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

Dries, M.H. van den

4 The application domain: use-wear analysis of prehistoric flint tools

4.1 Introduction

The subject of the knowledge-based application which is the case study of this thesis, is the analysis of use-wear traces on prehistoric flint tools. Use-wear analysis is a method that can render information on the function of prehistoric stone tools. It consists of the investigation of a stone tool edge and surface, for the purpose of tracing evidence of use. The premise of use-wear analysis is that the damage an archaeological artefact bears may relate to its former function.
This has been deduced from experiments with replicated artefacts. These demonstrated that the processing of organic and inorganic materials can cause the edges of a tool to be damaged and its surface to be modified. It is assumed that in case of congruent traces, the functional interpretation of the archaeological tool can be inferred from an empirical comparison of its damage pattern with those of the experimentally employed implement.[1]
Use-wear traces, which can sometimes be observed by eye but usually only by microscope, offers information that is used for the purpose of answering questions about human economic and behavioural activities. These questions can focus on tool-specific research, for instance, on the function and task-specificity of certain tool types (*e.g*. Juel Jensen 1982, 1988; Van Gijn 1988), or on site-related research, for instance to investigate site differentiation regarding subsistence and craft activities (*e.g*. Van Gijn 1989; Schreurs 1992). These studies do not have to be confined to intra-site situations, but can also be directed towards inter-site variation or similarities, either in a synchronic or a diachronic perspective (see also Cook & Dumont 1987; Juel Jensen 1988).
The analysis method is being applied on a worldwide scale and by a considerable number of archaeologists. This, however, has not been accomplished without serious struggles: it has taken quite a while before this method was academically accepted. Some scholars are still reserved about its reliability. Sceptics are mainly concerned about the fact that the method involves subjective observations and unformalized, unstandardised interpretations. Obviously, this concerned use-wear analysts as well, especially because it complicates and lengthens the training trajectory of students. Consequently, they have made various attempts to improve their method on these specific elements. It is also within this context that the expert system application (WAVES) has been developed.
In this chapter some background information will be given on the theoretical and methodical aspects of this method, in order to give an impression of the framework in which WAVES has been developed and which has influenced or almost dictated its design. It is for this reason that the emphasis will be lying on the difficulties that this method has encountered and on the benefits of the solutions that hitherto have been proposed by various researchers.
In outline, the emergence and development of use-wear analysis will be described in paragraph 4.2 and its methodical aspects are exposed in paragraph 4.3. The particular difficulties that use-wear analysis faces are discussed in paragraph 4.4, the solutions that have hithertho been developed and their achievements in paragraph 4.5 and paragraph 4.6.

4.2 The emergence and development of use-wear analysis

Use-wear analysis arises from a curiosity to learn how and for what tasks prehistoric man used his stone tools. This has interested archaeologists from the mid-nineteenth century onwards. They first tried to build stone tool classifications by means of ethnographic analogies and later on by experiments by which they tested the efficiency of a particular tool type for a particular task. Subsequently, they noticed that the use of stone tools left traces on the edges of the artefacts and assumed that this could provide hypotheses on tool use that would be directly referable to prehistoric man. When, at the end of the last century, it was acknowledged through serious experiments (Spurrell 1892) that there is a diagnostic relation between tool use and edge damage, use-wear analysis was born.
In the beginning of the twentieth century more studies followed (*e.g*. Curwen 1930), but use-wear analysis has only become a well-known and widely applied method after Semenov's 'Pervobytnaya Tekhnika' (1957) was published in English in 1964. In this famous and still useful book, Semenov discussed methods of functional analyses of palaeolithic tools. He argued that neither replication of stone manufacturing or of use would yield sufficient information on the actual function of an implement. He acknowledged

the importance of experiments and ethnographic information, but considered a systematic study of (microscopic) traces as the most valuable means to understand the whole range of activities for which prehistoric men employed his tools.

The major merit of Semenov was that he developed a critical methodology for the observation and subsequent interpretation of wear traces, which he applied systematically in gathering information and testing hypotheses on tools use. Starting from experiments, he made important observations with regard to the cause of edge damage and drew guidelines to distinguish between traces originating from use and from manufacturing. Moreover, he noticed a relationship between the degree of wear of a tool and the motion or contact material it was used for. He emphasized, however, that this relationship could not be considered linear because of the numerous factors that he had found to be of influence on the vulnerability of a tool to wear. Other virtues of Semenov are that he demonstrated that implements of other stone types, like chert, quartzite, and obsidian could be analyzed for wear-trace occurrence as well, and that he introduced the micro photography in this field of research as a means to systematically document all microscopically observed information.

The work of Semenov has laid a firm basis for use-wear analysis and has made archaeologists much more aware of the potential of this method. Its potential was for instance illustrated when this approach became involved in the *Mousterian debate*. This debate focused on the meaning of morphological variations between Palaeolithic artefact assemblages. The French archaeologist François Bordes was convinced that differences between assemblages in the occurrence of tool types could be interpreted as an indication of cultural change. He believed that use-wear analysis would be a meaningful aid for complementing morphological typologies but could not be helpful in explaining differences between lithic industries (Bordes 1967). Semenov, on the other hand, was not interested in constructing functional typologies at all. Instead, he tried to discover what the tools were made for and how they were used. He argued that variation in the macromorphological aspects of artefacts can be no evidence for cultural change if it is not accompanied by a change of technology ánd of use (Semenov 1970). Since that kind of additional information could only be provided by methods like use-wear analysis, it became clear that this method could play a more crucial role in the reconstruction of prehistoric tool use than many people had initially realized. As Tringham *et al.* once put it: "*In fact it can be argued that detailed information on the usage of an artifact can add a completely new dimension to the potential of lithic analysis as an indicator of cultural change and variation.*" (1974: 173). It meant a means to translate static typologies into hypotheses on human behaviour.

One of the shortcomings of Semenov's work, however, was that he had not given 'real' evidence for his interpretations, notwithstanding the fact that they had been based on the results of twenty years of microscopic and experimental research. It was argued that his interpretations had been based on insufficient experiments and that he had omitted to make a comparison of his archaeological traces with the experimentally obtained traces (see also Hayden & Kamminga 1979). The reason for this was that Semenov mainly had been interested in the technological aspects of stone tools rather than in exact reconstructions of tool use.

He was, therefore, not really to blame for his supposed shortcomings. In fact, his work created a demand for more research and inspired others to work on the construction of a theoretical and methodological framework.

