to be not the most desirable. But, since this is the first time that uncertainty in use-wear analysis has been handled in this manner, it needs to be experienced in practice first. In case it does need some adjustment, one possibility would be to assign a bonus value to combinations that have been experienced most, or that have been experienced experimentally. The main argument, however, that this adaption was not employed right away is that it would imply that experimentally obtained patterns are more reliable and more diagnostic than wear patterns from archaeological contexts. It must also be kept in mind that the wear patterns which were deduced from experiments, and which are therefore considered to be reliable, have not been objectively determined. Since these patterns were described by human analysts, they are composed on subjective grounds. If the analysts had used a slightly differently description for one of the attributes, this would have yielded other patterns. This raises the question whether these patterns deserve a bonus while they partly originate from coincidence.
Another argument for not giving bonus values to the experimentally obtained patterns is that the experimental program was unbalanced. There is a large discrepancy in the number of experiments with single contact materials. Bone, for instance, was studied best. It was subjected to no less than 53 experiments. Teeth and stone, on the other hand, are represented least: they were only used in one experiment. Assigning bonus values for observed patterns would therefore favour the interpretation of materials that were studied best and discriminate those that were studied less.

part two: WAVES in action

5.5 The composition of the analysis procedure

The goal of the wear analysing procedure is to guide students in interpreting the function of flint stone tools on the basis of a description of wear traces (see paragraph 5.1). The internal affairs of this procedure, *i.e.* the representation of the conceptual and procedural knowledge, the applied reasoning processes and the composition of the interface for communication with the user will be discussed in this paragraph.
The knowledge that had to be implemented for this procedure consisted predominantly of a list of wear features (variables and attributes), a list of materials and motions that cause these features and the relations between them. For this the lists of attributes and diagnostic values were used that were deduced from the analysis of the experimentally obtained data and the knowledge modelling (appendix I). The variables and attributes as well as the materials and motions were represented by means of hierarchical structured elements, called *Object-Attribute-Value-triplets* (fig. 16). The hierarchy

Object:	Attributes:	Value:
retouch distribution	overlapping	TRUE/FALSE
	clumped	TRUE/FALSE
	etc.	
polish topography	cratered	TRUE/FALSE
	pitted	TRUE/FALSE
	etc.	
soft wood	score on retouch distribution	numeric value
	score on polish topography	numeric value
	etc.	
Longitudinal motion	score on retouch distribution	numeric value
	score on polish topography	numeric value
	etc.	

Fig. 16. Example of the way in which WAVES variables, attributes and their values, materials and motions are represented by means of object-attribute-value triplets.

```
Rule name
IF attribute x OF object y IS value z
THEN attribute z OF object x RECEIVES value y

Rule on longitudinal motion 1
IF overlapping OF retouch distribution is TRUE
THEN retouch distribution OF longitudinal motion RECEIVES diagnostic value 40

Rule on longitudinal motion 2
IF overlapping OF retouch distribution is TRUE
THEN retouch distribution OF transverse motion RECEIVES diagnostic value 20
```

Fig. 17. The format of a decision rule. Several rules can be connected to one wear feature.

means that the objects consist of attributes and that the attributes have a value. Such a value can be predefined (True, False or a digit) or variable, like a word or a sentence. It is assigned by the application, as a result of a reasoning process, or by the user as input information. In this case, the variables, materials and motions are the *objects*, the attributes of the variables are their *attributes*. At the start of a session of the analysis procedure, the initial value of all attributes is FALSE or zero (in case of numeric attributes). If the user subsequently indicates that, for instance, the use retouch on an implement is distributed in an overlapping manner, then the value of the attribute ‘overlapping’ changes into TRUE. In order to interpret the features which are described by the user, they have to be linked to a material and motion. In expert system terms this means that the activated attributes (their value is TRUE) must be connected with the objects that represent the materials and motions. This requires a reasoning method. In WAVES this is carried out by IF-THEN decision rules (fig. 17).[12] Each wear feature that is described by the user activates the decision rules that are connected with it. In the example in figure 17 the value of the attribute ‘overlapping’ is changed into TRUE and both rules that relate to this attribute are activated. Since most wear features are diagnostic for several materials and motions, many are connected to several rules. In this case, the overlapping retouch is indicative for both a longitudinal and a transverse motion and they both receive a score.
The strategy by which the rules are applied in the analysis procedure is data-oriented (or forward directed): the input of the user, *i.e.* the description of the observed wear-character-

Observed features of a polish:

DISTRIBUTION:	BRIGHTNESS & TEXTURE:	TOPOGRAPHY:
isolated spots	very bright & smooth & matt	pitted

Associated diagnostic value:

soaked bone	**80**	siliceous plants	80	**soaked bone**	**50**
hard wood	60	cereals	80	**soaked antler**	**30**
soaked antler	**50**	**soaked antler**	**80**	hide	20
meat and fish	40	hard wood	50	leather	20
leather	30	soft wood	50		
hide	30	**soaked bone**	**40**		
soft wood	20	polish '23'	20		
fish scales	20				
siliceous plants	10				
cereals	10				
shell	10				
dry antler	10				
pottery	10				
dry clay	10				
stone	10				
tooth	10				

Final interpretation and total score (= diagnostic value):

soaked bone	170 (=80+40+50)
soaked antler	160 (=50+80+30)

Fig. 18. Example of how WAVES deduces the diagnostic value of the observed wear patterns. Exclusively the materials are included in the interpretation for which all wear features are diagnostic.

istics, determines the direction of the reasoning process and the outcome of a reasoning process, *i.e.* the interpretation. The combination of all activated rules determines the contents of the conclusion.

When all descriptions have been collected and all related rules have been activated, all the diagnostic values of the wear features that relate to a particular material or motion are summed. This final score, which represents the diagnostic value of the wear pattern, will be presented to the user. In figure 18 an example is given of how WAVES calculates the final diagnostic values. In this example the wear traces may have been caused by soaked bone or soaked antler. Some of the observed polish features were also diagnostic for other materials, like the isolated spots for hard wood, but these materials were excluded from the interpretation on the basis of other, non-fitting features. In the case of hard wood, for example, the pitted topography did not make sence. It may be noted that, since only materials and motions which relate to all features are included in the interpretation, the order in which the features are described is irrelevant. It does not make any difference whether a material is excluded on the basis of the feature that is described first or last.

For practical purposes it has been decided that most rules contain only one conclusion, *i.e.* each assigns a diagnostic value to only one material or motion. The reason is that most rules contain criteria on the IF-side which usually concern a particular material instead of all materials. An example of such a criterion is whether the previously described wear feature was diagnostic for a particular material, *i.e.* whether evidence has been found that that particular material has caused the wear pattern. If this is not the case, the THEN-side of that rule will not be carried out: this material is excluded from the interpretation. If this rule should also assign scores to other materials, this would fail due to its abortion. Obviously, this is undesirable.

It has been put forward (paragraph 5.2.3) that in some cases combinations of wear features are far more diagnostic than individual features. Since this could enhance the interpretation of the analysis procedure considerably, it has been included in the reasoning process. The termination of the edge removals is only assessed in combination with the degree of edge rounding and the texture of the polish with its brightness. Only particular combinations of wear features are diagnostic and, therefore, rewarded with a score. It may seem that this approach conflicts with the principle of optimal flexibility, since it does not permit unknown combinations. However, it was only applied in the two cases in which all possible attribute combinations of the involved variables were included.

The rules assign either an *objective* or a *subjective* value to a material or motion. An objective diagnostic value has been determined by means of the quantitative analysis of the experiments, a subjective value has been estimated on the basis of the knowledge that was elicited from the expert. In the decision rules, these two types of values are intentionally handled separately in order to keep a clear distinction between the aspects of the interpretations that are based on the different kinds of knowledge. The final diagnostic value of the wear pattern is the sum of the two scores, but it is indicated to the user which part of it has been deduced subjectively and which part objectively.

The analysis procedure enables separate analyses and interpretations of macro traces (retouch and rounding) and of the micro traces (polish and striations). The reason for this division is that in many cases the wear-categories are not simultaneously present or equally diagnostic (Van Gijn 1989; Van den Dries & Van Gijn, in press). In such cases it ought to be possible to get an interpretation of either one of them. An advantageous side-effect of a separated analysis of both categories is that two independent interpretations are obtained, which can either confirm each other or act as an extra control. Two very different interpretations can be a warning that due to whatever reason the interpretation is not very reliable.

In outline, 20 materials and 6 motions have been included in the analysis procedure[13]; it consists of 33 question screens, of 35 screens with additional information and of 7 screens on which the analyses results are displayed; it contains 209 illustrations, mostly photos; 372 rules handle the conceptual knowledge and 176 rules take care of the procedural aspects.

5.6 The composition of the hypothesis validation procedure

The hypothesis validating procedure can be used by more experienced students or by analysts for the purpose of evaluating their interpretations (see paragraph 5.1). Many aspects of this module resemble that of the analysis procedure. They use the same basic knowledge, though in opposite ways, they use the same knowledge representational method, *i.e.* a combination of rules and objects, and they make use of the same photographic material. Nevertheless, there are some major differences as well. This procedure has, for instance, not been designed to cover a broad range of relations between wear features and materials and motions, but to validate merely ideal types of wear. Validation comes down to verifying the presence of diagnostic features: a hypothesis is only confirmed when the use wear on an implement comprises all the features that the application expects. Consequently, this approach is less tolerant than the approach of the analysis procedure, which even interprets wear patterns that are not very diagnostic. This may seem very inflexible. However, when a feature is not present the application first verifies the presence of alternative features before a hypothesis is rejected (fig. 19).

```
RULE for bone 1

IF goal OF analysing process = “bone”
AND present OF use retouch = TRUE
AND step and hinge equal OF retouch termination = TRUE
OR step and feather equal OF retouch termination = TRUE
AND medium OF retouch width = TRUE
OR large OF retouch width = TRUE
AND close distribution OF retouch distribution = TRUE
OR clumped distribution OF retouch distribution = TRUE
AND isolated spots OF polish distribution = TRUE
OR thin line along edge OF polish distribution = TRUE
OR band along the edge OF polish distribution = TRUE
OR spread OF polish distribution = TRUE
AND very bright OF polish brightness = TRUE
OR bright OF polish brightness = TRUE
AND smooth and matt OF polish texture = TRUE
OR rough and matt OF polish texture = TRUE
AND comet tails OF polish topography = TRUE
OR pitted OF polish topography = TRUE
OR flat OF polish topography = TRUE
AND retouch exceeds polish OF invasiveness = TRUE
OR polish and retouch equal OF invasiveness = TRUE
THEN soaked bone OF identified material := TRUE
AND remark OF analysing process := “traces are probably caused
      by a longitudinal motion”
AND result OF hypothesis check IS positive
```

Fig. 19. Examples of a rule that handles the knowledge that is employed to validate the hypothesis that an implement was used for bone working.

Another difference concerns the way the knowledge is handled. Similar to the analysis procedure, the variables and their attributes have been represented by means of ‘Object-Attribute-Value-triplets’. The materials and motions, however, are not included at all. In this procedure, only the *criteria* have been represented that need to be met in order to permit an interpretation in favour of a contact material. These criteria are handled by IF-THEN decision rules (fig.19, appendix II). Furthermore, a goal-oriented reasoning strategy is applied rather than a data-oriented strategy. This means that the application reasons in a backward direction: it starts with a goal (the hypothesis) and subsequently searches for evidence that supports this goal. In the above case, the goal is set to ‘bone’ as the user indicated this as his hypothesis. Then the rules are activated which contain the criteria for its confirmation. If one of the conditions of the first rule cannot be met, *i.e.* if a feature is not present, the rule fails and — if available — another rule will be tried. The selection of another rule is based on the knowledge or facts that were already collected for the first rule. If these facts indicate that one of the criteria of the next rule cannot be met either, this rule is aborted as well and the application will search for an alternative rule. This continues until the known facts meet all criteria of the alternative rule. Only from that moment, the user may be asked for additional information. The validations process finishes when all conditions of one of the rules have been confirmed or when none of the rules was successful.

