Article

The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts

G. Rugg[1] and P. McGeorge[2]

(1) School of Computing Science, Middlesex University, Bounds Green Road, London N11 2NQ
G.Rugg@city.ac.uk

(2) Dept of Psychology, University of Aberdeen, King's College, Aberdeen AB9 2UB, UK
mcgeorge@abdn.ac.uk

Abstract: *Although sorting techniques (e.g. card sorts) are widely used in knowledge acquisition and requirements acquisition, they have received little formal attention compared to related techniques such as repertory grids and laddering. This paper briefly describes the main sorting techniques, then provides a detailed tutorial on one variety (repeated single-criterion sorts), using a worked example. Guidelines for choice and sequencing of techniques are given, both in relation to varieties of sorting technique and in relation to other techniques. It is concluded that the sorting techniques are a valuable part of the elicitor's methodological toolkit.*

Keywords: sorting, card sorts, repertory grids, laddering, Personal Construct Theory

1. Overview

One of the factors leading to the rise of Knowledge Acquisition (KA) as a research field was recognition of the problems caused by the "knowledge acquisition bottleneck" (e.g. Barr and Feigenbaum, 1982). Although the extent of this problem has been re-considered (Cullen and Bryman, 1988), the importance of KA is still undisputed, and many techniques have been imported into KA from other disciplines. A similar trend is apparent in the closely related field of Requirements Acquisition (RA) which has paid considerable attention to the KA literature, and has also imported techniques from other disciplines.

Although it is clearly essential to choose the correct technique for a task, and to use it correctly, there is surprisingly little guidance on this in the literature. There are some partial guides to choice of KA technique (e.g. McGeorge and Rugg, 1992), but the only published integrated framework for selection and sequencing of techniques appears to be in the RA field (Maiden and Rugg, 1996). There is also a scarcity of readily available tutorial papers on the individual techniques; most of the literature on individual techniques assumes previous knowledge of the technique being discussed.

This paper presents a tutorial on the use of the sorting techniques, intended for readers who have no previous knowledge of those techniques. Its primary goal is to give readers sufficient "hands-on" knowledge to allow them to use the techniques correctly; for this reason, quite detailed practical issues are discussed where necessary. The secondary goal of this paper is to prepare readers for the more sophisticated literature on technique use, so there is reference where appropriate to underlying semantic issues, etc.

This paper is written for both an RA and a KA audience. For brevity, a single case study is used, which could apply both to KA and to RA. The case study makes allowance for the comparison of results across multiple respondents. This may not be an issue in many KA situations, where only one or two experts may be available, and then for only short periods of time. In at least some RA cases, however, there may be a reasonably large sample of potential clients whose requirements need to be investigated and integrated, and for such situations comparing respondents' categorisation is an important issue, even if it does lead to greater complexities and problems than when dealing with a single respondent.

This paper concentrates on one variety of the sorting techniques. The reasons for this are partly practical (lack of space), partly to do with availability (the variety described here has been formally compared with other techniques in quantitative experiments) and partly theoretical (there are theoretical problems associated with the other varieties). All these issues are discussed in more detail below.

1. 1. *Introduction*

An important part of people's knowledge is the categories which they use. Experts are expert largely because they have a more extensive and sophisticated categorisation than non-experts (Chi, Glaser and Farr, 1988; Ellis, 1989). It is therefore important to be able to elicit respondents' categorisation.

One response to this has been the extensive use of repertory grid technique (Kelly, 1955) in KA (e.g. Shaw, 1980; Shaw and Gaines, 1988). Repertory grid technique is based on an object:attribute matrix (the "grid"). The cell values usually contain numbers for Likert-type semantic scales; for instance, the attribute line for "hardness" in a repertory grid on minerals would describe each of the minerals in the grid in terms of its hardness, probably on the 1–10 scale of hardness normally used by mineralogists. A major attraction of this approach is that it is highly suitable for statistical analysis, allowing sophisticated interpretation and manipulation of the results obtained, and numerous computerised versions have been produced (e.g. Boose, Shema and Bradshaw, 1989).

Although the repertory grid approach has many advantages, it is not well suited for all sorts of data. Repertory grid technique encounters particular problems when dealing with nominal values, i.e. data which do not form any sort of semantic scale, and which are divided into non-scalar categories (Yorke, 1983, Rugg and Shadbolt, 1991). Such categories are, however, well handled by the sorting techniques, described in this paper (McGeorge and Rugg, 1992).