Especially European and North-American archaeologists have been elaborating several of the aspects that were discussed in 'Lithic technology'. In particular Tringham *et al.* (1974) and Keeley (*cf.* 1980) have made noteworthy contributions to the research on the supposed relationship between activities and the subsequent damages. Their studies were based on extensive experimental programmes. The work of Tringham *et al.* concerned the assessment of edge-damage formation in relation to factors such as action, worked material, edge angle and grip, while Keeley described the sometimes characteristic patterns of wear that different materials would cause. He demonstrated the occurrence of similar patterns on prehistoric artefacts. The studies of others have been dedicated to the variability of wear patterns in relation to raw materials (*e.g.* Greiser & Sheets 1979; Beyries 1982), to the wear characteristics of stone types like basalt (Plisson 1982), quartzite (*e.g.* Sussman 1985; Knutsson 1986, 1988), obsidian (Hurcombe 1985), etc., to the mechanisms of wear development (*e.g.* Kamminga 1979; Anderson 1980; Andersen & Whitlow 1983; Mansur-Franchomme 1983) and to standardizing nomenclature, observation and recording (see Hayden 1979).

For several years now, the analysis of wear traces has established a firm place in archaeological research. Recent publications (*e.g.* Anderson *et al.* 1993) show that this method is being applied on an intensive and worldwide scale. Nevertheless, there is still a lot of research going on, both on its methodological and technological aspects.[2] For example, efforts are being made to refine and expand the experimental and ethnographic evidence for stone function (a.o. Keeley 1983; Owen 1993), to learn more about the process of polish formation (Yamada 1993), to trace the diagnostic values of single wear characteristics (*e.g.* Van den Dries & Van Gijn, in press), to gain knowledge about the influence of post-use and post-depositional processes on the development of use wear (*e.g.* Shea & Klenck 1993), and to develop new methods that enable more objective interpretations of wear traces (see paragraph 4.5 and chapter 5).

4.3 Methodical aspects

4.3.1 COMBINATION OF INFORMATION SOURCES

Use-wear analysis of prehistoric flint artefacts consists of the observation and subsequent interpretation of wear traces on stone artefacts, for the purpose of reconstructing their function. Apart from an optical investigation of the traces on the surface of the artefact, this involves also the morphological aspects of an artefact, experimental and ethnographic evidence, and sometimes even chemical analysis. Whereas the ethnographical information can give insight into the variety of tasks an archaeological implement may have been employed for, the experimental information offers a possibility to compare the traces on the archaeological implements and to deduce the materials that may have, or reversely, cannot have caused them.

The morphological aspects of an implement, such as the shape of its edge, can give an indication of its function or for the exclusion of functions. For instance, a blunt or concave shaped tool can be considered unsuitable for activities like cutting or boring. However, this must be done with care, because lithic analysts have frequently experienced that a tool's morphological characteristics may get them on the wrong track in deducing its supposed function. It has been demonstrated that there is no simple and constant relation between the shape of a tool shape and its function (*cf*. Juel Jensen 1988). Therefore, morphological characteristics can merely render a rough functional indication and may either confirm or contradict a hypothesis that is based on the other information sources.

A third source of information that is involved in a functional interpretation are the traces that the applied activity may have left on a tool and which can be investigated optically. Depending on the type of damage the tool incurred, different types of microscopic equipment can be utilized (see section 4.3.3). In comparison with the morphological characteristics of a tool, wear traces may give more detailed and reliable information on a tool's prehistoric function. It has actually been shown by blind tests with experimental tools, that it can be determined with high accuracy whether a tool has been used or not (*cf*. Keeley and Newcomer 1977; Odell & Odell-Vereecken 1980; Gendel and Pirnay 1982; Unrath *et al*. 1986; Bamforth *et al*. 1990). Moreover, the utilized part of a tool can be identified with similar precision (*e.g*. Tringham *et al*. 1974; Unrath *et al*. 1986; Grace 1989: 134). The interpretation of wear traces does, however, not have to be confined to these two aspects. Often analysts can even infer the actual function of an artefact, *i.e*., on which contact material it has been used and in which motion (see also chapter 7).

Unfortunately, it is not always possible to apply use-wear analysis for a detailed functional interpretation. Many archaeological tools do not show any trace of use, whereas others show wear that cannot be ascribed to a particular activity: they may be caused by factors other than use, they may not be diagnostic enough, they may be poorly preserved due to post-depositional processes, or they may not resemble any of the experimentally obtained traces from the reference collection. Despite the fact that this kind of wear may exclude an interpretation on the level of the applied contact material or motion, it may still enable to verify whether a tool has been used or not.

In case of absence or irreducibility of wear traces, a fourth source of information may be helpful in pertaining a functional interpretation, *i.e*., (organic) *use residues*. It is possible that the use of an implement did not cause damage, for instance due to a short duration of the activity, but that it has left some residue on the tool. The analysis of such residues, which originates in criminological research (see Briuer 1976), can be performed with optical (*e.g*. Shafer & Holloway 1979) and (bio)chemical means (*e.g*. Briuer 1976; Loy 1993). Repeatedly it has been shown that they can render important and unique information on prehistoric tool function (see for instance Anderson 1980; Fullagar 1988). In some cases residue even allows for the analysis of prehistoric DNA (*e.g*. Loy 1993).

Like the analysis of use damage, the analysis of residue has its limitations and interpretation related difficulties as well. It may, for instance, be laborious to determine whether the observed residue can indeed be ascribed to use or rather to contamination, for instance by the soil in which the implement was lying for hundreds or thousands of years (Hurcombe 1986; Loy 1993). Concerning optical analyses of residue it may also be difficult to separate residues from wear traces (Anderson 1980). Moreover, (bio)chemical analyses may corrode the sample or its unique character (Loy 1993) or may not be possible at all due to the minuscule amounts of the traces. The main limitation, however, is that residues are only occasionally preserved on prehistoric tools. Whereas wear traces are already vulnerable to obliteration, residues are even more easily destroyed because they quickly dissolve in a humid or acid environment.

Despite their difficulties and limitations, wear-trace and residue analysis are crucial sources of information regarding functional interpretations. They are means for retrieving functional and environmental information that archaeologists cannot obtain otherwise. Especially a combined use of both methods of analysis can yield unique information on subsistence and craft technologies in the stone age: there are numerous materials that are rarely preserved in archaeological context but, fortunately, which have often left their traces on stone tools.