One of the advantages of this approach of verifying the applicability of alternative rules by means of the already collected information, is that the user does not have to answer the same question twice. He is not even bothered by questions of which the answer can be deduced from the answers that were given already. If, for example, the user has indicated that the use retouch on a tool is distributed in a regular manner, the application concludes that all other retouch distribution types are not applicable. In case an alternative decision rule is employed, this information is used and the user is not questioned about the other retouch distribution types. An additional advantage is that it quickens the validation process, which, in most cases will be pleasant for the more experienced user. A disadvantage, however, is that in some cases the user might lose track of the reasoning process as it may not always be clear why a particular line of reasoning is followed instead of another.

The contents of the decision rules have not been based on a particular hypothesis validation method that the expert applies, but have primarily been deduced from the experimentally obtained data. Obviously, the resulting lines of reasoning have been verified by the expert. In comparison with the rules of the analysis procedure, they are more straightforward. Uncertainties are not taken into account; a wear feature is either present or not.

The number of features that has been used to validate a hypothesis varies from one material to another. For instance, implements that have presumably been used on soaked bone are verified on at least nine variables (see figure 19), whereas others may be validated on just five variables. This depends on the amount of knowledge that is available on the diagnostic features of the various materials. Those that are known best, because they were subjected to many experiments, will be checked in more detail during the validation process than others. This implies that these validations may be more reliable than those that concern materials and motions of which is known less.

At the moment, the procedure is able to validate the hypotheses concerning 22 materials. These are the 20 materials that were included in the experimental program, supplemented with polish ‘23’ (Van Gijn 1990: 85) and another type of polish of which the origine has not been identified yet, called polish ‘10’ (Schreurs 1992: 147). Furthermore, the verification of a hypothesis does not yet involve motions. In outline, the hypothesis validation procedure consists of 65 question screens, of 27 screens on which photos of wear patterns of different materials are shown, and of one screen on which the result of the validation is displayed. It contains approximately 150 illustrations, almost exclusively photos. Furthermore, 53 rules handle the conceptual knowledge and 96 rules take care of the procedural aspects.

5.7 A session

5.7.1 GETTING STARTED

WAVES runs on an ordinary personal computer.[14] The entire application consists of three knowledge bases. One gives an introduction to the two procedures and subsequently manages their activation and closes the application, one contains the analysis procedure, and one comprises the validation procedure. When the application is started it initiates the first knowledge base and introduces the two procedures. Subsequently the user can select the procedure he wants to utilise. As the procedures run completely independent of each other, only the selected one will be activated. After each session with that procedure is finished, the user may opt for another session, or for a session with the other procedure, or may close the application (fig. 20).

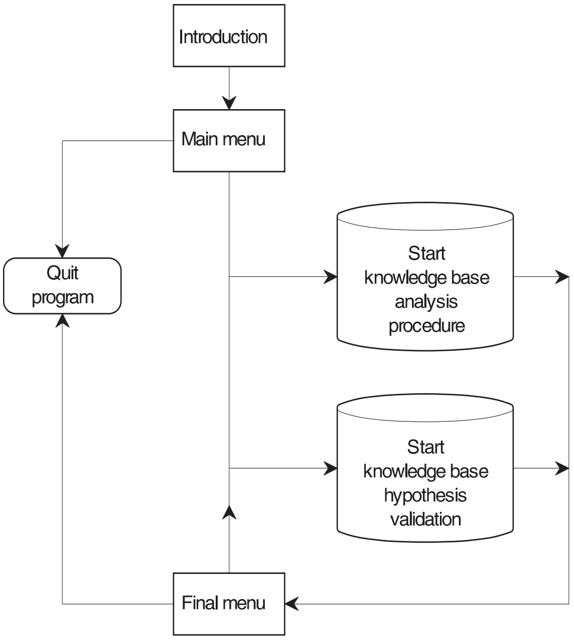


Fig. 20. Main menu of WAVES

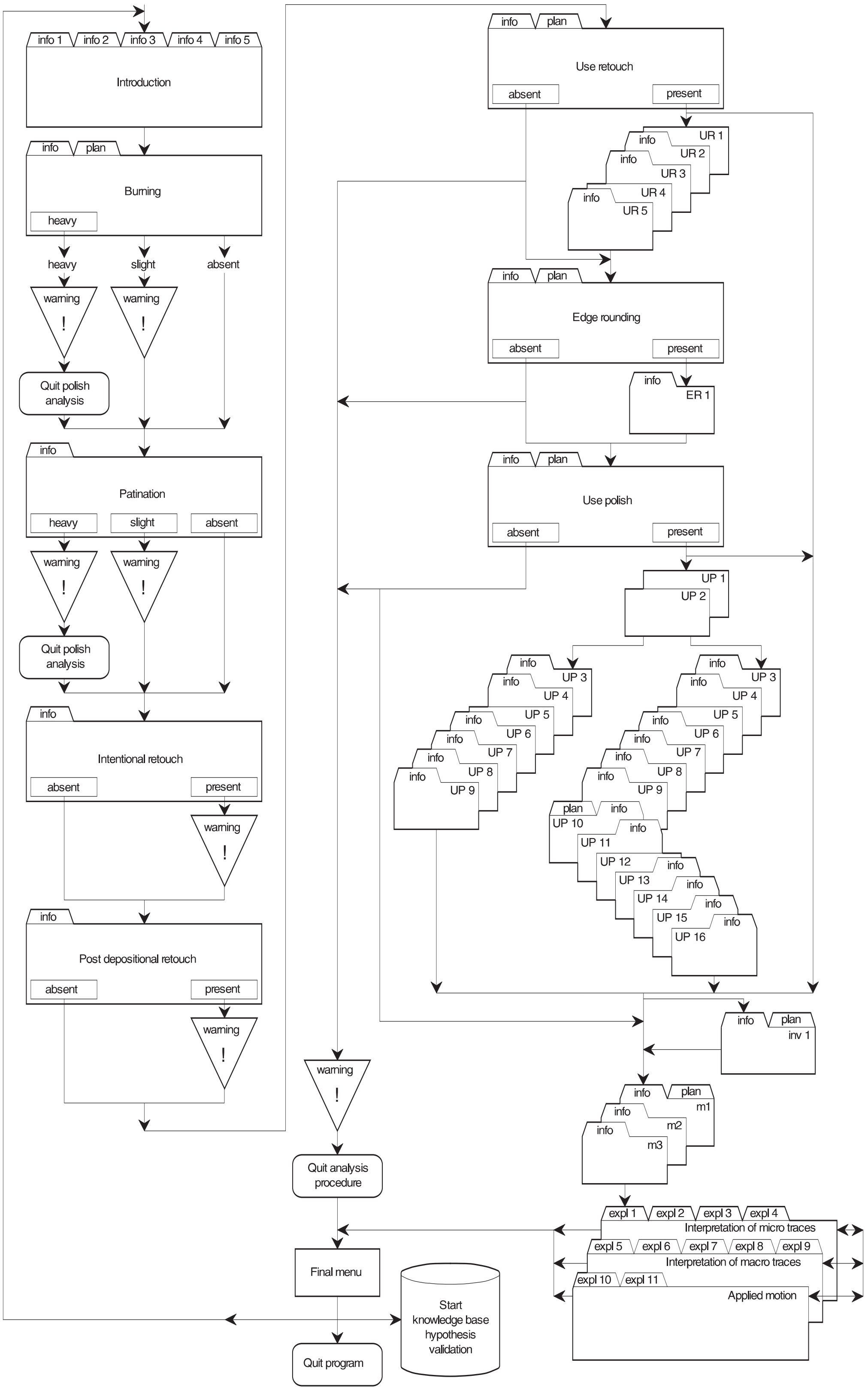
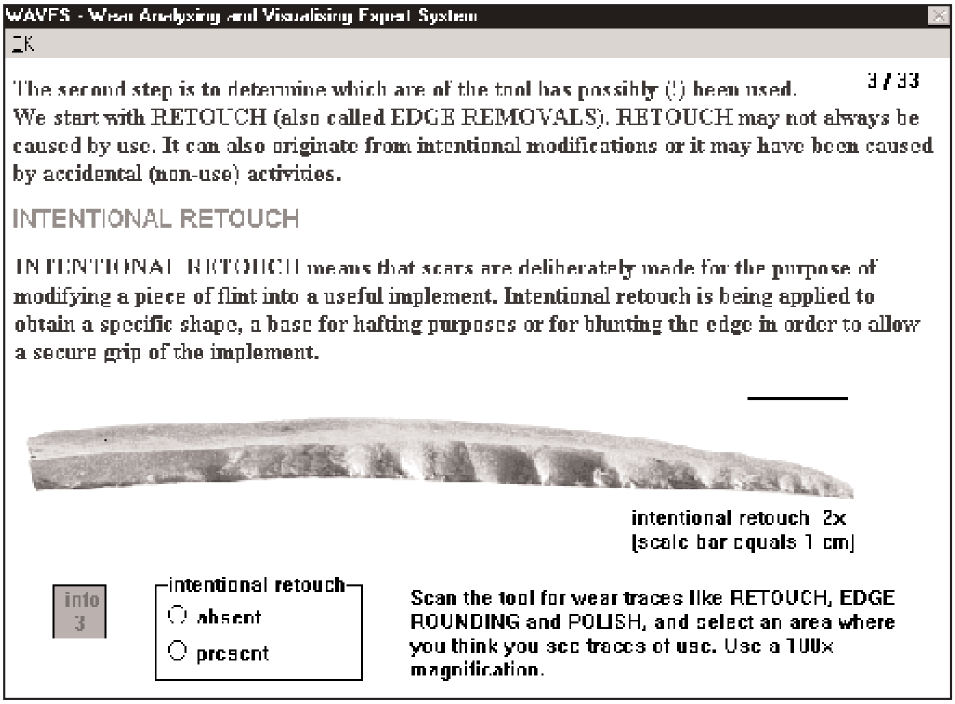


Fig. 21. System scheme of the analysis procedure of WAVES. UR = use retouch, ER = edge rounding, UP = use polish, inv. = invasiveness, M = morphology, expl. = explanation screen



The second step is to determine which are of the tool has possibly (!) been used. We start with RETOUCH (also called EDGE REMOVALS). RETOUCH may not always be caused by use. It can also originate from intentional modifications or it may have been caused by accidental (non-use) activities.

INTENTIONAL RETOUCH

INTENTIONAL RETOUCH means that scars are deliberately made for the purpose of modifying a piece of flint into a useful implement. Intentional retouch is being applied to obtain a specific shape, a base for hafting purposes or for blunting the edge in order to allow a secure grip of the implement.

intentional retouch 2x
[scale bar equals 1 cm]

intentional retouch
○ absent
○ present

Scan the tool for wear traces like RETOUCH, EDGE ROUNDING and POLISH, and select an area where you think you see traces of use. Use a 100x magnification.

Fig. 22. Question screen handling intentional retouch.

5.7.2 RUNNING THE ANALYSIS PROCEDURE

The analysing part of the expert system gives direction to the process of use-wear analysis and interprets the wear traces which are observed by the user. The process of analysis is executed according to a specified trajectory. The reason for this is twofold: it not only forces analysts to work consistently, but enables a comparison of the analysis results of one analyst as well as of several analysts. In figure 21 the flow chart of this procedure is shown.

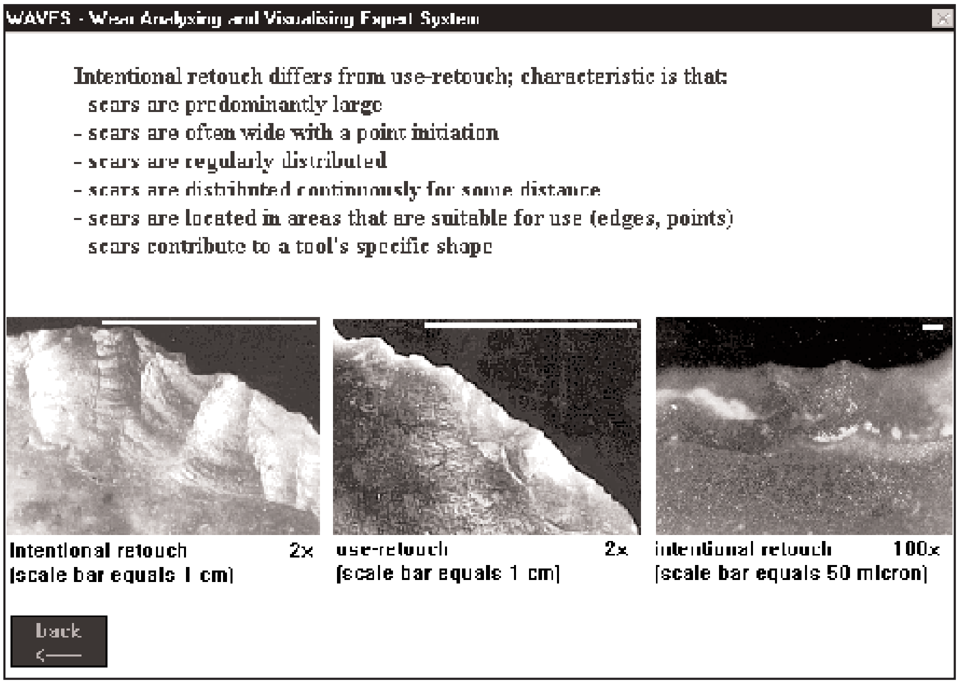
It starts by giving a general introduction on the program and its subject. The user can get information on how the application works and on the meaning of the buttons. Moreover it provides a list of references for further reading, an instruction for using a microscope, and information on the procedures to clean a flint implement prior to its microscopic analysis.

