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

Dries, M.H. van den

acknowledgements

Several people have contributed to the realization of this thesis and I would like to thank them all.
First of all, I highly appreciate the effort made by Annelou van Gijn, José Schreurs and Channah Nieuwenhuis to initiate me into the world of use-wear analysis and to put their knowledge at the disposal of this study. Moreover, they kindly and patiently corrected the bloomers that I made, they checked the knowledge bases for biases and they helped to compose and to carry out the blind tests. On top of this all we had valuable and stimulating discussions.
Richard Fullager, Laurance Astruc, Xavier van Dijk and Serge Polman were brave enough to submit their selves to the blind test with WAVES. I kindly thank them for participating in the test and for giving suggestions to improve the application.
Milco Wansleeben shared his computer with me for several years and gave me hard- and software support. With pleasure I look back at the discussions that we had on all kinds of subjects concerning information technology. Martijn Bulk allowed me to use the scanning equipment of the National museum of Antiquities (RMO) in Leiden, Jan Pauptit reproduced the photographs, Henk de Lorm made the hand made drawings, and Peter Heavens produced some of the computerized drawings.
Finally, I would like to thank my family, friends and colleagues for their continuous interest in the progression of this thesis. Most of all, however, I am grateful to Harry Fokkens and Rustin Ramadhin. Harry not only commented on draft versions of my chapters, but also supported me both morally and practically. Thanks to their patience and willingness to share me with my computer, I was able to finish this thesis.

addendum
Welcoming WAVES

Annelou van Gijn and Richard Fullagar

1. Introduction

Usewear analysis of stone artefacts has come under fire from several directions, most often challenging the reliability of claims purporting to precisely identify what materials the stone tools actually processed in the past. Ironically, the strongest critics have been specialists in the field (Newcomer *et al.* 1988; Grace 1996). A central problem has been the lack of a definitive key upon which all analysts can agree, and which anyone could use to identify particular tool functions. Partially this is inherent to the method because there is an endless number of possible past activities, creating a large variation in wear traces. At the same time, traces from contact with different materials sometimes overlap in their constituent attributes, resulting in so-called and/or interpretations (see Unrath *et al.* 1986). This is especially the case with tools which have been used a short time; here diagnostic attributes fail to develop. Wear traces also alter through time, the character of the postdepositional alterations varying according to the matrix an artefact is buried into. Because of the lack of a definitive key, analysts have developed different systems of both recording and interpreting microscopic observations. They attach variable importance to different forms of usewear: scarring, striations, polish, bevelling and rounding. Moreover, countless descriptive attributes, including such colourful terms as 'melting snowfield' and 'comet tails', have been proposed and analysts tend to develop their own subjective hierarchies of diagnostic attributes. As Van den Dries (1998) argues, an expert system is an important means of formalizing and structuring the interpretation process, because it can organise the rules and hierarchies of a particular approach.

We focus our discussion here on the viability of WAVES. Our perspective is essentially that of expert users, with a commitment to training students. Although each of us has a different research background, we share a common interest in teaching applications and validation of hypotheses.

2. The Leiden approach to usewear

It is clear both from its theoretical underpinnings and key results of the second blind test that WAVES is primarily designed to apply a particular approach, hereafter referred to as the Leiden school, which is somewhat different from approaches developed in France and Australia. The Leiden approach basically started in the early eighties from the premises as originally outlined by Keeley (1980), using similar variables (see Van den Dries 1998, chapter 4). With respect to micropolishes the following aspects were noted: location, distribution, texture, brightness, topography, width and directionality. Striations were described in terms of their location, definition (length, width and depth) and directionality. In contrast to common belief, which associates Keeley with polishes only, Keeley did incorporate edge removals in his functional inferences. In the Leiden approach, however, the low power analysis of the so-called use retouch plays a more prominent role; variables related to edge removals included location, distribution/regularity, width, form and termination (using the Ho Ho committee's suggestions, see Hayden (ed.) 1979: 133-135). This was partially due to the realization that postdepositional changes had frequently affected the use polishes, leaving only edge removals as a clue to use. The Leiden approach also incorporates the morphology of the used edge into the functional inference: edge angle, profile, cross section, shape and outline, as well as the presence of retouch. The interpretation of past function has to be in accordance with the morphological aspects of the edge. As the first assemblages studied largely lacked specific tool types, the method of analysis was very much directed at individual edges, instead of entire tools.