4.3.2 USE-WEAR PHENOMENA

The type of analysis that was focused on for the development of the knowledge based application that will be discussed in chapter 5 is wear-trace analysis. Residues have not been

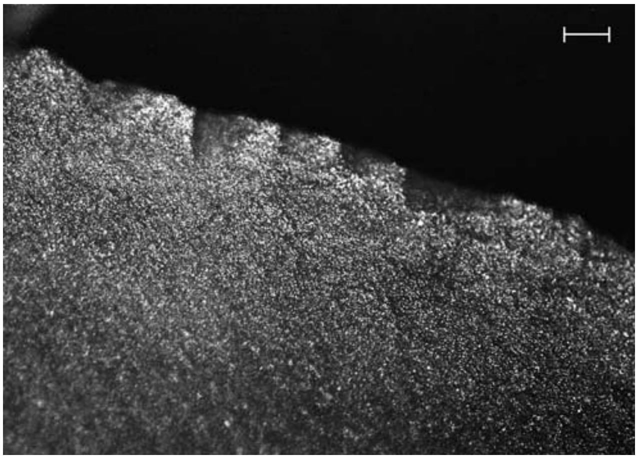

Fig. 5. Use retouch.
(scale bar equals 50 micron)

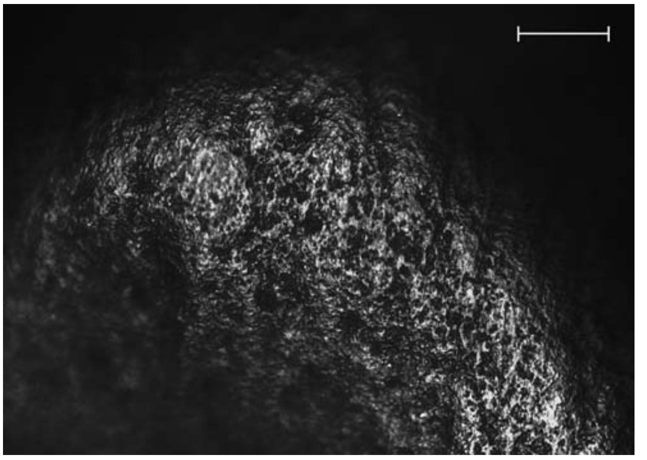

Fig. 6. A rounded and polished edge.
(scale bar equals 50 micron)

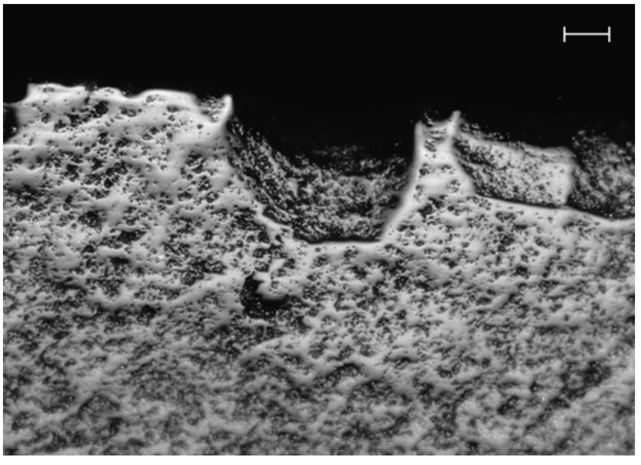

Fig. 7. Polish.
(scale bar equals 50 micron)

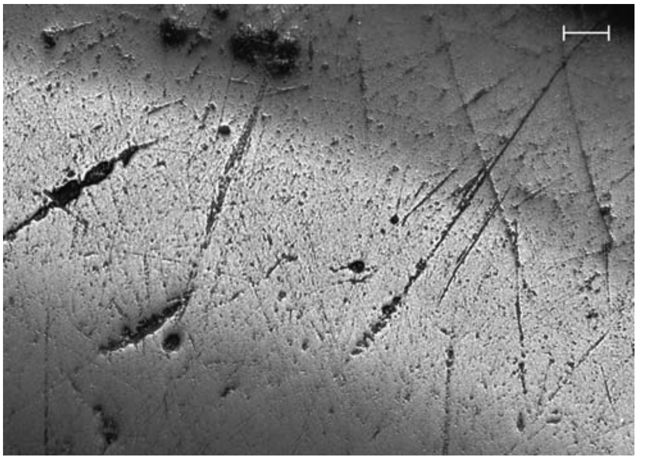

Fig. 8. Striations within a polish.
(scale bar equals 50 micron)

incorporated. The traces of use that micro-wear and macro-wear analysis involves are *edge removals*, *edge rounding*, *polish*, and *striations* (fig. 5-8). Edge removals or micro fractures form due to (too heavy) pressure on a tool's brittle edge. It is assumed that specific patterns of micro chipping are diagnostic for the hardness of the worked material and the applied motion (*cf*. Tringham *et al*. 1974; Lawrence 1979; Odell & Odell-Vereecken 1980; Odell 1981). Others, however, have emphasized the difficulties of interpreting edge removals. For instance, Moss (1983b) and Vaughan (1985) showed that single retouch features, such as their location or distribution, often do not correlate with specific motions or materials, and that the morphological aspects of a tool influence scar development as well. After an extensive study on this issue, Cotterell and Kamminga (1987) concluded that "*...the mechanisms responsible for flake formation will not always be apparent*." (*ibid*. 1987: 704). Furthermore, it has repeatedly been stressed that non-use factors such as post-depositional processes complicate the analysis of edge removals, because they may be to a large degree responsible for their development (*e.g*. Lévi-Sala 1993; Shea & Klenck 1993). Other problems with the analysis of use retouch concern the representativeness of the traces and the preciseness of the interpretations. With reference to the representativeness it has been demonstrated by experiments that even after considerable use on relatively hard materials like wood or bone, implements do not always sustain damage (Van den Dries & Van Gijn, in press). The damage that did develop was found to be more diagnostic for the hardness of the worked material than for particular contact materials (*ibid*.). Retouch patterns mainly give an indication of the hardness of the contact material and in some cases whether it is vegetal, animal or inorganic (Shea 1988). As a consequence, edge damage does not provide as precise information as may be wanted. One of the reasons, however, that some analysts are clear advocates of the analysis of these macro traces is that it allows for rapid functional analyses and, therefore, it is possible to process large samples in limited time periods (*ibid*.) with less expensive equipment (Odell & Odell-Vereecken 1980: 88).