Prior to the actual analysis, it is verified whether a tool and its wear traces are analysable. The user is asked to describe the condition of the implement. If it is heavily burned or patinated or otherwise damaged by post-depositional processes, the user is advised not to proceed with the analysis. Subsequently, the user is warned not to confuse intentional retouch with retouch that originates from use (fig. 22, fig. 23). If eventually the implement appears to be suitable for analysis, the actual analysis starts.

As was said, the wear categories are analyzed separately. This yields two independent interpretations: one on the basis of the macro traces (retouch and rounding) and one on the micro traces (polish and striations). At Leiden University the wear categories are observed by means of different approaches: edge removals by low magnifications, edge-rounding, polish and striations by high magnifications (see also chapter 4.4.3). Hence, the wear analysing procedure of WAVES has been designed in accordance with this method. For the observation of the retouch the user is advised to utilise a stereo-microscope with magnifications up to 150×, for the observation of the other features an incident-light microscope with magnifications of 100× - 400× for the high-power method.

The user is not obliged to pursue the entire trajectory and to analyze all wear categories. If, for instance, one of them is not present, the part that relates to it can be skipped. Or in case a user is only interested in the interpretation of only one category of wear, this is possible as well. It must nevertheless be stressed that in such a case the amount of evidence for an interpretation decreases. It is therefore recommended to include as much information as possible.

The first wear category that is analyzed is use retouch. The features that are examined of the scars are: their location on the implement, the character of their distribution and termination, their average length and the main direction of their orientation. The required information on these features is obtained by an interactive game of questions-and-answers. This means that the system poses the questions and the user makes the observations and provides the system with the



Intentional retouch differs from use-retouch; characteristic is that:

scars are predominantly large
- scars are often wide with a point initiation
- scars are regularly distributed
- scars are distributed continuously for some distance
- scars are located in areas that are suitable for use (edges, points)

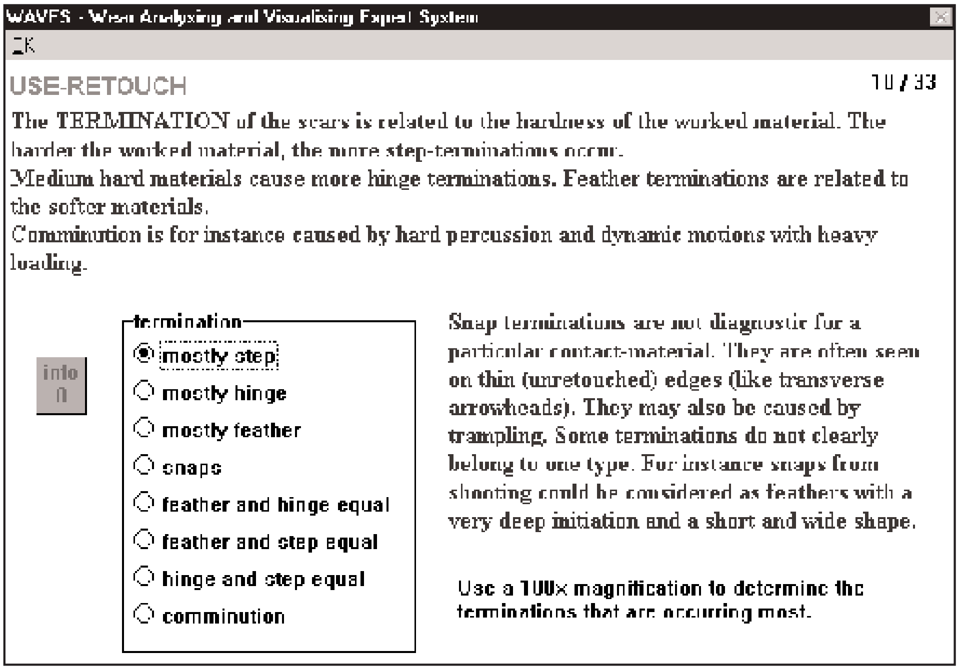
scars contribute to a tool's specific shape

intentional retouch 2x
[scale bar equals 1 cm]

use-retouch 2x
[scale bar equals 1 cm]

intentional retouch 100x
[scale bar equals 50 micron]

Fig. 23. Background information concerning intentional retouch.



USE-RETOUCH 10 / 33

The TERMINATION of the scars is related to the hardness of the worked material. The harder the worked material, the more step-terminations occur.
Medium hard materials cause more hinge terminations. Feather terminations are related to the softer materials.
Comminution is for instance caused by hard percussion and dynamic motions with heavy loading.

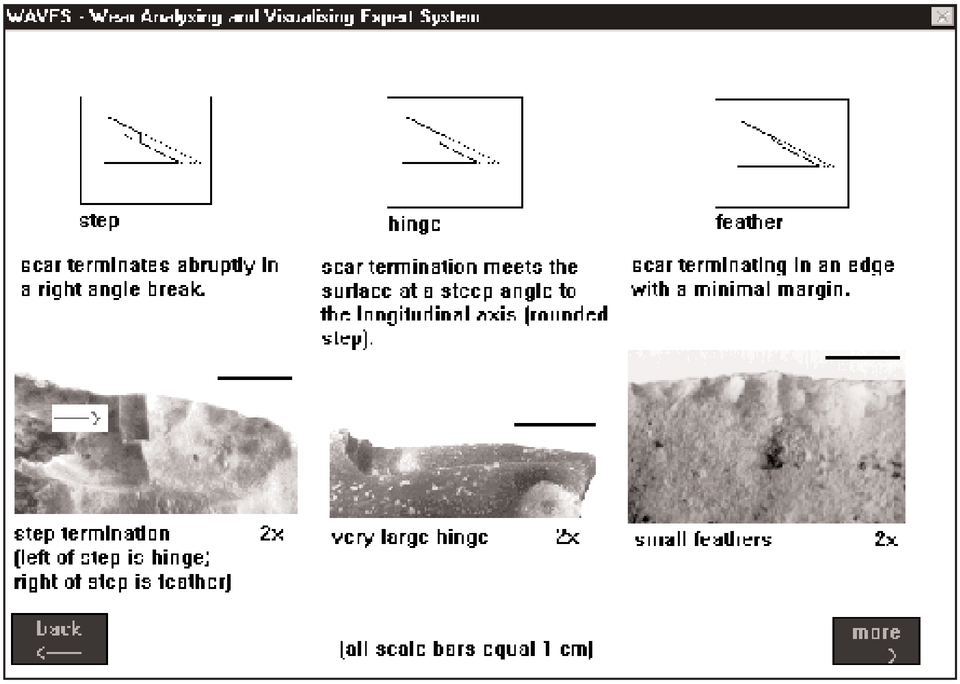
termination
- ● mostly step
- ○ mostly hinge
- ○ mostly feather
- ○ snaps
- ○ feather and hinge equal
- ○ feather and step equal
- ○ hinge and step equal
- ○ comminution

Snap terminations are not diagnostic for a particular contact-material. They are often seen on thin (unretouched) edges (like transverse arrowheads). They may also be caused by trampling. Some terminations do not clearly belong to one type. For instance snaps from shooting could be considered as feathers with a very deep initiation and a short and wide shape.

Use a 100x magnification to determine the terminations that are occurring most.

Fig. 24. Question screen for a retouch variable.

required answers. Since in this approach the user's role is crucial, considerable attention has been payed to guide the process of information input. The user is, for instance, advised on the type of magnification that should be used to answer a particular question. Moreover, the user is provided with predefined lists of wear attributes from which he only has to select the one that resembles his observation best (fig. 24, fig. 26).
This not only eliminates the chance of a semantically wrong input, but it also guarantees that descriptions which are



step

scar terminates abruptly in a right angle break.

hinge

scar termination meets the surface at a steep angle to the longitudinal axis (rounded step).

feather

scar terminating in an edge with a minimal margin.

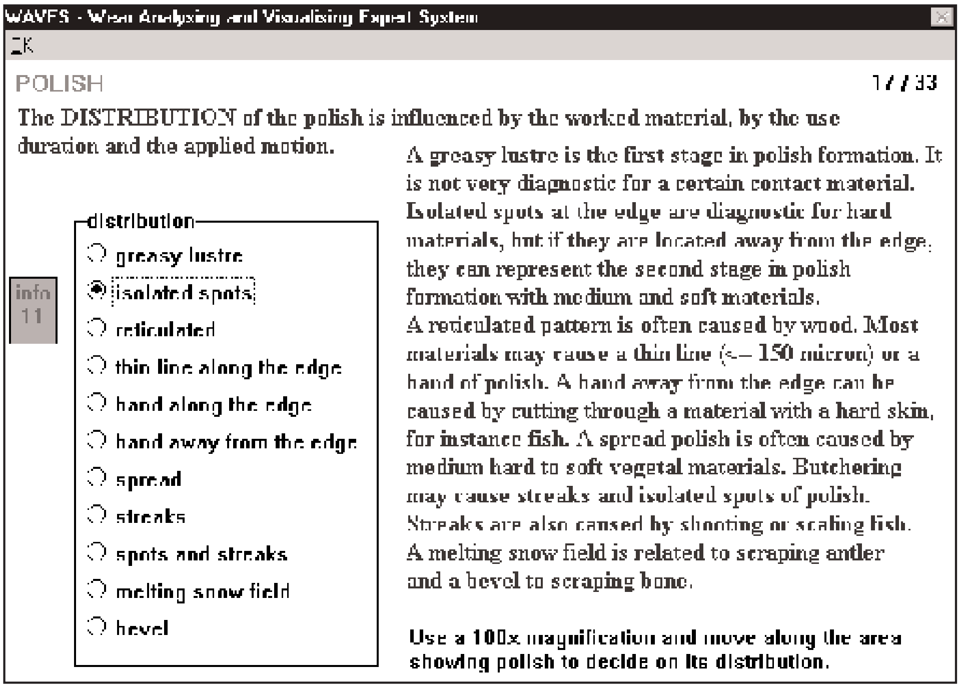
step termination 2x
(left of step is hinge; right of step is feather)

very large hinge 2x

small feathers 2x

(all scale bars equal 1 cm)

Fig. 25. Background information screen for the attributes of a retouch variable.



POLISH 17 / 33

The DISTRIBUTION of the polish is influenced by the worked material, by the use duration and the applied motion.

distribution
- ○ greasy lustre
- ● isolated spots
- ○ reticulated
- ○ thin line along the edge
- ○ band along the edge
- ○ band away from the edge
- ○ spread
- ○ streaks
- ○ spots and streaks
- ○ melting snow field
- ○ bevel

A greasy lustre is the first stage in polish formation. It is not very diagnostic for a certain contact material. Isolated spots at the edge are diagnostic for hard materials, but if they are located away from the edge, they can represent the second stage in polish formation with medium and soft materials. A reticulated pattern is often caused by wood. Most materials may cause a thin line (<= 150 micron) or a band of polish. A band away from the edge can be caused by cutting through a material with a hard skin, for instance fish. A spread polish is often caused by medium hard to soft vegetal materials. Butchering may cause streaks and isolated spots of polish. Streaks are also caused by shooting or scaling fish. A melting snow field is related to scraping antler and a bevel to scraping bone.

Use a 100x magnification and move along the area showing polish to decide on its distribution.