The French approach, as well as the Spanish, is somewhat different, and uses a number of different variables, such as the *trame* and *reticulation* (cf. Gassin 1996; Gonzáles Urquijo & Ibáñez Estévez 1994).[1] In a next version of WAVES some of these variables can hopefully be incorporated, since this would make the system much more useful for the French and Spanish researchers.

In contrast with the focus on the relatively fine grained European flints, Australian based studies have developed a less specific approach to accommodate a wider range of commonly used coarse-grained siliceous stones. Australian stone technologies generally demonstrate a scarcity of distinctive tool types and there is widespread ethnographic evidence of multiple uses for many stone tools. A third feature is the less distinctive nature of micropolishes, arising in part because the silica content in many timbers causes overlapping patterns of polish development for wood and other plants (Fullagar 1991). For these and other historical reasons, Australian studies have incorporated residues with usewear in a broad approach using light and electron microscopes at various magnifications (low, high and very high) (e.g. Allen *et al*. 1997; Kamminga 1982; Fullagar 1986; Fullagar & Field 1997; Furby 1995). This approach depends less on highly distinctive polishes and recognises considerable overlapping of polish patterns from processing wood and other plants. Training in this approach at the Australian Museum and the University of Sydney has proved useful for analysing stone materials like quartzite, quartz, obsidian, silicified tuff, silcrete and other raw materials. WAVES is nevertheless potentially useful in the broad Australia-Pacific region for at least two reasons. First, there are indeed fine grained flint and other chert artefacts found in Australian and Pacific archaeological sites, with some distinctive polishes (Fullagar 1988), making the Leiden approach directly relevant. Second, WAVES does integrate the four main forms of usewear: scarring, striations, rounding and polishes and permits evaluation of these forms, although values for extent of wear development may require adjustment for different lithic materials. Consequently, the Leiden approach may be indirectly relevant because it provides a useful introductory approach to the range of variables commonly found on stone tools in this part of the world.

3. Testing WAVES

Van den Dries has outlined the motivation to construct an expert system for wear trace analysis (1998, chapter 4). Our expectations of an expert system like WAVES are:

1. it should successfully analyse the range of observations to interpret essential elements of tool function (mode of use, relative hardness of worked material (e.g. soft, hard), class of worked material (e.g. wood, flesh, shell, bone);
2. it should accurately record observations and provide a standardised system;
3. it should be simple to use, so that an expert does not have to spend considerably more time studying each artefact than is already required;
4. it should assist in teaching, so that undergraduate students can learn the basic methodology. Students should find it acceptable to learn the basic methodology from a substitute teacher.

Additionally, we were interested in the possibilities of using WAVES for research purposes. Although this was clearly not the primary aim of the expert system, the trajectory of the validation of hypotheses offered a possibility for asking a second opinion to the system concerning an interpretation. Two blind tests were initiated during the construction of the expert system, the first in conjunction with testing the neural network prototype, the second directed at testing the finalized version of WAVES. It is this second test we will discuss a bit further, as it provides a way to evaluate whether the system meets our expectations. The set of tasks selected included contact materials commonly used and with which European analysts ought to be familiar (wood, grass, bone, antler, hide, shell, flesh and clay). Fifteen artefacts were selected from van Gijn's reference collection, nine of which contributed to the knowledge composing WAVES, the remaining six were more recent experiments (Van den Dries 1998, chapter 7). There were four participants, including the second author, a student trained at Leiden, an untrained student from Leiden and a microwear analyst from France. Van Gijn, who contributed to structuring the WAVES rules with her own experiments, set the standard for recording observations (the control set).