A second category of use wear is edge rounding. The absence or presence of edge rounding, and especially its degree, is also considered to be an indication of the hardness of the worked material. Experiments have repeatedly demonstrated that hard materials hardly allow an edge to round, because they cause a continuous edge scarring that obliterates all (if any) previous rounding. Soft materials, on the other hand, are often too yielding to cause an edge to round. Thus, if severe edge rounding originates from use, it is usually caused by materials of medium hardness, like wood, dry clay and soil (see also Van den Dries & Van Gijn, in press), or by working dry hide or hide polluted with abrasives like powdered ochre (Van Gijn 1989: 28-29).

A third use-wear phenomenon is polish. Many analysts consider this to be the most diagnostic wear category. Vaughan once defined polish as "..*an altered flint surface which reflects light and which cannot be removed with acids, bases and solvents*." (Vaughan 1981: 132). Since it has been proved that the latter is not always true, Van Gijn proposed to add 'weak' before 'acids, bases and solvents' (Van Gijn 1989: 5). It is known that a polish constitutes a change of the topography and reflectivity of a stone surface. Moreover, different activities may cause specific wear patterns that in most cases can be replicated in experiments. But it is still unknown what exactly causes polish formation and its apparent material-related variation. Therefore, research on the process of polish formation remains an interesting issue (see for instance Yamada 1993).

The advantage of an interpretation based on polish as compared to one based on retouch, is that it can yield more detailed information. While retouch patterns indicate the relative hardness of the contact material, polishes can give information on the level of the exact worked material (hide, bone, soil, etc.). Another advantage of polish is its tendency to develop somewhat more frequently than edge rounding and use retouch. This means that under optimal circumstances, use polishes may be slightly more representative for prehistoric activities (Van Gijn 1989: 50; Van den Dries & Van Gijn, in press).

A fourth wear category are striations. These are grooves and scratches within a polished surface. They were already noticed by Semenov, but their diagnostic value for functional analyses is still being discussed. A complicating factor is, like with polish, that it is unknown how exactly they develop. Despite the fact that the formation of striations has also been the subject of detailed research, clear relations between specific contact materials and the occurrence or characteristics of striations have not been found (*e.g*. Del Bene 1979). One of the suggestions is that they are mainly caused by abrasive particles or grit that are being rubbed across the surface of the tool while it is being used (Lévi-Sala 1993). However, there is no decisive definition and striations are, therefore, not always recognized on either archaeological or experimental tools (*cf*. Moss 1983a; Van Gijn 1989). It is especially difficult to distinguish striations from other linear features within a polish. For these reasons, most wear analysts consider these scratches merely as an indication for the direction in which the tool was employed. It is more or less generally accepted that the direction of striations corresponds with the applied motion. For instance, an edge showing striations that run parallel with it, probably was used in a longitudinal movement, while perpendicular oriented scratches indicate to a transverse motion.

4.3.3 MICROSCOPY

Wear traces can be investigated by means of a *stereo microscope* and an *incident-light microscope*. Usually, a stereo microscope is used for magnifications up to 150×, and the incident-light microscope for magnifications up to 560×, although this also depends on the preference of the individual analyst. Working by means of the former is called the *low-power approach* and the latter method is known as the *high-power approach*. The expressions *macro-wear* and *micro-wear* analysis mirror the difference between these approaches. In most case, magnifications up to 150× are employed to locate wear traces. The interpretation of the traces, however, usually requires a more detailed view. It is based on specific characteristics that can only be observed by means of high magnifications. Only with regard to edge removals, it may be sufficient to use a low-power microscope both to identify their presence or absence and to infer a functional interpretation, because the overall pattern usually reveals more than of single scars. However, these gross patterns usually do not show enough detail to allow for an interpretation of the polish.

Apart from the above mentioned devices, analysts sometimes use very high magnification equipment, such as a *scanning electron microscope* (SEM). This enables detailed observations with magnifications of up to 400.000×, which are for instance required for research on polish formation (a.o. Kamminga 1979; Anderson 1980; Mansur-Franchomme 1983; Yamada 1993). However, this instrument does not belong to the analyst's basic equipment, due to the financial aspect and to practical, time-consuming problems, like the necessity of coating.

	scanning	interpretation
retouch	stereo 40× incident light 100×	stereo 100× incident light 100×
polish, edge rounding and striations	stereo 40-160× incident light 100×	incident light 200-400×

Fig. 9. The microscopic equipment and magnifications that are used at Leiden University for a first scan of a tool and the subsequent analysis for an interpretation.

At Leiden University, the presently followed method consists of the analysis of all four wear categories. A tool is first being scanned for traces by means of either an incident-light microscope or a stereo-microscope, but the magnifications are usually less than 100×. Subsequently edge removals are analyzed by the same magnifications, but edge rounding, polish and striations are interpreted by means of high magnifications (200×) of the incident-light microscope (fig. 9). Occasionally a magnification of 400× is used. Edge damage and polish are considered equally diagnostic and are used as basic information sources. Information concerning edge rounding and striations is only used additionally. In addition to these aspects, the morphology of a tool is studied as well. However, this is also merely used as additional information. In Leiden, the main expertise lies in analysing implements of the flint type. Since 1995 residue analysis is being applied as well.

4.4 Difficulties encountered

Although the interpretation of each wear category has its specific difficulties, I will concentrate on the methodical problems that use-wear analysis has hitherto faced. The difficulties will be discussed because they have affected the way WAVES has been constructed to a large extent (see chapter 5). Moreover, it will also illustrate which problems still exist and the effect that they have on the functionality of WAVES. The difficulties that will be discussed concern the polemic between the advocates of the high-power approach and those of the low-power approach, methodical inadequacies that caused disappointing blind-test results, the interpretation of overlapping or non-diagnostic wear patterns, the representativeness of wear traces, and the subjectivity of the observation and subsequent interpretation.

Based on the achievements of Semenov, wear-trace analysis was considered to be a promising method and several archaeologists started using it. Unfortunately, few of them awaited the constitution of a sound methodical base or tried to build one themselves. In 1974, after ten years of research since the appearance of Semenov his book, Keeley expressed his concern on the technical and methodical state of the wear-trace analysis of that moment. He was convinced that the disappointing productivity and validity of many studies were primarily due to a combination of unnecessary technical omissions and methodical inadequacies of the investigators themselves (Keeley 1974). As a guideline for future research he therefore urged for systematic testing of interpretations by means of an experimental or ethnographic framework relevant to the archaeological site being studied, for studies on wear-obliterating effects and for more serious quantification attempts.