Fig. 26. Question screen for a polish variable.

given by different users, are comparable. Furthermore, this approach is user friendly because the user does not have to type anything. The entire application is operated by mouse. Another aspect of the guidance of the user is that explanations are given about the meaning of a variable. And, on demand, background information is given on the differences between the attributes. Especially for apprentices the terms that are being used may not be sufficiently meaningful. This background information consists predominantly of photos illustrating the different attributes (fig. 25, fig. 27).

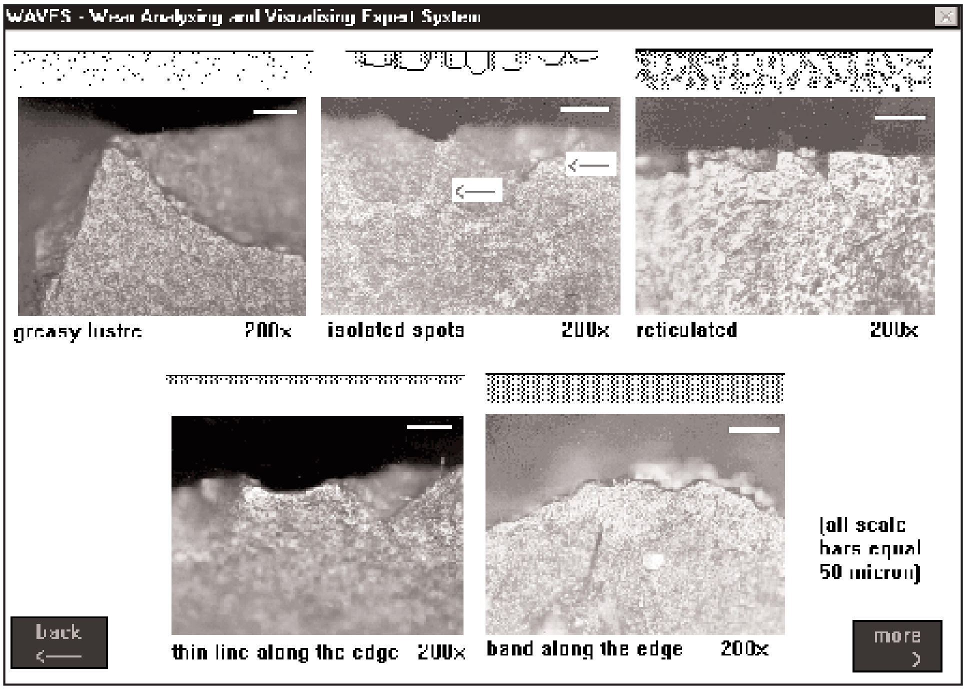
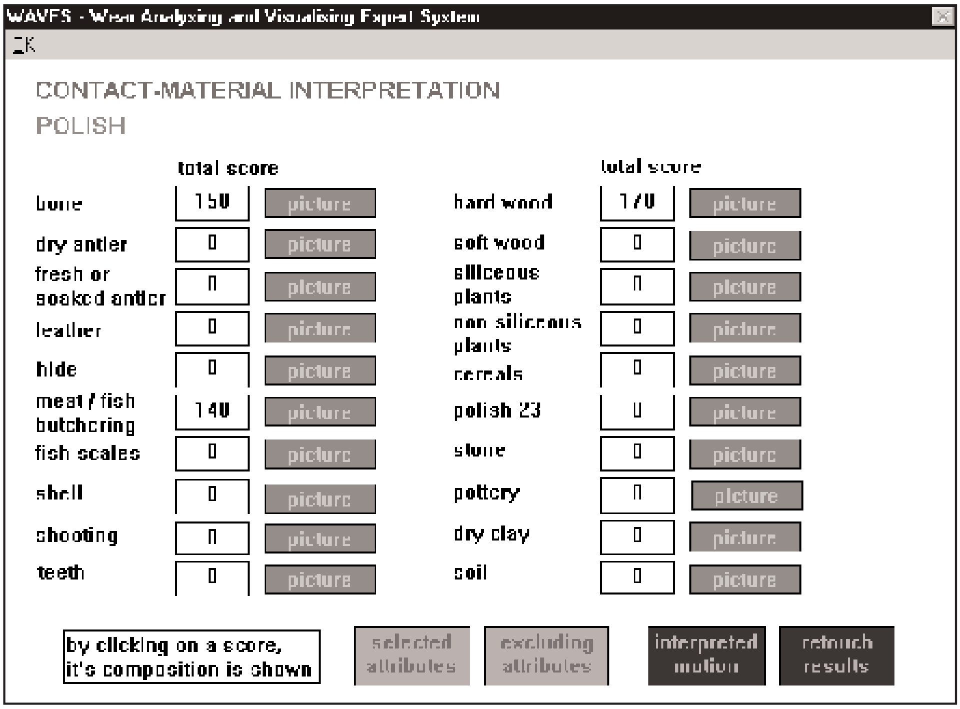


Fig. 27. Background information explaining the polish distribution types.

When the characteristics of the retouch have been described, the analysis of the edge-rounding follows. If present, the degree of rounding is established. Subsequently, the characteristics of the polish are examined: its location on the implement, the resemblance of the character of the polish on the dorsal and the ventral side, its directionality, its distribution, its width, its texture, its topography, its brightness and the amount of striations in the polish. The reason for asking about the resemblance of the polish on the dorsal and ventral side is that this may be an indication for polish ‘23’. In case there are clear and distinct differences between the polish on the dorsal and the ventral side, the user is asked to describe both sides in succession to enable the traces to be interpreted as the mysterious polish ‘23’. With regard to the topography of the polish, the user is allowed to select several attributes to describe it. Since especially topographical features seem to display a stage of polish development and many implements show several stages of polish development, the variety of topographical features may be indicative for the material by which an implement has been in contact.
The invasiveness of the use retouch versus that of the polish is the last aspect of the wear traces the user is queried on. This can be an important indication for the resistivity of the worked material. Obviously, this question is only posed when the user has described both the use retouch end the polish. Finally, when the description of the wear traces has been completed, the system continues the analysis by collecting information on the morphological aspects of the edge. The user is asked to measure the edge angle of the edge of which the wear traces were described, to give an indication of the grain size of the flint and to describe the shape of the edge. These elements are not of major importance for the composition of the interpretation, but they are merely used to check whether the result of the wear-analysis, *i.e.* the supposed activity, is in accordance with the morphology of the tool. If they suit, some of the morphological information may be used as a reinforcement of the interpretation. If they diverge, the user is notified of this by means of a remark or, in exceptional cases, the interpretation is revised. In that case a material or motion can still be excluded from the interpretation.

5.7.3 INTERPRETING THE INTERPRETATION

On the basis of the collected information, the system presents its final interpretation regarding the activity that may have caused the traces. It consists of a value (called ‘score’ in WAVES) which represents the diagnostic value of the wear pattern in reference to that particular contact material. This value is the sum of all the scores that were assigned to that particular material on the basis of the wear features which the user described. The analyzed wear pattern is most diagnostic for the material with the highest value. The value does not represent a probability, however. It can only be used for a comparison with the other diagnostic values that the analysis may have yielded. It does represent a weighed value: the wear attribute which is the most characteristic of



CONTACT-MATERIAL INTERPRETATION

POLISH

	total score		total score
bone	150	hard wood	170
dry antler	0	soft wood	0
fresh or soaked antler	0	siliceous plants	0
leather	0	non siliceous plants	0
hide	0	cereals	0
meat / fish butchering	140	polish 23	0
fish scales	0	stone	0
shell	0	pottery	0
shooting	0	dry clay	0
teeth	0	soil	0

by clicking on a score, it's composition is shown

Fig. 28. Interpretation screen showing the diagnostic values ('scores') of the wear traces in relation to the contact materials.

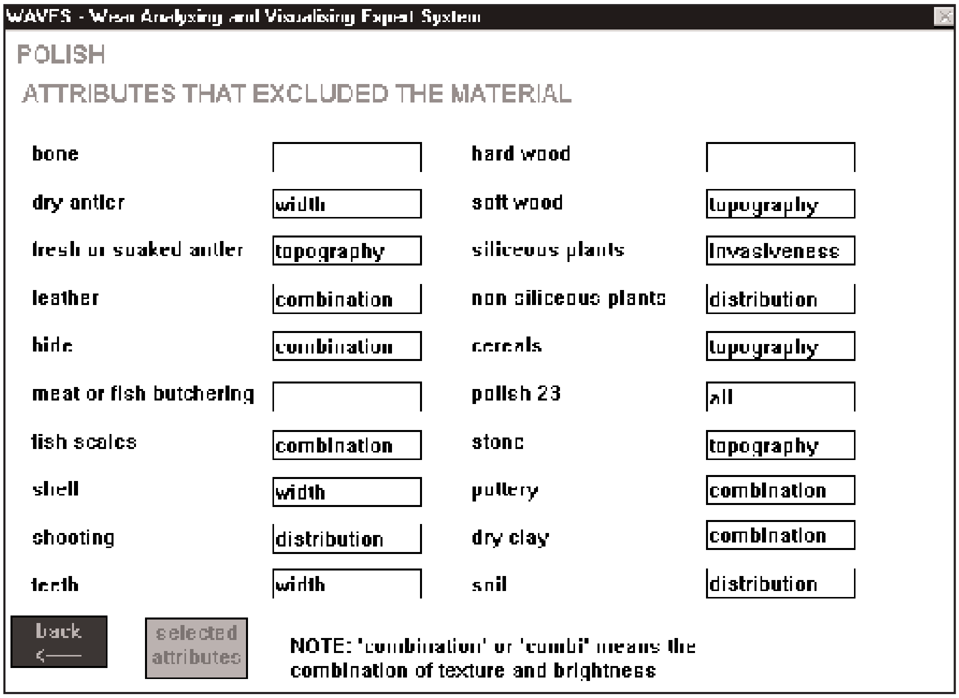
a pattern is the one which occurs most frequently. Automatically this attribute has the highest *frequency of occurrence* (see paragraph 5.2.4).

The reasoning process is based on the principle that 'the strength of a chain is determined by its weakest link'. This means that the chain of inference stops as soon as one element is missing. As a result, a path of inference that leads to a particular contact material or motion can only be completed if these materials gain a score on all variables, *i.e.* if all the observed wear-characteristics are diagnostic for that material.

In paragraph 5.2.3 and 5.2.4 it has been argued that wear traces are not always sufficiently diagnostic to enable an interpretation with a high level of specificity. Moreover, some contact materials may cause similar traces. Therefore, the system presents its interpretation as a list of all materials and motions which the user should take into consideration (fig. 28). The interpretational spectrum consists of 20 materials on the level of the polish, three on the level of the use-retouch and of five motions on both levels. In relation to the experimental program one material was added, the so-called polish '23' (Keeley 1977, Van Gijn 1989: 85). Each of these materials is accompanied by a score which indicates the diagnostic value of the analyzed wear pattern for that particular material. In theory, this score may reach as high as 1500. But this is only possible if a wear pattern from the reference collection received the highest frequency score (100%) on all of its 15 characteristics and if the analyzed wear pattern resembles this pattern from the reference collection with 100%.

If a particular contact material or motion did not receive a score at all, this means that the system did not find any indication that the artefact had been in contact with that particular contact material or applied in that motion. In that case the cell on the screen remains empty. If, on the other hand, a cell contains a zero this refers to the fact that the system found a few, but not enough, indications in favour of that contact material or motion. By clicking on the button called 'excluding attributes', the user can look at the wear characteristics that were responsible for the exclusion of particular materials (fig. 29).