The basic idea behind the sorting techniques is simply to ask respondents to sort things into groups. The things may be *objects*, such as different types of mouse, or *pictures*, such as screen dumps of various screen layouts, or may be *cards*, with the names of objects or situations on the cards, such as the names of different editors. The groups may be ones chosen by the questioner, or ones chosen by the respondent, or a mixture of both. The sorting techniques are a useful way of eliciting respondents' groups, and of finding out how much agreement and disagreement there is between respondents about which the categories.

The sorting techniques are therefore useful both for identifying relevant categorisation and for investigating commonality and differences between experts in the use of that categorisation. These techniques have the further advantages of being quick, systematic, and easy to use, both for respondent and questioner (Rugg, Corbridge, Major, Burton and Shadbolt, 1992). They have been quite widely used in knowledge acquisition (e.g. Gammack, 1987), but have received comparatively little systematic attention compared to techniques such as repertory grids. This paper examines reasons for this omission, and describes one variety of sorting technique in detail.

1. 2. *Background*

The sorting techniques are aligned with the constructivist approach, and more particularly to Kelly's Personal Construct Theory (PCT) (Kelly, 1955): they assume that people make sense of the world by categorising it, and that people can describe their own categorisation of the world with reasonable validity and reliability. It is clear that people are not always able to do this (Seger 1994; Kahneman, Slovic and Tversky, 1982), so before using a sorting technique it is usually advisable to use at least one observation-based technique to check whether these assumptions hold true for the area about to be investigated.

One attraction of the techniques associated with PCT techniques is that their use of a common underlying conceptual framework makes it easy to integrate them in a structured, systematic way (e.g. McGeorge and Rugg, 1992). This issue is of particular importance in view of the long-standing and continuing interest in integrating software tools to support KA (e.g. Shadbolt and Wielinga, 1990).

Although sorts have received comparatively little attention in the Personal Construct Theory (PCT) literature, they have been part of that literature from its outset, and both fit into its framework and complement other PCT techniques well. For that reason, we have used PCT terminology throughout this paper. For clarity, the relevant terminology is defined here.

A **construct** is an attribute used by an individual to describe something (Kelly, 1955, Bannister and Fransella, 1980). For example, the construct may be "easy to use" or "expensive". "Construct" can be used in a wider range of contexts than "criterion" or "category", which only make proper use when used to describe sorts.

A **criterion** is the attribute used as the basis for a sort when using the sorting techniques. For example, the criterion may be "place of manufacture" or "cost". The criterion provides the basis for sorting things into categories (see below).

A **category** is a group into which things may be classified, using a criterion. For example, the categories under the criterion of "place of manufacture" may be "USA,", "Japan", "Europe", etc.

A **facet** is the viewpoint used for a particular set of classifications. For example, computers may be categorised in terms of criteria relating to hardware features, or in terms of criteria relating to usability (Vickery, 1960; Rugg and McGeorge, 1995).

Range of convenience is the range of settings in which a construct can be used meaningfully (Kelly, 1955, Bannister and Fransella, 1980).

1. 3. *Varieties of sorting techniques*

There are several varieties of sorting techniques. Unfortunately, although these have been quite widely used, they do not appear to have been systematically surveyed. The descriptions and categorisations below are our own.

Although several varieties of sorting technique are described in this paper, not all of these are suitable for inexperienced users. In some cases, this is because the variety described is one which involves complex theoretical presuppositions, which need to be understood before the technique can be used properly; in other cases, the variety described depends on sophisticated statistical analysis, which may be misinterpreted by anyone unfamiliar with statistics. The main varieties are therefore described for the sake of completeness, but the main emphasis of this paper is on the variety which is in the authors' opinion most suitable for the majority of users, namely repeated single-criterion sorts.

Q sorts Q sorts derive from Stephenson's (1953) Q methodology, and have been extensively used within fields such as personality theory in psychology. Q sorts normally involve use of quite a large set of cards, each of which bears a different statement or phrase. Respondents are asked to fit the cards into a normal distribution pre-defined by the investigator. For instance, the cards may each bear a description of a personality attribute, and the respondent's task may be to fit each card onto a scale ranging from "strongly agree" to "strongly disagree", with only a few cards being used at each end of the scale, and most cards being placed somewhere near the middle. Statistical analysis is then used across results from different respondents, or results from different sessions with the same respondent, to infer higher-order clusterings, etc.

This approach imposes a semantic distribution on the respondent's categorisation (i.e. a normal distribution) which may not be appropriate, and which could lead to distortion of the results. In addition, the technique depends on statistical analysis, and requires significant preliminary work to establish the appropriate contents for the cards, both of which require significant extra effort on the investigator's part. Q sorts also take a significant amount of time for a single sort. Because of these disadvantages, we do not recommend Q sorts for routine KA use.