In the same year, Tringham *et al*. (1974) published the first study on the formation of use retouch that was based on systematic experimentation. The aim of their work was to deploy their descriptions to the interpretation of edge damage that is analyzed by means of low magnifications. Although they clearly described the relation between the observed wear patterns and the worked materials and applied motions, they did not actually demonstrate it their selves, nor did they involve the aspect of representativeness of the replicated patterns in their research. It was not until 1980, that the conditions and results of a large experimental programme were described that focused on use polish and the high-magnification approach (Keeley 1980). From the outcome Keeley concluded that, indeed, there is a high correlation between the appearance of particular types of polish and materials that are being worked.

Although Keeley said to have demonstrated the folly of relying on one class of wear phenomena only (Keeley 1980: 176), from that moment on the polemic between the high-power approach and the low-power approach augmented. Each camp had its advocates and there was a mutual disbelieve in the claimed potential of these methods. For instance, Odell & Odell-Vereecken (1980: 89) stressed that the low-power method could in most cases answer questions equally precise as the high-power method, though faster, and that it could be applied to a greater variety of raw material types. The high-power camp, on the other hand, did not agree with the supposed equivalence of interpretations. Moss, for instance, stated that "*In most cases the force necessary to produce only microscopic edge damage is so minimal as to be completely undiagnostic*." (1983a: 231), and that "*The edge damage which can only be seen properly at 75-100× magnification is usually meaningless unless accompanied by polish or striations*." (*ibid*.).

In the mid eighties, the debate culminated when two blind tests were presented at the Teubingen conference of 1985. They lead to conflicting conclusions on the potential of the high-magnification method for the analysis of polish. Unrath *et al*. acknowledged the need for much more research, but were positive about the results that could already be obtained at that point. They were also confident about the progress that was lying within reach. Oppositely, Newcomer *et al*. (1986) were clearly disappointed in the outcome of their blind test and seriously doubted the interpretative abilities of the high-power approach. This started a severe methodical discussion (Moss 1987; Bamforth 1988; Hurcombe 1988; Newcomer *et al*. 1988), which continued for several years. One of the results of this discussion was that eventually functional analyses became more explicitly based on the entire range of sources, including tool morphology, edge damage, polish etc. In fact, few analysts are nowadays clear advocates of one of both approaches. The majority seems to deduce its interpretations from the combination of macro- and micro-wear evidence (*cf*. Van Gijn 1989; Grace 1993).

Although use-wear analysis does no longer suffer from crucial methodical struggles, there are still various difficulties that need to be surmounted. To begin with, many tools cannot be functionally interpreted because there are several factors that affect the visibility of the apparent traces. For instance, use polishes are often altered or even obliterated by post-depositional surface modifications like patina, or bright spots (Stapert 1976; Lévi-Sala 1986, 1993). But apart from the difficulties concerning the 'readability' of the traces, their interpretation may in some cases be problematic as well.

First of all, it may be difficult to relate traces to specific activities. It is, for example, not always possible to distinguish traces caused unintentionally by use from those caused intentionally due to manufacture or from those resulting from 'natural' processes such as transport (bag carrying) or post-depositional processes. Not just macroscopical damage, but edge rounding, striations, polish and residue may all relate to non-human activities.

Another problem is that of non-diagnostic wear patterns. This means that they are not exclusively representative for one particular activity or contact material. Keeley (1980) and many others have shown that there is not always a direct correlation between a performed activity and the resulting traces, but that different materials may cause similar traces, and *vice versa.* The cause of this variety is not very clear.

It is known that the development of non-diagnostic traces often relates to the duration of an activity or to the softness of the contact material. It has, for example, been experienced that if a tool is only briefly in contact with a material that the resulting damage usually does not reach a well-developed stage.[3] But the observed variety probably also has to do with the fact that the development of wear traces is a more complex phenomenon than it may seem at first sight. The characteristics of wear traces can be highly affected by factors other than the resistance of the worked material, the applied motion and the duration of work as well. For example, the quality of the raw material of the implement, the robustness (angle) of the edge, the angle in which the edge is held during work and the rate at which the edge is loaded may also play a determinate role. Furthermore, a tool may have been employed for a variety of tasks. This can cause physically overlapping traces, which may result in a non-diagnostic wear pattern.

The phenomenon of the non-diagnostic traces may cause serious difficulties for the functional interpretation of archaeological implements. Since yielding materials will often fail to leave diagnostic traces, these materials are underrepresented in the interpretations of archaeological tools (*cf*. Shea 1991; Van den Dries & Van Gijn, in press). That is why the interpretation of less diagnostic traces has received serious attention during the development process of WAVES (see chapter 5).

Quite a different problem is that of characteristic but unreproducible traces. In some cases there are no equivalent experimental traces of archaeologically encountered patterns. Therefore, the materials that are responsible for them have not yet been discovered (see Van Gijn 1989; Juel Jensen 1989; Schreurs 1992). With respect to such implements, the only thing which can be inferred with some certainty is the function that they certainly had not and the motion with which they have been employed. Two frequently observed mysterious polishes have been incorporated in WAVES.

A final problem that has been focused on during the construction of the knowledge based application is the subjective nature of wear-trace analysis. In fact, it has been one of the main reasons for its development. Analysing traces of use depends on a visual observation that is subsequently interpreted by means of analogies with traces of known origin. Consequently, this involves a considerable degree of subjectivity on three levels. First, the analyst makes a subjective observation of the wear features. He or she then describes the observations by means of his or her own subjective and qualitative terms and, finally, draws a conclusion on the basis of their personal reference collection of experiments. The subjective nature of this procedure has been considered to be problematic for building a solid foundation because it affects the comparability of analysis results of individual analysts (*cf*. Grace *et al*. 1987). Additionally, the lack of objective descriptions and interpretations lengthens the process of acquiring the ability to recognize and interpret wear traces for apprentices.

4.5 From a qualitative to a quantitative method?

4.5.1 INTRODUCTION

Various researches have hitherto been carried out in an attempt to obviate the difficulties that were discussed in the former paragraph. In particular, minimizing the subjectivity of the method has turned out to be one of the most challenging aspects of use-wear research and has resulted in all kinds of suggestions. This paragraph will be confined to these suggestions.