Although the diagnostic values of the analyzed wear patterns may give a reliable impression of the materials that may have caused the wear, it is recommendable that the user also verifies the resemblance between the observed wear-traces and those of the reference collection. For that purpose, wear traces of each material and motion are illustrated in WAVES. By clicking on the 'picture'-button that accompanies a contact material, photos are displayed of the different traces that this material causes (fig. 30). They can be enlarged into full screen vision (fig. 31). Such a verification is always advisable, but necessary in case the wear pattern is equally diagnostic for several contact materials. It is always conceivable that the material scoring best did not actually



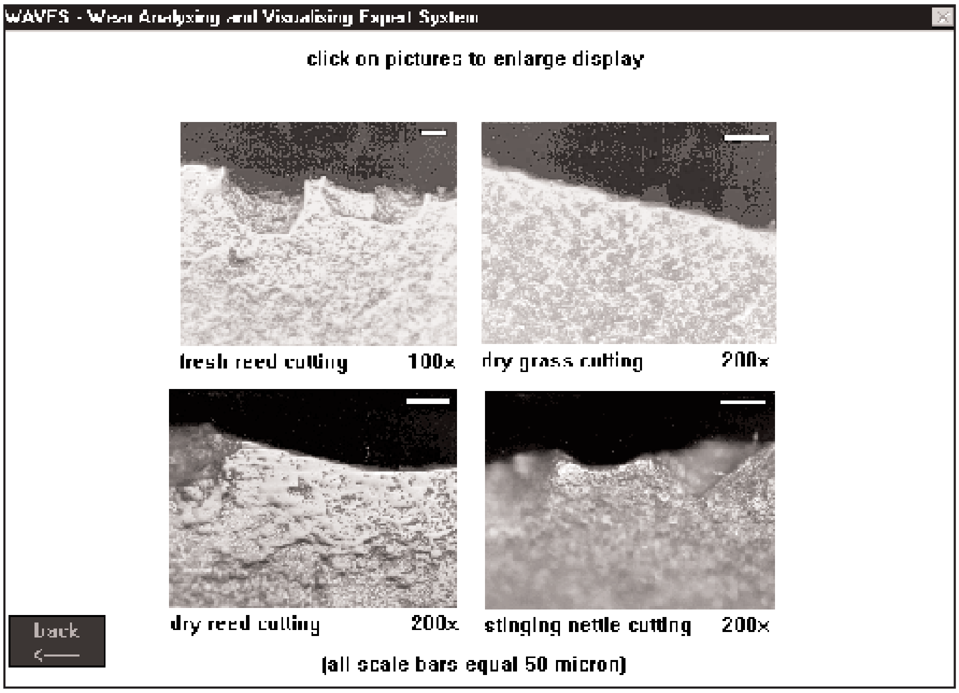
POLISH

ATTRIBUTES THAT EXCLUDED THE MATERIAL

bone		hard wood	
dry antler	width	soft wood	topography
fresh or soaked antler	topography	siliceous plants	Invasiveness
leather	combination	non siliceous plants	distribution
hide	combination	cereals	topography
meat or fish butchering		polish 23	all
fish scales	combination	stone	topography
shell	width	pottery	combination
shooting	distribution	dry clay	combination
teeth	width	soil	distribution

NOTE: 'combination' or 'combi' means the combination of texture and brightness

Fig. 29. A screen showing the wear features that excluded materials from the interpretation.



click on pictures to enlarge display

fresh reed cutting 100x

dry grass cutting 200x

dry reed cutting 200x

stinging nettle cutting 200x

(all scale bars equal 50 micron)

Fig. 30. The variation of wear traces that can be caused by a particular contact material.

cause the traces, but a material scoring less. This is possible as the height of a score does not refer to the probability of the interpretation. Based on the photos of the wear traces, the user may decide to give preference to the material scoring second best or even less, if this corresponds better with his own observations.

By using this approach of a plural interpretation, the competence of WAVES has been restricted: it only gives suggestions

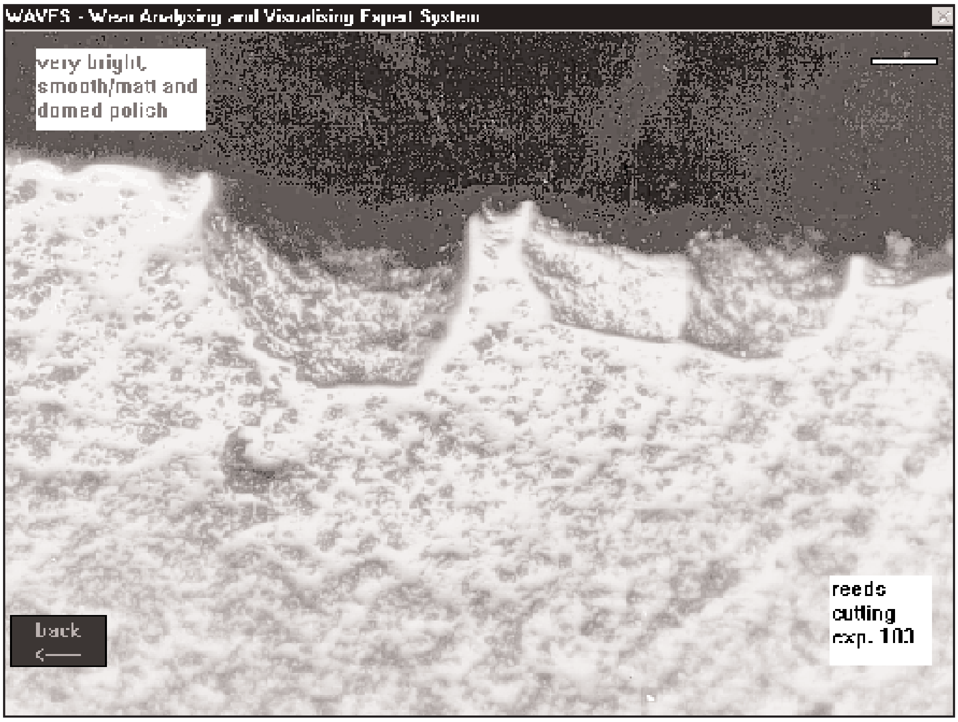


Fig. 31. Detail of wear traces.

about the possible interpretations. Its competence has deliberately been focused on the guidance of the data input and of the process of the analysis, because it is believed that the final interpretation always has to remain the responsibility of the user. He or she is the only one who is able to decide whether the expert system's interpretation is in accordance with the wear traces that were observed on the implement.

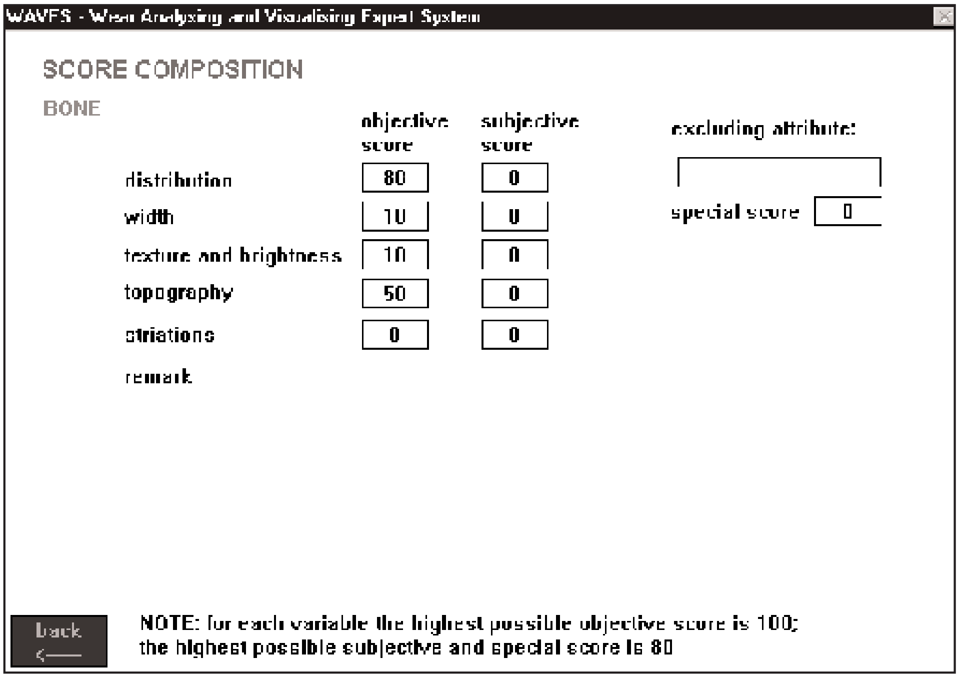
Another facility is the outline of the composition of the diagnostic value (fig. 32). This is also meant to support the validation of the interpretation. It explains the underlying reasoning processes and the relation between the wear features on one side and the materials and motions on the other. Of all materials and motions, even of those that were eventually excluded from the interpretation, it is shown for which material, if any, the observed attributes are diagnostic. On this screen it is also shown to what extent the interpretation is based on objective, calculated values or on subjective, estimated values (see also paragraph 5.5). Additional remarks may give some further elucidation of the composition of the diagnostic value. It may for instance occur that a particular combination of attributes is an extra indication for a particular conclusion. In that case, this is explained by a remark.

Furthermore, the user can ask for the results of the retouch analysis and of the supposed motion. These interpretations are structured comparably with the one that is based on the polish characteristics (fig. 33). They are accompanied by the same kind of explanatory information and illustrations (fig. 34, fig. 35).

5.7.4 RUNNING THE HYPOTHESIS VALIDATION PROCEDURE

The second module evaluates interpretations in order to support the more advanced student (or experienced analyst) who is already familiar with the procedures but wants to validate his or her interpretations. Hypothesis validation implies that the user presents his interpretation to the application and, in its turn, it will try to confirm the user's interpretation by questioning him or her about the observed wear features that have led to that interpretation. Figure 38 shows the system scheme of this procedure.

Like with the analysing part of the application, detailed information is needed from the user. Again this is obtained by a game of question-and-answer. The procedure starts by asking for the degree of detail of the user's interpretation, *i.e.* whether it concerns a hypothesis on the level of the hardness of the material (hard, medium or soft) and on the kind of material (animal, vegetal or inorganic), or on the level of the precise material. Based on the answer, lists are shown of the options of which the application has knowledge (fig. 36). The user can select his hypothesis from these lists. Photos of wear traces are provided to enable the user to



SCORE COMPOSITION

BONE

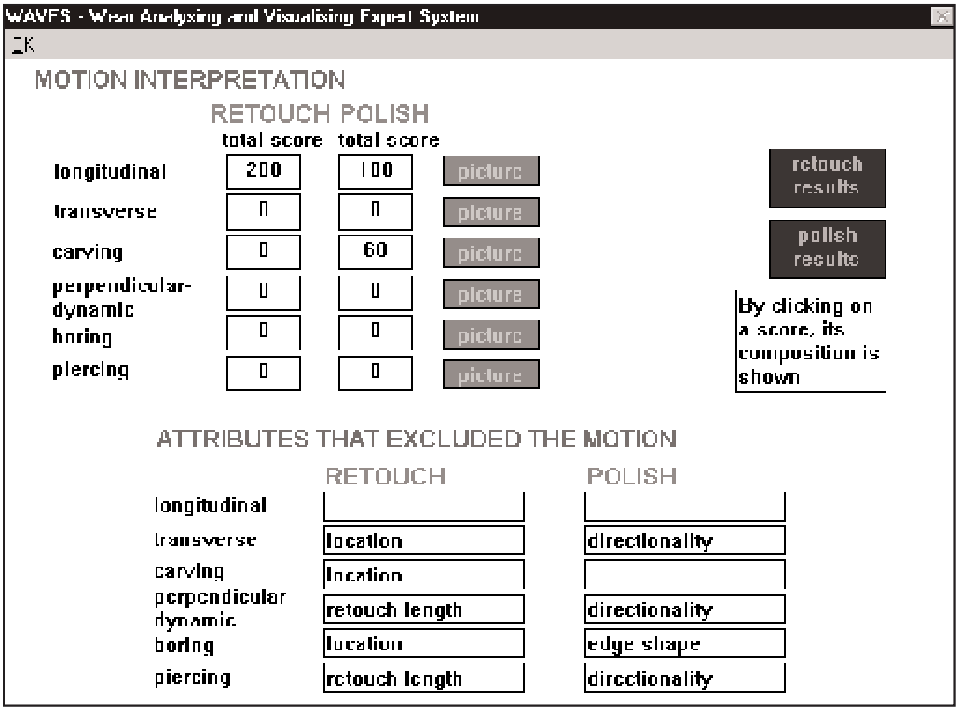
	objective score	subjective score
distribution	80	0
width	10	0
texture and brightness	10	0
topography	50	0
striations	0	0
remark		

excluding attribute:

special score 0

NOTE: for each variable the highest possible objective score is 100; the highest possible subjective and special score is 80

Fig. 32. Interpretation of the applied motion.