In testing WAVES of interest are relationships between personal interpretations and familiarity with the Leiden approach, consistency of observations, completeness of WAVES rules for conceptual knowledge and the ability of WAVES to interpret beyond the experimental data set, particularly in handling non-observed combinations of attributes.

The first part of the analysis, that of the use retouch, displayed considerable variation in the way the analysts described the edge removals. The identification of bifacial versus unifacial retouch seems to pose no problems for the analysts. Within the category of bifacial, interpretations vary with respect to which surface has more edge removals and the presence of alternating retouch. Retouch distribution was not consistently recorded except in a few instances. Categories of 'overlapping', 'uneven' and 'close' were not easily distinguished, nor was retouch orientation consistently recorded. Retouch termination varied for feather, step, hinge and snap. Retouch width, the only quantitative measure, was reasonably consistently recorded, with most variation in the small and very small categories. Edge rounding was also recorded with some consistency with variations at one or the other end of the scale (absent and slight), except for carving dry clay, which covered the full spectrum. Invasiveness created problems for a few experiments with inconsistency over whether retouch exceeded polish or polish exceeded retouch.

The second part of the system, that of polish and striations, was recorded in an even less consistent manner between analysts. Polish location was recorded reasonably consistently, especially with respect to whether or not polish was bifacial or unifacial. Variation occurred in specifying whether polish was mostly on one side or the other. Polish directionality was not recorded consistently, although mostly this was because directionality was not identified (absent) in many interpretations. Polish distribution caused problems, displaying a wide variation; this had critical implications for the resulting interpretations, because it concerns diagnostic attributes for specific contact materials. Polish texture was recorded more consistently, but still with some problems over ‘smooth and matt’ versus ‘rough and greasy’. Polish brightness was not consistently recorded, but variations were at one end of the scale such as dull to bright or bright to very bright. Polish topography caused problems with categories of ‘flat’ versus ‘domed’. Polish width was not consistently recorded, although variation (with a few exceptions) was in the same part of the scale. Presence of striations was recorded reasonably consistently with most variation in the same part of the scale (again with a few exceptions) (see also Van den Dries 1998, appendix IV).

The results of the blind test show that WAVES still has difficulties with the interpretation of certain contact materials, especially mineral substances like clay and with materials like shell. This can be attributed entirely to the lack of knowledge of the system; only a few experiments were carried out with these materials, so WAVES had no rules for slightly different variability. The system seems to deal quite well with silicious plants. The difficulties with bone are remarkable, certainly considering the large number of experiments with this contact material. They are probably due to problems with the description of polish distribution. The incidental resemblance of bone, antler and woodworking traces is a well known problem in wear trace analysis (see Unrath *et al.* 1986), and cannot be attributed to WAVES itself. The mistakes with hide and wood are attributable to the variability wear traces from these contact materials can display. Certainly traces from materials like bark are likely to pose interpretation problems, because WAVES does not have enough knowledge, as traces from this material deviate to some extent in appearance from regular soft wood traces. Insufficient knowledge is also behind the difficulties with the interpretation of the fish working tool. The results show that it is necessary to include many more experiments, in order to have more combinations of attributes to cover the variability of, for example, woodworking traces.

As Van den Dries (1998) argues, there is a strong case to be made that the system works best for those with some familiarity with the Leiden system of recording, rather than experience with other systems of analysis. Indeed it seemed best to have no experience at all and to work under qualified supervision. The reason is of course that the system is entirely dependent on a way of recording highly similar to the rules the system is composed of. Given the subjective manner wear traces are still described across the various laboratories, analysts vary in their descriptions. This shows the difficulties with building an expert system for less standardized knowledge fields. It can to some extent be solved by adding many more photographs, supporting the users in their decisions.