Attempts to improve the objectivity of wear-trace analysis are nearly as old as the method itself. For instance, Semenov (1964) already emphasized a systematic collection of evidence. He probably expected academic resistance towards the subjectivity of his approach and urged for a systematic documentation of observations by means of micro photography. In fact, photography was an utterly fruitful first attempt to obtain objective evidence. Up till this moment it is one of the most important ways for communication between use-wear analysts. It enables comparisons of interpretations of different analysts and it is very useful for training students. But photography alone was not enough. In order to enable the comparison of different wear traces on objective grounds, an objective classification was needed.
Basically, all subsequent studies on the objectivity of the method aimed to obtain a more objective *identification* or *determination* of wear patterns or polish 'types' that could replace a subjective *interpretation*. With reference to the three levels of subjectivity of the analysis procedure different approaches have been applied. A first line of approach concerns attempts to establish an objective method of the first step in the analysis procedure, *i.e*. the observation of the traces (paragraph 4.5.2). This has been tried by means of optical and mechanical devices, and by means of image processing and pattern recognition techniques. A second line of approach is directed towards a more objective interpretation of archaeological wear patterns by building a solid quantitative framework of experimental data (paragraph 4.5.3). This not only allows for a better understanding of the circumstances that cause wear to develop but, if the framework serves as a formal reference collection for the interpretation of archaeological wear, it may yield more controlled and objective interpretations. A third line of approach focuses on a formalization and standardization of the interpretation procedures by simulating the inferencing process by means of artificial intelligence methods (paragraph 4.5.4). These three lines of approach will be exposed in more detail in the following sections.

4.5.2 AUTOMATION OF THE OBSERVATION PROCESS

To my knowledge, attempts to identify wear traces by means of the automation of observations have up till now been directed to polish only. There seems to be less priority to incorporate use retouch and edge rounding in this line of approach. Probably, the reason for this is that the analysis of flakes is somewhat less subjective than that of polish: flakes can be counted and the descriptions of shapes and terminations are already fairly well defined.[4] Moreover, compared with polish it is more difficult to obtain an image that represents the variety of the scars on an entire edge. Often the distribution of scars is less uniform than that of polish. It is therefore important that especially use-retouch interpretations are based on the analysis of the entire edge of a tool.

Keeley (1980) was one of the first to stress the need for a more objective method of polish observation. He investigated the distinctiveness of experimentally obtained polishes by means of *light reflection measurements*. He equipped a microscope camera with a light meter to measure the average amount of light that various polishes reflected. In fact, he measured their *brightness*. Unfortunately, he could not compare his measurements with each other because he could not use a standard area of measurement due to differences in polish extensions. In a similar manner he also tried to measure polish *roughness*, but encountered the same problem.
Many attempts followed in the line of Keeley and gradually more sophisticated instruments were deployed. While Keeley simply measured reflectivity, Dumont and Bauche tried to obtain a three-dimensional picture of a polished surface. The former did this also by means of an optical measurement, the latter by means of a mechanical measurement. Dumont (1982a,b) applied the industrial optical technique of *interferometry*. With this technique small variations in the distance between two surfaces can be measured, for instance between the surface of the flint and the surface of its polish. This enables the measurement of the physical dimensions of the features that are found on polish surfaces, such as striations, depressions, ridges, protuberances, etc. From such measurements an interference pattern can be made, which resembles a contour map. Finally, by counting the light and dark bands of such a map quantitative data are obtained.
The basic idea of applying this technique was that it would enable a quantitative assessment of polish morphology. It would show the existence of consistencies in the peculiarities of the observed phenomena and could eventually be used to differentiate between polishes. However, it turned out that one of the limitations of interferometry is that the measured object must have a regular surface and that it could not be applied to unused flint or to implements with irregular surfaces (Dumont 1982a: 209). This implies that variations in grainsize of different types of flint may be a complicating or even a problematic factor. An even more important drawback is that the quantitative data that this technique yields, could not simply be extrapolated. It was not explained, for instance, how these contour maps could be translated in terms of activities responsible for the particular contours. Consequently, it merely provided a method to study mechanisms of polish development, but was not experienced as a practical aid to support the analyst in interpreting wear traces.
Bauch (1986) suggested an alternative method. He proposed to apply an electrical contact-instrument (*perthometer*) to measure surface roughness mechanically rather than optically. The idea was that the needle of the device was led across the surface of a polished stone tool, that the vertical movement of the needle would be digitized and the subsequently registered

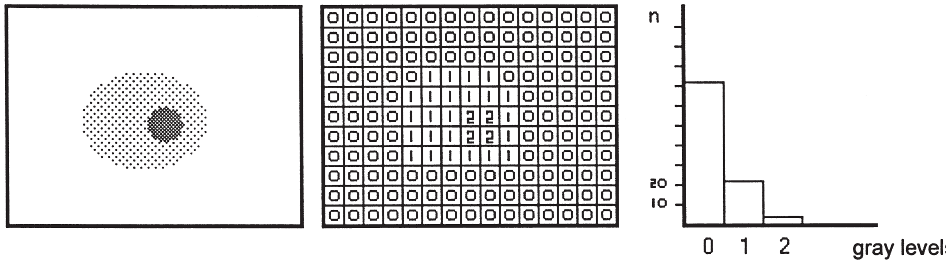

Fig. 10. Simple image processing approach: translation from a digitized image via a data matrix into a grey-level histogram.

contours would be translated into a *roughness profile*. Bauche had applied this technique in a study of surface variations between mill- and grindstones, and thought that it could be of benefit for polish analysis as well. To my knowledge, he did however not succeed in measuring polish variation mechanically.

It can be doubted whether mechanical measuring is applicable to polish analysis. Irrespective of the problems of interpretation and correlation of the obtained contour maps, the actual registration of roughness would surely be problematic or even impossible. One would need extremely sensitive equipment to measure the minute unevenness of polishes mechanically. Polish is in most cases only visible at high magnifications (200-400×) and its extension onto the surface of the tool is measured in microns. Moreover, the problem Keeley had experienced would be encountered again: the extension of the polishes varies too much between individual tools to allow for a comparison of measurements.

Eventually, the mechanical approach was not pursued, but the optical approach evolved further. For instance, Grace *et al*. (1985, 1987) started to apply a computerised *image processing* technique. This image processing approach consisted of four steps. First a microscopic image was digitized. This can be done by scanning an image by means of a video camera that is attached to a microscope and a computer, or by digitizing a micro photograph. The digitized image was subsequently translated into a grey level histogram. Such a translation can be done according to various calculation methods, but each of these concentrates on one particular aspect of the image. For instance, one method shows a plain distribution of light and dark tones (fig. 10), while another highlights the relief within the image. The next step was to apply a mathematical normalization function to the histograms in order to eliminate the effects of lightning conditions and photographic and digital processing, and to increase the contrast between discrete features. There are various normalization functions available. Grace used co-occurrence matrices (*e.g*. Grace 1989). Finally, the histograms of different images were compared with each other and statistical functions were applied of which the results were presented in a scatter diagram. These diagrams visualize the differences or similarities of the images.