Hierarchical sorts Some practitioners and researchers use sorts to establish semantic hierarchies within a domain. This may be accomplished either by using cards representing entities at different semantic levels, or by using cards from the same semantic level.

An example of cards representing entities at different semantic levels would be a pack in the domain of zoology where some cards represented species, such as otters or pine martens, and others represented classes, orders, phyla, etc, such as "carnivora" and "mammalia". The respondent's task would be to organise these cards into the same type of hierarchical organisation as in the familiar Linnean taxonomy.

The main drawback with this approach is that choosing the appropriate range of entities for the cards would be a major KA exercise in its own right, while use of an inappropriate range of entities would lead to distortions in the knowledge elicited. We therefore do not recommend this approach for normal KA use.

An example of cards representing entities at the same semantic level, again in the domain of zoology, would involve all the cards representing species, such as otters, pine martens, etc. The cards would then be sorted into phylum, class, order, family, etc. This might be accomplished by sorting the pack once for each semantic level, and then re-sorting the same cards for each successive level: one sort for the phylum level; another for the class level, and so forth.

This approach is preferable to using cards representing different semantic levels. However, it still does not guarantee complete domain coverage for the hierarchies involved: entire branches of the hierarchy may be missed if the cards do not cover all the relevant categories. A more sensible approach to hierarchy elicitation would be the use of laddering (Hinkle, in Bannister and Fransella, 1980; Corbridge, Rugg, Major, Shadbolt and Burton, 1994; Rugg and McGeorge, 1995).

"All in one" sorts This category covers a range of overlapping approaches, which share the common attribute that the respondent only performs one sort on the entities being sorted. For instance, respondents may be asked to sort the entities into a matrix layout using one attribute for one axis of the matrix and a second attribute for the second axis. A hypothetical example of this would be to ask a medical expert to categorise cards bearing the names of illnesses into a matrix with seriousness of the illness along one axis, and rarity of the illness along the other.

Another approach would be simply to ask respondents to sort the entities into clusters in terms of overall similarity (i.e. similarity across a range of attributes simultaneously, rather than conducting separate sorting sessions for each individual attribute). A hypothetical example of this would be to ask a vehicle expert to sort a set of cards bearing the names of vehicles into clusters. It is likely that the expert would categorise the four-wheel drive vehicles as similar to each other for the majority of salient attributes, although they might differ widely across some individual attributes (e.g. place of manufacture).

The main disadvantage with both these approaches is that they do not attempt to elicit individual attributes systematically, but instead look for underlying factors, usually via statistical analysis. This may be useful when searching for nomothetic regularities in the social sciences, but in both KA and RA it is important to be able to elicit the knowledge of individuals (whether experts, users or clients) as accurately, validly and reliably as possible. We would not therefore recommend the "all in one" techniques for routine KA use.

Repeated single-criterion sorts In repeated single-criterion sorts, the respondents sort the same entities repeatedly, categorising in terms of a different single attribute ("criterion") each time.

Repeated single-criterion sorts are more flexible and easier for most elicitors to handle; for these and the theoretical reasons described above we favour repeated single-criterion sorts, and the following descriptions will refer to repeated single-criterion sorts unless otherwise specified.

As with other varieties of sorting, the entities involved in sorting may be objects (object sorts), pictures of entities (picture sorts) or the names of the entities (card sorts).

Somewhat surprisingly, empirical research to date has found no significant semantic differences in the types of criteria and categories elicited by card, picture and item sorts for repeated single-criterion sorts (Rugg et al., 1992) - for instance, item sorts were not significantly more likely to elicit "concrete" criteria and categories than card sorts. The guidelines for choice between these three varieties are therefore based primarily on logistical grounds.

Sorting may be used to elicit criteria and categories from a respondent, or may be conducted using supplied categories (the latter is useful for assessing agreement and disagreement between respondents).

1. 4. When to use sorts and when not to use sorts

When to use sorts Sorts are appropriate when the emphasis is on finding out the categories which people use - for instance, finding out which symptoms of a problem are considered by an expert to be significant.

Sorts may be used both as an exploratory technique and as a main technique. Sorts should only be used as a main technique after proper groundwork to establish that there are no significant differences between front and back versions of the topic (Goffman, 1955). They may, however, be used as an exploratory technique as part of the piloting work when deciding on the main technique, if it is clearly understood that the results from such use may be distorted.