MOTION INTERPRETATION

	RETOUCH total score	POLISH total score
longitudinal	200	100
transverse	0	0
carving	0	60
perpendicular-dynamic	0	0
boring	0	0
piercing	0	0

By clicking on a score, its composition is shown

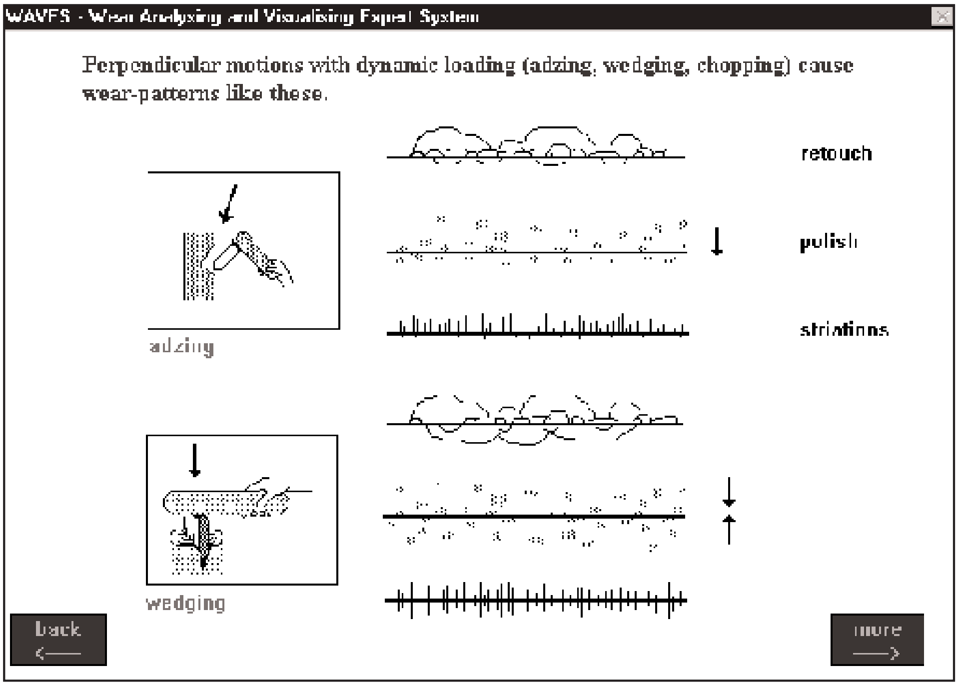
ATTRIBUTES THAT EXCLUDED THE MOTION

	RETOUCH	POLISH
longitudinal		
transverse	location	directionality
carving	location	
perpendicular dynamic	retouch length	directionality
boring	location	edge shape
piercing	retouch length	directionality

Fig. 33. Interpretation screen showing the composition of a diagnostic value.

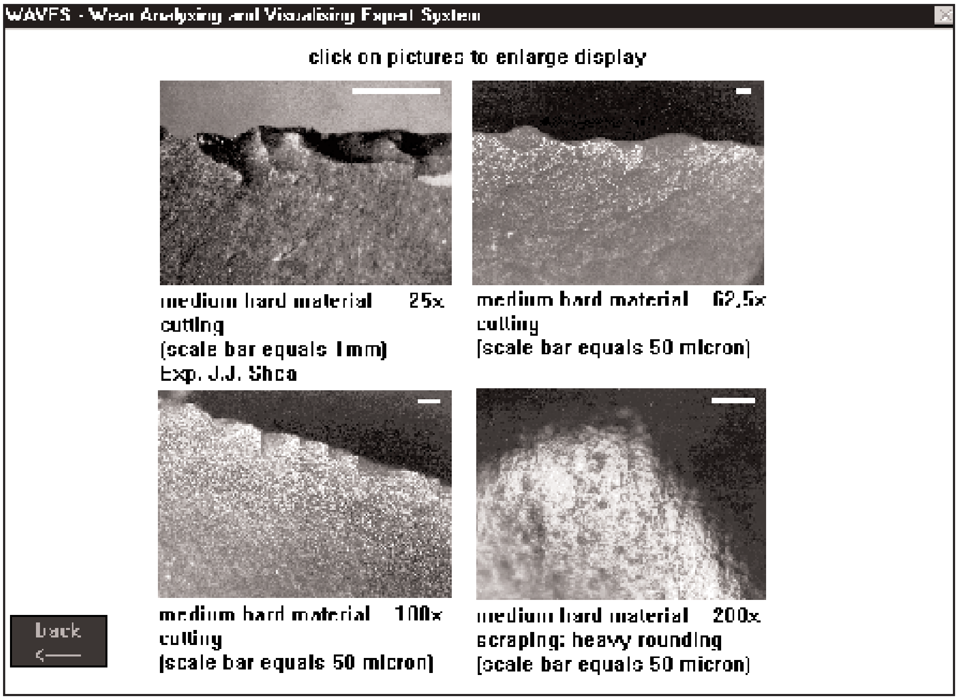
make a first validation of his hypothesis before the actual validation process is started. These are the same photos which accompany the interpretation of the analysis procedure.

Subsequently the actual verification starts. It consists of queries on the presence of particular wear features. These queries are tuned to the hypothesis. In order to make sure that the system receives the appropriate information, and to avoid invalid input,



Perpendicular motions with dynamic loading (adzing, wedging, chopping) cause wear-patterns like these.

Fig. 34. Traces which are caused by perpendicular applied motions.



click on pictures to enlarge display

medium hard material 25x
cutting
(scale bar equals 1mm)
Exp. J.J. Shea

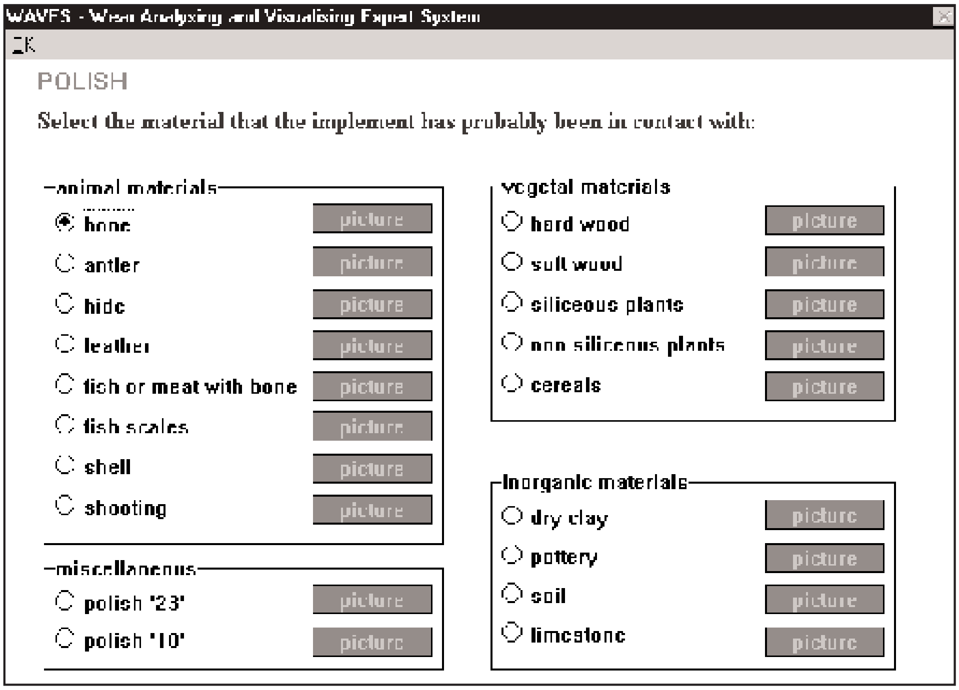
medium hard material 62,5x
cutting
(scale bar equals 50 micron)

medium hard material 100x
cutting
(scale bar equals 50 micron)

medium hard material 200x
scraping: heavy rounding
(scale bar equals 50 micron)

Fig. 35. Traces which are caused by working medium hard materials.

the answers to the questions are of a simple TRUE/FALSE-format (fig. 37). The user only has to select TRUE if his observation corresponds with the question the system asks, or FALSE. If a diagnostic feature appears to be absent, there may be an alternative which, assuming that it can be acknowledged, may still allow a confirmation of the hypothesis. This implies that in many cases the validation session is not immediately interrupted if a question is answered negatively.



POLISH

Select the material that the implement has probably been in contact with:

animal materials
- ● bone
- ○ antler
- ○ hide
- ○ leather
- ○ fish or meat with bone
- ○ fish scales
- ○ shell
- ○ shooting

miscellaneous
- ○ polish '23'
- ○ polish '10'

vegetal materials
- ○ hard wood
- ○ soft wood
- ○ siliceous plants
- ○ non siliceous plants
- ○ cereals

inorganic materials
- ○ dry clay
- ○ pottery
- ○ soil
- ○ limestone

Fig. 36. List of materials from which the user can select a hypothesis.

Opposite to the analysis procedure, the validation procedure does not provide additional explanatory facilities. If students or analysts use this procedure they are expected to have sufficient knowledge of the subject at hand to perform the procedure without the help of background information. For instance all wear attributes that are included in the application are expected to be known, but to make sure that the user and the application are talking about the same thing, a photo of the relevant attribute accompanies the query. Another contrast with the analysis part is that the wear categories are not handled as separate blocks which can be skipped on the initiative of the user. The interpretation of use-retouch, polish, edge rounding, and striations are fully integrated. In various cases both micro and macro traces have to be present in a particular combination in order to justify an interpretation.

Eventually, when all questions have been answered, the conclusion is given. If the presence of the diagnostic wear features cannot be acknowledged, a message is given which proposes to reconsider the hypothesis. Subsequently the user may revise and retest his hypothesis. For instance, WAVES may falsify ‘hard wood’ while it considers ‘soft wood’ an acceptable interpretation. Alternatively, the user can employ the analysis procedure in order to obtain an alternative interpretation. If, on the other hand, all diagnostic features are present, the user is informed that there are no reasons to doubt his hypothesis. This interpretation is not subscribed by a diagnostic value: a hypothesis is either fully confirmed or not.

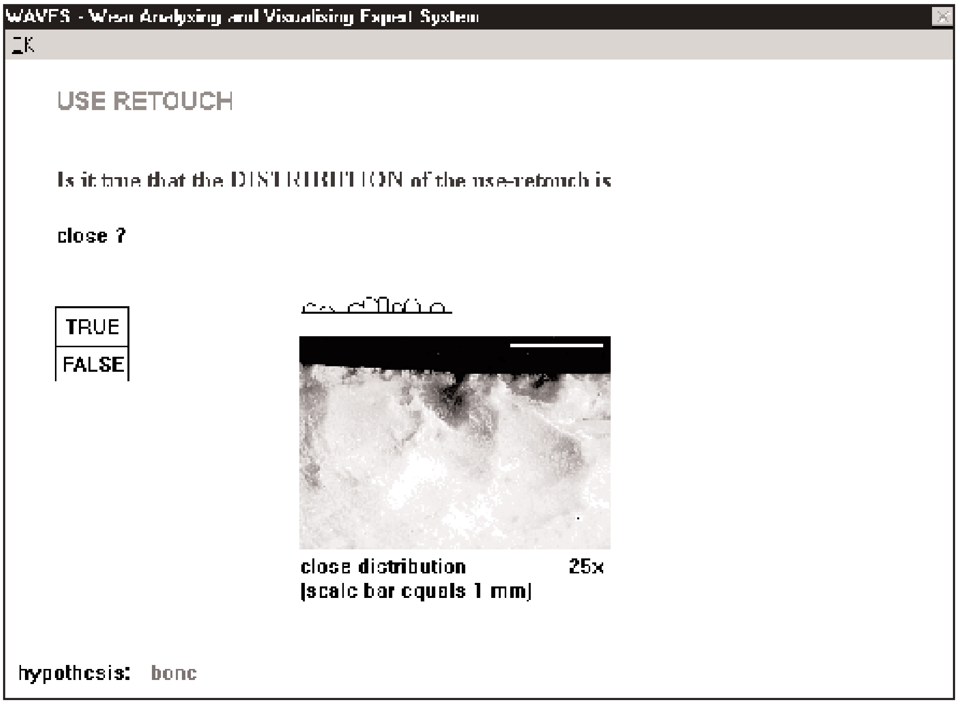
5.8 An assessment of the composition of the application

5.8.1 INTRODUCTION

The demands that archaeologists, and in particular use-wear analysts, pose on expert system applications (paragraph 2.4.2 and 4.5.6) have been kept in mind during the design and implementation of WAVES. As it is clear that most of these demands are crucial for the functionality of the application and thus for its acceptance by analysts and students, it must subsequently be assessed on these aspects. Furthermore, knowledge engineers have requirements as well. But the approaches that were followed in this respect will not be discussed in particular, as in this study only well-known methods have been applied. Besides, the choice of these methods was highly dictated by the shell that was selected.