The image processing approach was a promising innovation but did not yield the expected results. Like Keeley, Dumont and Bauche, Grace *et al*. kept on focusing on the analysis of the *texture* of polishes (Grace *et al*. 1985, 1987) and they could not establish significant differences between the textures of the polishes that resulted from different contact materials either. Unfortunately their disappointing achievements not only made them abandon texture analysis as a satisfactory means to identify polishes but it temporarily placed use-wear analysis in a awkward position as well. They extrapolated that the difficulties they had encountered would cause inaccuracies in ordinary optical identification of polishes and they claimed to have demonstrated that human analysts would not be able to deduce contact materials from polishes (Newcomer *et al*. 1986: 216).

Despite these conclusions and Grace's rejection of this type of texture analysis for polish identification, others continued in this line. For instance, Knutsson *et al*. (1988) applied the same technique for the purpose of building a classification of different polishes, which they used for interpreting a Middle Neolithic settlement assemblage. Their conclusion was somewhat more optimistic than that of their predecessors. They had experienced that texture measurements could indeed distinguish variations in the analyzed images. Concerning the archaeological material, the texture analysis could not yield more information on the identity of the worked materials than a non-automated analysis.

Another example of optical texture-analysis is a study of Rees *et al*. (1988), in which Grace was involved as well. This time they applied the technique of *fractal geometry*, which was borrowed from geology. According to Rees *et al*., "*Fractals are essentially spatial distributions or patterns which possess self-similarity – i.e. there is a statistical equivalence between small-scale and large-scale fluctuations in these patterns*." (1988: 177). A coastline is a classical example of a fractal. Fractal analysis is a statistical measurement of the self similarity of a digitized image and was believed to be a suitable technique for calculating surface roughness. The expectations were high because the basic principle of fractal analysis corresponds with the human perception of natural texture. If a quantitative link between contact materials and use wear could be proven, "*...this would support the view that contact-material can be determined from visual examination of microscope images alone*." (*ibid*.: 183). Indeed, they found that micro-wear images tended to structure in a fractal manner and that the measurements of polished and unpolished surfaces differed significantly. It must be stressed, however, that this outcome was based on merely seven images and that correlations between fractal dimensions and contact materials were not found. Even though further research on the issue was announced, nothing has been heard from it since.

Although the analysts all employed other techniques or variations of the same, it is clear that they have all been wrestling to eliminate the effects of measurement disturbances and that none yielded the expected result. Nevertheless, the image processing attempts have certainly not been abandoned yet. For instance Shea recently proposed to measure and quantify polish variation by means of luminance profiles (Shea 1992). This approach would also consist of calculating grey tones of digitized images, but the difference with the method followed by, for instance, Grace *et al*. (1985, 1987) is that the measurement entity is a transect across the edge instead of a single spot.

Furthermore, Vila & Gallart (1993), Yamada (1993) and Bietti *et al*. (1994) have been working on the improvement of the image processing approach by making adjustments of the standard procedures. But their suggestions were also small variations on the same theme, and major advances have still not been accomplished, despite the fact that image processing techniques have evolved considerably.

4.5.3 QUANTIFICATION OF RECORDINGS

Simultaneously with the above attempts to develop devices for objective polish observations, it was tried to build a representative and quantitatively underpinned experimental framework to allow for a more objective interpretation of subjectively observed phenomena. Many analysts considered the quantification of experimentally obtained wear pattern information of both polishes and use retouch as an absolutely vital step towards a more objective method. For instance, Keeley urged for more serious attempts to quantify micro-wear data (1974: 332, 1980). He was one of the first to systematically quantify the occurrence of both macro- and micro-wear in relation to different contact materials and to the morphological aspects of the implement (Keeley 1980). Two of his conclusions were that both high and low magnifications are required to study implement function, and that his frequency tables had made clear "…*that many variables other than the material worked and the method of use [...] will affect the type and size of the utilization damage formed*." (*ibid*.: 83). Furthermore Ahler (1979) quantified the edge damage on 140 end scrapers and successfully applied multivariate analyses to distinguish functional variation between the artefacts.

Also in this line, Keeley was followed by several scholars. For instance, Akoshima (1987) gave an extensive quantitative description of flake pattern variability. He counted and measured 3840 experimentally obtained scars and by means of simple calculations of percentages he demonstrated significant correlations between the observed patterns and the performed activity.

Due to the increasing number of quantification of the recorded wear-patterns, it became clear that the overlap of wear patterns is a truly complicating factor: few materials produce exclusively diagnostic wear attributes (*e.g*. Keeley 1980; Vaughan 1985; Van Gijn 1989; Van den Dries & Van Gijn, in press). It also became clear that neither the low-power nor the high-power approach yields the best results. The quality of the interpretations depends heavily on the quality of the traces, which varies from tool to tool and from assemblage to assemblage. Moreover, it was noticed that the traces of particular materials (such as meat) are most probably underrepresented in the archaeological record, because these materials damage an implement less frequently.

In conclusion, the quantification of the recorded wear-phenomena has yielded important information on wear variability and representability and allows for more objectively based interpretations of archaeological wear traces. There is, however, one important limitation of this approach that should be kept in mind: it may not always be possible to carry out comparative studies of the quantified data of different analysts, because of the fact that their observations and subsequent recordings of the wear phenomena remain subjective. This means that one does not know whether the data are fully comparable. One analyst may describe a polish as 'bright', and the other may call it 'very bright' (see also chapter 7).

4.5.4 AUTOMATION OF THE INFERENCING PROCESS

The third line of approach is the youngest. It is directed towards the deduction of interpretations and attempts to obtain objectivity by formalizing and standardizing the analysts inferencing or reasoning processes. This can be done in different ways, but basically it means that a computer is programmed to simulate (parts of) the interpretation process. The difference with the image processing approaches as discussed in paragraph 4.5.2, is that the interpretation is still based on a subjective description of traces rather than on a purely objective observation.

An automated interpretation can be obtained by means of artificial intelligence techniques, such as *expert systems* and *neural networks*. The expert system approach implies that the line of reasoning of the analyst is made explicit and translated into a computer program. Subsequently this program can be used to give an interpretation of descriptions of wear (see chapter 5). A neural network, on the other hand, does not follow the analysts method of inference, but its own (chapter 6). It is furnished with a statistically based inferencing method, which automatically matches the observed phenomena to an interpretation. The reference collection of a neural network simply consists of examples of observed phenomena and the activity that caused them.