The results also demonstrate a great need for more quantitative measures in wear trace analysis. One advantage of the construction of the expert system was that it clearly exposed this necessity. It may be too labourious to follow Kamminga (1982), who made careful sketches and measurements for each form of edge removals. Perhaps an improvement can already be made by further explicating with drawings and photographs apparently difficult terms referring to polish distribution and polish topography, such as ‘melting snowfield’, ‘comet tails’, ‘pitted’ and so forth. It is exactly these attributes, subsumed under the variables polish distribution and topography, which are usually diagnostic for specific contact materials. If they are described in a different way from WAVES, the system may arrive at a wrong interpretation, even if the personal interpretations (unguided by WAVES) of the analysts are right.[2] The variations in description between analysts clearly is meaningful, because personal interpretations varied from the actual experimental situation.

Lately, in the research context, it has also been experienced that the variation in archaeological traces exceeds the available descriptive terminology based on experiments, so that we really require a typology of archaeological traces, with less emphasis on detailed descriptions of individual attributes based on experimental polishes. The reliance of WAVES on an extensive system of descriptive terms is clearly problematic, at least for experienced analysts, because they have developed their own idiosyncratic descriptive system. This is not so problematic for the teaching application of WAVES, but it is for the hypothesis trajectory. The system would be greatly improved by adding archaeological traces, perhaps only in a photographic way.

WAVES is very easy to use, and user friendly. Especially when run on a fast computer it does not really interfere with the normal analysis procedure, although it is used only incidentally by the more experienced users. The good performance in the second blind test of the totally inexperienced user shows that the system is very suitable in assisting beginning analysts in their learning process. For the most part it is acceptable to the students as a substitute teacher as well (Van den Dries 1998).

4. Conclusion

As a teaching aid WAVES is highly useful. The students greatly appreciate all the information it contains and they also use it with the initial series of analyses. It provides an introduction into usewear analysis, with pictures helping the student to get acquainted with the various relevant and diagnostic attributes. It saves the expert a lot of time in explaining, especially because of the presence of the various information screens. Clearly many more pictures are needed, because for some students it remains difficult to describe the visual phenomena in the same way as the expert. Small variations in the recording of observations can interfere with an otherwise correct line of thought (see for instance experiment 8, analyst IV of the blind test) and thus lead to faulty answers. This is rather disconcerting. Maybe, in the initial stage of getting acquainted with the method there should still be an important role for the expert to make sure the students describe the visual phenomena in the proper way. In Leiden, a small blind test at the beginning of the training is used to ascertain whether the student is describing the traces in the correct way. It should be remarked however, that an expert system cannot solve the weaknesses inherent in the analysis of wear traces. Although it can standardize the line of reasoning, it remains dependant on a subjective description of the wear traces. The difficulties with the description of polish distribution are illustrative. It may be useful to separate this variable in two components, like degree of linkage and degree of intrusion as suggested by Juel Jensen (1994: 24).

As a research aid WAVES is much less useful. Clearly, this was not the objective of the expert system, but the hypothesis testing option could possibly assist in the interpretation of problematic traces, because of the very fact that the system is consistent in its reasoning process. However, contrary to the evaluations of the expert, who includes the latest findings in his or her evaluation, the system does not have the necessary knowledge to evaluate the more problematic traces. The time lapse between the acquisition of new knowledge by the expert and the inclusion of this knowledge into an expert system can amount to several years and is one of the most important drawbacks of using expert systems. WAVES, for example, is largely based on experiments performed and described between 1984 and 1988, supplemented and modified by expert knowledge up to 1994.

By definition expert systems are impossible to have creative thoughts without the expert having them. In this sense, it may be the neural network (see Van den Dries 1998, chapter 6) that could be more useful for research purposes, because it always provides an educated guess, even if it does not have the necessary knowledge.