5.8.2 ANSWERS TO EXPECTATIONS OF *COMPUTER ARCHAEOLOGISTS*[15]

The most important complaint from archaeologists in relation to knowledge based applications is that they usually offer only one interpretation without leaving an opportunity for uncertainty. As this is hardly in accordance with an academic approach, they expect from an application that it offers an alternative approach for this ‘one-solution attitude’.



USE RETOUCH

Is it true that the DISTRIBUTION of the use-retouch is

close ?

close distribution 25x
(scale bar equals 1 mm)

hypothesis: bone

Fig. 37. A question screen of the hypothesis validation procedure.

In designing WAVES, it was therefore decided to allow the same uncertainty in the conclusion as often is experienced with interpretations of human analysts. The consequence of this, however, is that the user becomes responsible for the final conclusion rather than the program. For students this may be a somewhat confusing approach, especially in the beginning. They will certainly draw some wrong conclusions or may even be afraid to draw a conclusion at all. Here, some guidance may be needed from a more advanced analyst or an expert, or they may even double check their conclusion by consulting the hypothesis validation procedure. Nevertheless, it is assumed to be a valuable approach, because it represents the real world situation. It confronts students with the ordinary problems of use-wear analysis and forces them to think of good arguments that favour the various options.

Another major concern with knowledge based systems is their *black box* property. It means that their communication with the outside world merely consists of receiving input data and of returning an answer. It will be clear that this approach only tolerates a user to supply data: questions about the arguments that underlie interpretations are not to be asked. Although this is a rather arrogant approach, many of the first knowledge-based applications were designed in this way. From a scientific or an educational point of view this is not acceptable and knowledge engineers devised alternative approaches. An important improvement was the development of explanatory facilities such as discussed in section 5.3.3. For this reason it has been tried to make the analysis procedure of WAVES as transparent as possible: it shows how it composes the interpretations and how the wear features relate to the contact materials or motions.

Closely related to the ‘black box’ problem is the fear for *knowledge fossilization* (see chapter 2.4). It is expected that once knowledge is captured in an automated system, it will be applied for many years on the assumption that it reflects the state of affairs, while in fact it may already be outdated. Especially in disciplines that are still evolving, like use-wear analysis, this is indeed an important point of concern, although no knowledge is definite or final. It must also be kept in mind that fossilization of knowledge is inherent to science. Every article or book that is written, every database that is filled, causes knowledge to fossilize. Nonetheless, it has never led to disastrous effects and has not kept scientists from writing or from filling databases. If there were no points at which knowledge is ‘frozen’, moments of assessment would not exist either. Besides, by distributing the state of affairs by means of knowledge based computer systems, others are enabled to assess it and to continue from that point. It only requires the acknowledgment that a particular article or knowledge-based application merely reflects the state of affairs of one particular moment.

The best way to avoid the fossilization of the knowledge of an expert system application is to make sure that the knowledge

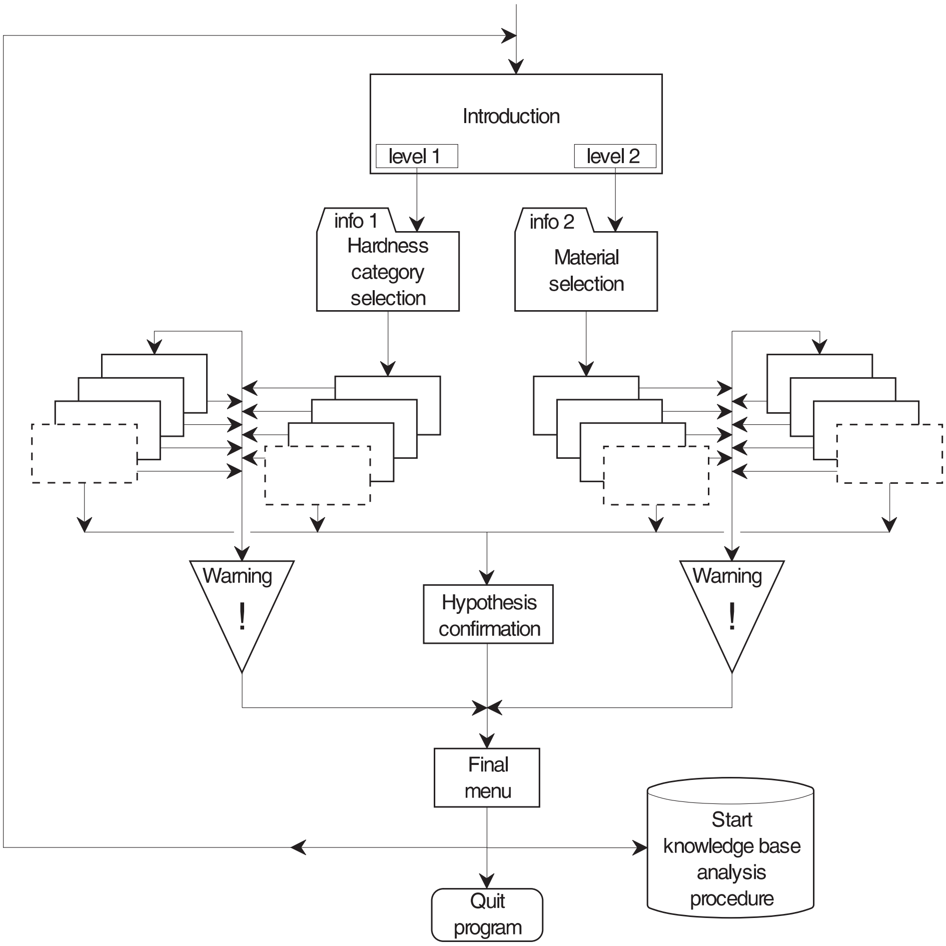


Fig. 38. System scheme of the hypothesis validation procedure of WAVES.

is evaluated and adjusted periodically. Unfortunately, this is often not realised by those who initiate the application building process and no financial means are reserved for this purpose. It is, nonetheless, recommendable to take the aspect of maintenance into account at the start of any application development process, and to incorporate facilities for this purpose. It might be a good suggestion to reinvest (part of) the sales profits in maintenance and improvement. With reference to WAVES, it is not expected that the present design will have to be adapted radically. But, it will absolutely be necessary to maintain its conceptual knowledge repeatedly, because this is expanding rapidly. In order to meet the fossilization fear, the knowledge has been represented by means of easy accessible decision rules and by easily adaptable diagnostic values. In order to stimulate and guarantee that the present knowledge in WAVES will indeed be validated and expanded, it would be useful to create an international database for experimental achievements that can act as a basic source of knowledge and which represents the state of knowledge.

A last demand that archaeologists pose on knowledge-based applications is a *user-friendly communication interface*. This has been tried to meet in WAVES by including explanatory facilities (see also sections 5.3.3, 5.3.4 and paragraph 5.7) and by using as much ordinary language as possible.

5.8.3 ANSWERS TO EXPECTATIONS OF USE-WEAR ANALYSTS

The main reason for use-wear analysts to be interested in automated approaches is that they may provide a means to standardize the method of analysis and to formalize the knowledge. But from the responses on previous attempts (see paragraph 4.5 and 4.6), it can be concluded that such an approach must meet several additional requirements in order to be eligible for actual use. Preferably, it is based on the analysis of various wear categories, it supports the learning process of students, it is suitable for practical use, it is maintenance-friendly, and applicable to different archaeological assemblages.

Probably the most important expectation of use-wear analysts is that the process of analysis becomes more standardized and the interpretations less subjective. For this purpose several automated systems have hitherto been designed and this was an important reason for the development of WAVES as well.

From the point of view of most use-wear analysts, an optimal analysis and therefore an ideal application incorporates all aspects of use-wear traces. Previously, many of the automated approaches were based on only a few characteristics, on a tiny area of the implement's edge, or on just one wear category (see chapter 4.6). In WAVES we have, therefore, included all wear categories and the wear characteristics on the entire edge of a tool. Furthermore, analysts increasingly deduce their interpretations from a combination of low-power and high-power observations (*cf.* Van Gijn 1989, Hurcombe 1992, Lemorini 1997), because this can yield more balanced and well-considered interpretations. We have kept this in mind when building WAVES: according to an analyst's preference, he or she is enabled to reach an interpretation by both methods: the edge removals and edge rounding are studied by the 'low-power' approach and the polish and striations by means of the 'high-power' approach.

Another important demand, which is however not typical for the field of use-wear analysis only, is that a tutoring system should indeed be able to support and maybe even quicken the learning process of students. In order to help students to learn from WAVES, illustrations and explanatory facilities have been incorporated. This is predominantly the case in the analysis procedure, as this is meant to help students to understand the reasoning process and to stimulate them to eventually simulate it.

Closely related to the former demand is the requirement that an application is user friendly. One of the main problems analysts have with many computer-based approaches of use-wear analysis, such as image analysis, is that they are technically very sophisticated but hardly practical. Especially a system that is being designed with the intention to be used in practice, ought to be easy to handle. Any archaeologist should be able to understand and work with it without intensive guidance or knowledge of expert systems. WAVES is, therefore, entirely menu-based and the communication between the user and the application consists of clicking buttons. The user does not have to type anything.

In my opinion, user friendliness also means that all parts of the application are easily adjustable. In principle, both procedures of WAVES are maintainable by a domain expert due to the fact that the decision rules are easy to understand, the diagnostic values are part of these rules, and the rules can be adjusted and expanded individually without this having affect on other parts. The part that is probably the most subject to change, *i.e.* the decision rules and the diagnostic values that are associated with them, is the easiest to accommodate. One can add or delete a rule, add or delete a material or motion, and adapt a diagnostic value. It is slightly more complicated to add new variables or attributes, because this implies that new question screens must be created and that the reasoning process must be expanded. This requires somewhat more knowledge of the internal affairs. It is possible to add illustrations too. Although this may require an adjustment of the infrastructure of the screens, it is not too complicated since it does not influence the reasoning process.

The expectation that an automated approach of use-wear analysis is applicable to various archaeological assemblages may seem more obvious than it really is. If the morphological characteristics of a tool play a decisive role in the method of analysis, like with the FAST program (Grace 1989), it is necessary to adapt the application's decision rules when the archaeological context changes. If, on the other hand, the method of analysis concentrates on the characteristics of the wear traces on edges only, like with WAVES, the same decision rules are applicable to wear traces from various assemblages. Furthermore, one can imagine that a system based on experimentally obtained information does not yield optimal results when employed for archaeological data.

Due to the discrepancy between the range of wear patterns in the real world and those obtained by replication, it is possible that the 'artificial' knowledge is not sufficient. In order to avoid that the decision rules of WAVES would not be useful for slightly divergent wear patterns, it has been designed to interpret combinations of features rather than fixed patterns (see paragraph 5.2).

In addition to the above requirements, several analysts asked about a possibility to turn WAVES into an automatically learning application. They wished that it would add knowledge about any new wear pattern that it encounters. Unfortunately, WAVES cannot *learn*. It would not only be a very hazardous and time-consuming task to create a system that is capable of adapting and expanding its own knowledge base, but in my opinion it is undesirable as well. Technically, it is possible to create such a system, but methodically there are major problems to be conquered. Given the difficulties that were encountered in creating relevant and reliable decision rules (paragraph 5.2) and the problem of the subjectivity that is involved in describing wear features (paragraph 5.3.2.1), it is hard to check the validity of the information that the user presents. Without a thorough verification such 'new' patterns should not be added. A pollution of the knowledge base with 'unreliable' facts would corrode its credibility. Methodically, it is recommendable that the domain expert supervises all maintenance activities.

5.8.4 UNANSWERED QUESTIONS AND SUGGESTIONS FOR ADDITIONS

Although several expectations may have been met in WAVES, there are still many difficulties to be solved. In this paragraph, the attention will be concentrated on the methodical aspects rather than on technical problems, since the latter belong to the scope of artificial intelligence research. Most methodical wrestlings concern the range of the conceptual knowledge and the subjectivity of the input data. These are, unfortunately, hard to settle. Most of the suggestions for additions are more easy to achieve since they relate to procedural aspects.