Since these artificial intelligence approaches have been designed to handle complex, non-linear, heuristic expert-knowledge, they seem to offer interesting possibilities for use-wear analysis. Grace was the first to apply a knowledge-based method, *i.e*. an expert system application, for use-wear analysis (Grace 1989, 1993) and his attempts have not been without success. He developed a program that yielded encouraging results in a small test case (Grace 1989: 223). Grace demonstrated that in principle the inferencing process of wear-trace analysis could be formalized and that this approach could indeed be helpful to obtain more standardized interpretations. The recent development of the neural network technology has not yet been thoroughly tested, but may give another impulse to the attempts to formalize the inferencing process. Since a neural network uses a statistically based inferencing process rather than a line of reasoning that is deduced from a human, it may offer an even more objective approach than an expert system.

One of the advantages of both a neural network and an expert system application is that, in contrast with image processing techniques, they can be more easily applied by analysts and students. Moreover, expert systems have a larger educational value, because they are able to make the inferencing process that is involved accessible and assessable. For this reason, they can be employed to support the training process of a student. Nevertheless, it must also be kept in mind that the potential of these artificial intelligence approaches for use-wear analysis depends on the achievements that will be obtained with the other two lines of approach, especially with the attempts to lower the degree of subjectivity in the observations of wear traces. Since the results of automated inference processes are dependent on data which can only be obtained and recorded through subjective ways this will remain one of the main limiting factors that keeps use-wear analysis from becoming a quantitative method. This limitation will be encountered again in chapter 7.

4.6 Discussion

Each of the above lines of approach has yielded a positive contribution to the methodical evolution of use-wear analysis. Nonetheless, not one of them has proved to be the most suitable approach to develop use-wear analysis into a truly quantitative or objective method. The majority of the attempts have been directed towards the development of a device for polish identification by means of image processing, and this is also the line of approach that is tested best. As a whole, the image processing attempts have yielded less significant results than was expected. They have only acknowledged the fact that there is indeed a lot of overlap between the texture of polishes which originate from different contact materials and they have shown their potential for determining the presence of polish traces. Unfortunately, they have not yet been able to speed up the process of use-wear interpretation, nor have they complied with the prospect that they would support the learning process of apprentices (Grace *et al*. 1987: 69).

The image processing methods as they have been applied hitherto, turned out to have some major limitations. The most important problem lies in the fact that they focus on one or a few aspects of a polished surface, while a human analyst always bases an interpretation on a whole range of aspects. It has repeatedly been stressed by wear analysts that single wear attributes are not diagnostic for a particular contact material and in recent quantitatively oriented studies, this was once more confirmed (Yamada 1993; Van den Dries & Van Gijn, in press). By quantifying the presence and absence of particular instances of wear attributes on experimentally used implements, it was shown that the occurrence of almost none of the instances of the examined attributes relates exclusively to a particular contact material. In practice, this means that an image processing approach that is based on one attribute only, will fail to identify polish 'types'.

A second difficulty that needs attention is the fact that the image processing approaches all try to arrive at an interpretation on the basis of the analysis of a minute spot of polish: they process one or a few observations at a 200× or 400× magnification. Regarding the fact that the entire edge of a tool contains vital information (like the polish distribution pattern), the investigation of only a few micron can hardly be considered as an approach that is able to produce a reliable result.

A third problem concerns the archaeological applicability of the method. If it is impossible with the image processing technique to distinguish between clear and well-developed experimentally obtained polishes, it will certainly not be able to identify wear patterns which do not exactly resemble one of the patterns of the applied reference collection. In other words, the hitherto used techniques cannot yet be applied as a reliable method for the interpretation of archaeological polishes which are often not very well-developed or affected by post-depositional surface modifications.

Apart from the above mentioned problems, there is a fourth major drawback of the image processing applications. They are not very practical. It requires sophisticated equipment and skills not only to perform the analyses, but also to interpret the histograms or luminance profiles in terms of contact materials. It is therefore not very likely that all wear analysts are immediately willing to apply these kinds of techniques. These drawbacks or limitations do not mean, however, that the image processing studies have not been important. On the contrary, they have made the complexity of wear analysis visible and have made it clear that the complex nature of polish formation processes causes such variations, that it does not allow for the development of a polish identification device with the techniques that are presently available. This means that the ability of mortal humans — which is present without doubt — to distinguish between patterns visually, can not yet be simulated or even approached by artificial methods. In the eyes of some scholars this notion will certainly not constitute to the academic acceptance of use-wear analysis, because they believe that the human ability to recognize wear patterns can only be acknowledged if it has been proven empirically (see for example Rees *et al*. 1988). In the eyes of others this notion only confirms the opinion that these techniques cannot constitute satisfying alternative approaches as long as the issue itself is a poorly understood phenomenon (Juel Jensen 1988: 81).

When looking back at the three lines of approach that were discussed in the previous sections and when assessing their achievements, I am inclined to conclude that it will be very difficult to make use-wear analysis a truly objective or quantitative method by means of one of them. Even without running ahead of the results of a test with our expert system application and neural network prototype, that will be presented in chapter 7, it can already be concluded that it is impossible that one approach can single-handedly turn use-wear analysis into a truly quantitative method. This is not only a result of the fact that they all have limitations, but it is intrinsic to the fact that each approach covers only one of three subjective elements. Consequently, a combined effort will be required to break this impasse. Like a combined use of low-power and high-power microscopes improved wear-trace analysis, it may turn out to be equally fruitful to combine an image processing technique with the interpretational abilities of an expert system or neural network.

This implies, however, that research should continue on all three lines of approach. If one of them stays behind, the other will be hindered as well. For instance, it is of crucial importance for the improvement of the inference process that the quantification of recordings will be proceeded. Only by collecting more examples of wear patterns that result from various circumstances it may eventually be possible to map the variations and to unravel the mystery of the formation of wear phenomena and, subsequently, to build a knowledge base that truly and adequately covers the domain.

notes

1 This is a more or less academically accepted procedure, although it has also been questioned whether archaeologically found traces may indeed be directly compared with those experimentally obtained or ethnographically induced (see Gould & Watson 1982).

2 Technical aspects concern the actions and equipment involved in the observation (microscope) and recording (photography) of wear traces: methodological aspects relate to the interpretation of the observations.

3 When they concern use polish, badly developed traces are called ‘generic weak’ polishes (Vaughan 1985).

4 See for instance the *Ho Ho-classification and nomenclature committee report* in Hayden (1979: 133-135).