Van den Dries (1998) argues that expert systems can play an important role to alleviate the high work load of experts, arisen from the implementation of the Malta convention. She is probably very right when it concerns identification of well defined bone or plant remains, but wear trace analysis provides an interpretation, not an identification of tool function (see Van Gijn 1990). It concerns a creative process of evaluating the variability of traces, resulting from an endless number of possible prehistoric activities in order to arrive at a meaningful interpretation of tool use. What WAVES can do is formalizing our chain of reasoning, but it can never deal in a routine way with large amounts of data, if only because the variability in archaeological traces turns out to exceed the experimental one. It is our contention therefore that WAVES cannot play a role in the commercializing of wear trace analysis. However, WAVES is highly useful as a teaching aid, because of the very fact that it formalizes the way students arrive at an interpretation, without the constant need of supervision in the learning stage. This is also the case for wear trace analysis on stones other than flint (see above). In this sense it frees the expert to do innovative work. The expert will therefore always remain at the cutting edge of the method, and can never be replaced. The system depends on his or her knowledge growth to be updated according to the latest ideas and findings and can never be as creative as the most uncreative and unimaginative analyst of wear traces. For those experts being suspicious of having their knowledge incorporated in a computer system, this should be a consoling thought.

references

Allen J.A., S. Holdaway & R. Fullagar 1997. Identifying specialisation, production and exchange in the archaeological record: the case of shell bead manufacture on Motupore Island, PNG, *Archaeology in Oceania* 32, 13-38.

Fullagar, R. 1986. *Usewear and residues on stone tools. Functional analysis and its application to two southeastern Australian archaeological assemblages*, PhD thesis La Trobe University, Melbourne.

Fullagar, R. 1988. Recent developments in Australian use-wear and residue studies, in: S. Beyries (ed.), *Industries lithiques: Tracéologie et technologie* (=BAR International Series 411), 133-145. Oxford.

Fullagar, R. 1991. The role of silica in polish formation, *Journal of Archaeological Science* 18, 1-25.

Fullagar, R. & J. Field 1997. Pleistocene seed grinding implements from the Australian arid zone, *Antiquity* 71, 300-307.

Furby, J. 1995. *Megafauna under the microscope: Archaeology and palaeoenvironments at Cuddie Springs*, PhD thesis, University of New South Wales, Sydney.

Gassin, B. 1996. *Évolution socio-économique dans le Chasséen de la grotte de l'Église supérieure (Var)*, CRA 17, Paris.

González Urquijo, J.E. J.J. Ibáñez Estévez 1994. *Metodología de análisis funcional de instrumentos tallados en sílex,* Bilbao.

Grace, R. 1996. Use-wear analysis: the state of the art, *Archaeometry* 38, 209-229.

Hayden, B. (ed.) 1979. *Lithic use-wear analysis*, New York.

Juel Jensen, H. 1994. *Flint tools and plant working. Hidden traces of stone age technology*, Aarhus.

Kamminga, J. 1982. *Over the edge: Functional analysis of Australian stone tools* (=Occasional Papers in Anthropology 12), Brisbane.

Keeley, L.H. 1980. *Experimental determination of stone tool uses. A microwear analysis*, Chicago.

Newcomer, M.H., R. Grace & R. Unger-Hamilton 1986. Investigating microwear polishes with blind tests, *Journal of Archaeological Science* 13, 203-218.

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

Van den Dries, M.H. 1998. *Archaeology and the application of artificial intelligence*; case studies on use-wear analysis of prehistoric flint tools, this volume.

Van Gijn, A.L. 1990. *The wear and tear of flint. Principles of functional analysis applied to Dutch neolithic assemblages* (=Analecta Praehistorica Leidensia 22). Leiden.

Annelou van Gijn
Faculty of Archaeology, University of Leiden

Richard Fullagar
Australian Museum, Division of Anthropology, Sydney

notes

1 In a research project financed by the CNRS, France (coordinator B. Gassin), concerning the use of flint tools for the harvesting and processing of plants, the approaches were integrated; this integration could unfortunately not be incorporated into WAVES because it took place after WAVES had been finished.

2 It should be stressed that WAVES does not provide one interpretation but a number thereof, each with a diagnostic value. It is not obligatory to choose the interpretation with the highest diagnostic value; it is possible to evaluate the photographs of the traces of the different contact materials interpretated and make a choice on that.

colophon:

Graphic design:	H.A. de Lorm
Copy editor:	M. Wanders
Photography:	J. Pauptit
Computer-drawings:	P. Heavens and M.H. van den Dries
Handmade drawings:	H.A. de Lorm
Printing:	Orientaliste, Herent, Belgium