A first problem that needs serious attention is the *range of the conceptual knowledge* of the application. To a large extent the range of any application is determined by the expert's field of expertise and, in this case also by the scope of the experiments the knowledge was deduced from. Experts carry out experiments that fit in their field of interest. In reference to WAVES, especially the scope of raw materials and vegetal materials is restricted. Since the reference collection almost exclusively involved experiments with flint implements, WAVES has no knowledge about wear traces on tools made of quartzite, chert or obsidian. The same applies for the variety of worked materials that have been included. Although it is known that, in particular, vegetal materials may cause various traces, WAVES has knowledge of only six, including one that causes the mysterious polish '23' (Keeley 1977; Van Gijn 1989), which is only assumed to be vegetal. It must be stressed that the experimental program comprised several more vegetal materials, but the wear patterns that they caused were not distinct enough to include them separately. They were therefore treated as two groups, siliceous and non-siliceous plants. Moreover, the expert has knowledge about the traces that modern Dutch plants cause, but not of plants from other climatic zones. Although this confines the applicability of WAVES, it is not as problematic as it may seem. Many vegetal materials tend to cause somewhat universal wear traces. This implies that, although WAVES may not be able to give a suggestion on the exact material, it may still indicate whether the material is 'vegetal' or 'animal' and whether it is 'hard', 'medium hard' or 'soft'.

The restrictions of the conceptual knowledge not only relate to the composition of the experimental program, but also to the range of the included expert knowledge. The expertise in WAVES has predominantly been withdrawn from one expert. Consequently, it is automatically specialised on the same issues as this expert and it contains the same biases and premisses that the expert may have. From an educational point of view this may both be beneficial and undesirable. A benefit is that the system confirms the expert's method of analysis and vice versa. During their training period students will appreciated this kind of consistency. A disadvantage, however, is that they learn to look at the traces from one perspective only, while other experts may describe particular traces differently (see chapter 7).

One way to deal with such limitations is to *expand the experimental program* and to elicit the supplementary knowledge from various experts. Although this might obviate part of the missing information, it must be realized that it is impossible to incorporate knowledge about the traces of, for instance, all vegetal species that exist on our planet. Especially because we do not even have knowledge of all species that occurred in prehistoric times. It might, however, be worthwhile to add the best known of the various climatic zones. But incorporating the knowledge of various experts also raises a fundamental point of discussion. Since it is impossible to know everything of all items in the application, a much wider range of worked materials would be accomplished at the cost of profundity. Therefore it needs consideration whether one prefers specialization or generalization. But even if this choice would not be made, the latter may evolve automatically. It is to be expected that it may not be easy to get decision rules that all experts agree on and, eventually, the knowledge base may predominantly consist of compromises. On balance, students would not win anything with this approach: a compromise too shows only one perspective and presumably not the most inspiring one.

A second major problem to which no adequate answer has been found yet, is the subjective nature of the input the system receives from the user, *i.e.*, the descriptions of the observed wear features. In my opinion, this is the weakest spot of the application. Although it has been tried to accommodate to this problem by offering illustrations and definitions of the variables and attributes, and by using as many quantitative variables as possible. Nevertheless, this aspect of the application is the least under control and the performance of WAVES remains dependent of the competence of the user. Although there are a few suggestions that may improve this aspect, such as adding more photo's (paragraph 5.3.2.2), it is not expected that the best solution will be found in the direct line of the present approach. Either use-wear analysts may think of a means to adapt the method of description or it must be tried to find a combined approach in which automated pattern recognition is involved.

A final problem with the conceptual knowledge is that there is no objective definition of what is meant with 'soft', 'medium hard' or 'hard' materials. Again, this is not a problem that is exclusively related to WAVES, but in general the domain of use-wear analysis remains indeterminate on this matter. As this raises questions and creates confusion it would be better to employ universally accepted scales of hardness that chemists and physicists use. This could improve the comparability of interpretations as well.

Apart from these difficulties to solve methodical problems, WAVES has some imperfections that may be more easy to improve. For instance, it cannot analyze both the dorsal and the ventral edge of a tool at the same time. This is not problematic when both edges display a similar wear pattern, but in some cases the wear traces show differences. Since a discrepancy between the traces on the ventral and dorsal side may be highly indicative of the applied motion, it is important to include this aspect as well. In the present version of WAVES, the user has to analyze both sides independently and to use his imagination in order to combine both interpretations into one. In this approach, however, the added value that the specific combination of the traces may yield cannot be exploited.[16] The reason for this imperfection is, unfortunately, not only a procedural matter. There is simply not yet sufficient data to deduce rules on the relationship between the variance of the wear patterns on the dorsal and ventral face of an implement and the applied contact material and motion. In order to implement WAVES with this kind of data, it must be collected systematically first.

Another point of attention is the need for more photographic material. It is, for instance, known that especially polish develops through several stages, dependent of the duration of the activity the tool has been involved in. Since WAVES is a tutoring system, it is important that the whole range of these stages is incorporated. Even if it could only interpret two activities, it should be able to display the entire range of wear traces that these activities can cause. Co-operation with other use-wear analysts is required, however, to collect good examples of various wear-development stages of all the materials incorporated.

With reference to the application's user friendliness, some facilities ought to be added. For the sake of comparisons or statistical analysis of the results, it would for instance be useful if the interpretations of the analysis procedure could be saved in a database. Moreover, the hypothesis validation procedure may be equipped with some more advanced explanatory facilities, for instance with regard to the selection of the lines of reasoning. It should also give the user an indication of the value of the validation of the hypothesis, analogous to the certainty factors of the analysis procedure. Since none of these facilities are technically very difficult to program, they can be integrated in a next version.

A final aspect which needs attention in the future is the integration of additional interpretational means. Especially residue-analysis would be an interesting candidate. Increasingly, use-wear analysts and residue analysts tend to co-operate because the methods yield complementary results (Fullagar, Furby & Hardy 1996). This is therefore an argument in favour of an expansion of the application with this approach.

To conclude, it may be clear that some of the problems that were encountered during the development of WAVES are inherent to the method of use-wear analysis: their solution may require fundamental adaptions. Others may be solved by making clear choices and by simply accepting their consequences. In either case however, the choices will always be disputable.

5.9 Comparison with FAST

WAVES is not the first knowledge based application that has been developed for the analysis of use-wear traces. Roger Grace has been the pioneer in this field. In the late eighties, he constructed a system, called FAST (Functional Analysis of Stone Tools), by which he could interpret the function of lithic tools on the basis of use-wear traces (Grace 1989). Like WAVES, it was built for the purpose of assisting students.

There are, nevertheless, some major differences between the two applications. In fact, the only thing they have in common is that both are knowledge-based systems that operate on the domain of use-wear analysis. Therefore, this paragraph provides a comparison. This is, however, not meant as a means for judgement but merely as an informative inventory of the main discrepancies. The main reason for the differences between both applications is that FAST was developed for the purpose of functional analysis of entire tools rather than the interpretation of use-wear traces. This means that Grace regarded use-wear traces as one part of the analysis process only, whereas in WAVES they are the main point of interest. This is illustrated by the fact that in the analysis process of WAVES eight characteristics concerning the polish are involved, versus five in FAST (Grace 1989: appendix 4, figure 92-94).

In FAST, however, a tool's morphological aspects are more important: FAST includes nine morphological features, WAVES three. Since our intention was not focus on the relationship between tool shape and tool function, we employed only those morphological aspects which proved relevant for the development of wear traces. Moreover, WAVES focuses on the functional analysis of edges rather than of the implement as a whole. An additional implication of our approach is that it is more easily applicable to different archaeological assemblages. Since, the traces of working a hide from a paleolithic deer do not differ from traces of working a neolithic deer, WAVES should be able to interpret both. Grace, on the other hand, has focused on morphological aspects and has to adapt his program to each archeological period it is employed for: "*...the analyst can assess each variable individually and change the parameters of the variable ranges along with the relevant functional indications to fit the archaeological assemblage that is to be analysed.*" (Grace 1989: 224).

Furthermore, there are major differences regarding the procedural aspects of the applications. First of all, FAST gives an interpretation that consists of one final answer (*ibid.*:

appendix 4, figure 94). This is a drawback, since many use-wear analysts have demonstrated that traces of wear are seldom exclusively diagnostic for one particular activity (see chapter 4). This was handled differently in WAVES, because it has an educational task and has to make apprentices aware of the methodical difficulties that are inherent to this approach. Therefore, it shows all materials that could have caused the traces and the diagnostic value of these traces. Subsequently, the user has the opportunity to consider and validate the final interpretation. Another procedural difference is that WAVES consists of both an analysis and a hypothesis validating procedure, whereas FAST confines itself to the first.

Regarding the transparency of the internal reasoning processes, there are some dissimilarities as well. FAST has more or less been constructed like a 'black box' without explanatory facilities. Although the user has access to the underlying rules, the reasoning process and the interpretation is not really explained. A final, technical difference, is that FAST runs on an Apple Macintosh computer, whereas WAVES was developed for IBM-compatible platforms.

notes

1 The use-wear experts A.L. van Gijn and J. Schreurs of the Faculty for Pre- and Protohistory of the Leiden University have been so kind as to place their knowledge at the disposal of this study.

2 The results of these analyses have already been published in Van den Dries & van Gijn (in press).

3 An additional argument that a selection of variables and attributes was admissible on the basis of the cross- tabulations, was provided after the analysis had been finished. The data were then submitted to multi-variate statistics (*e.g.* principle components analysis), by prof. Richard Wright, University of Sydney, Australia. The results of these test were that on the basis of these variables and attributes no clear patterns could be distinguished between the occurrence of the wear patterns and the worked materials or motions. The results of the statistics confirmed that no single variable has an exclusively diagnostic value (pers. comm.).

4 An expert system shell is a special software package that is equipped with all facilities necessary to build an operational program (see also chapter 3.5). Basic features are already present, so they do not need to be programmed anymore. In order to get an operational system you only have to implement knowledge within a shell.

5 Level 5 Object is a registered trademark of Information Builders, Inc.

6 Windows is a registered trademark of Microsoft Corporation.

7 For the display of the photos incorporated in WAVES a Super VGA-card and a colour monitor are required that can support a graphics mode of 256 colours with a (minimal) resolution of 640 × 480 dots.

8 These requirements relate to the version WAVES was built with. More recent versions have other hardware requirements.

9 This does not mean that an application that has been built with this tool requires the same configuration; dependent on the dimension of the application it may require a more powerful or even a less powerful computer.

10 The photos were made by A.L. van Gijn and J. Pauptit and subsequently scanned by the author by means of a high quality (resolution) flat bed scanner.

11 The diagnostic values that are derived from the experiments are called 'objective', although it is realized that the description of these traces are not purely objective. The criterion for calling them objective is that they have been calculated, rather than estimated. The diagnostic values that were derived from the expert's expertise are estimates that could not be verified quantitatively and are therefore called 'subjective'.

12 In Level5 Object forward reasoning rules are called 'demons'.

13 These material and motions were part of the experimental program of Annelou van Gijn (1989: 168-174), except for one that has only been experienced archeologically (polish '23').

14 WAVES requires a Windows-based computer with at least 40 megabyte of free disk space, due to the large amount of images, and 2 megabyte of RAM memory. All versions of Windows (2.0 or higher, Windows for workgroups and Windows95) are allowed.
In order to get high quality illustrations, the computer must have a graphics card with at least 1 megabyte of memory in order to obtain a resolution of 640 by 480 dpi in combination with 256 grey levels. WAVES can only be used within a Level5 software environment, which is supplied with the application. It is possible to consult the knowledge bases via a network, but the Level5 runtime environment is incompatible with a network and has to run on a hard disk of a stand alone computer.

15 *Computer archaeologists* refers to people who are specialized in applying and developing quantitative and automated methods for archaeological issues.

16 In relation to polish '23' the reverse approach is applied. If a user indicates that the wear traces differ completely on the ventral and dorsal side of a tool, he can give a description of the features on both edges which the system will interpret in relation to each other. Only if the description of both sides correspond with the expectations of the system, the traces may be interpreted as being caused by this mysterious material. The reason for this exception is that this type of wear is very distinct. Especially the discrepancy between the traces on the ventral and dorsal face are characteristic.