It is possible to use sorts on a surprising range of entities, including abstract or complex entities, such as "programming in C", or " being lost in a deeply nested set of menus". This use of abstract entities extends to using sorts reflexively to categorise criteria themselves, which raises the intriguing possibility that this sort of meta-categorisation may provide insights into the higher-level construing used by the respondent, perhaps including faceting. If so, this would provide a further link between sorting and laddering.

When not to use sorts Sorts only address static, flat, explicit knowledge. They cannot conveniently access knowledge about sequencing procedures, about trade-offs, about knowledge structures such as hierarchies, or much tacit knowledge.

Sorts only make proper sense when comparing like with like: comparing "a hard disk" with "running late on a software development project" is meaningless. The entities need to be at the same semantic level as each other, and evenly semantically spread. Otherwise there is a real risk of the sort of semantic problems described in a chapter entitled "Is a boulder sweet or sour?" (Osgood, Suci and Tannenbaum, 1957), where constructs were being pushed far beyond their range of convenience.

1. 5. Choice of appropriate variety

Object sorts involve the most concrete, specific entities.

Advantages:
Respondents can use all of their senses to investigate the objects when deciding on the categories they will use.

It is possible to use object sorts in a domain new to the elicitor or new to the respondent, as in the case of user assessment of a new product.

Disadvantages:
The objects may have all sorts of irrelevant and distracting features unconnected with the purpose of the questioning session. It is surprising how much time experts will spend looking at such features.

There are also practical considerations: some entities are too big to be sorted in object sorts.

Picture sorts are more restricted, in that respondents can use only visual information, but are often more practical than object sorts.

Advantages:
Picture sorts share the advantages of object sorts (apart from the disadvantage of reduced sensory input). They have obvious practical advantages over object sorts, while providing more information than card sorts to the user. It is possible to trim out unwanted extraneous detail from pictures, which might not be possible with objects.

More subtly, it is possible to use picture sorts with slightly different pictures of the same object for different respondents, to see what effect is caused the changed feature in the pictures.

Disadvantages:
Picture sorts cannot show all the information that is available in an object: weight, for instance, is something which gives users an impression of solidity and sturdiness in an object, but can only be guessed at in a picture.

Card sorts are the simplest form of sorts, in that the entities being sorted are simply names on cards.

Advantages:
The main advantage of card sorts is the inverse of their main disadvantage. Since the cards contain so little information, there are no problems with extraneous detail (which may cause difficulties in object sorts or picture sorts). In addition, it is possible to investigate the respondent's recall knowledge about the entities in the domain, as opposed to the respondent's recognition-based categorisations of the domain.

Card sorts and picture sorts are both amenable to computerisation (Major, 1991).

Disadvantages:
Card sorts require the respondent to know about the entities named on the cards; it is not possible to use entities unknown to the respondent.

2. Using sorts

2. 1. Choice of entities

The most convenient way of ensuring appropriate semantic coverage is to use laddering as a preliminary technique, laddering on the appropriate facet and finding entities at the same level. The entities for the card sort can then either consist of one entity from each set of nodes sharing the same parent, or of all of the children of the same parent.

Alternatively, it is possible to use pre-existing classification schemes, or to ask the respondent to draw a hierarchy, and then use that. The following diagram illustrates this.

The basic principle is simple: the entities should be from the same horizontal level in the hierarchy, and should be either evenly spread across the whole thing, or be a complete set of one of the sub-trees.

The higher up the hierarchy the choice is made, the more abstract and general the categories which are likely to be used by the respondents. The lower down the hierarchy, the more specific the categories are likely to be, with a correspondingly narrow range of convenience.

If the entities all come from a very low level, then it is likely that all of them will share various attributes inherited from higher levels in the hierarchy, and that these attributes will therefore not emerge during the sorting (which can only identify categories to which some entities belong and others don't). For instance, a sort at the level of "computers", "printers", "CD-ROMs" will probably produce categories such as "has a keyboard" or "uses paper", whereas a sort at the level of "Mac Centris 610", "Mac Classic" and "Toshiba T3400CT/250" would produce much more specialised constructs. This is something to beware of with object sorts, which are by definition at the most specific level possible. It may be that all of the objects shown to the respondents share some feature which the respondents consider highly important, but which is never elicited in the sessions for the reason above, precisely because it is common to all the objects and therefore cannot be used as a criterion for sorting them into different groups. In an ideal world, this risk can be reduced by doing a series of sorts at progressively more detailed levels; if this is not possible, then it is a case of keeping an eye open for significant absence of features which were mentioned when other techniques were used.

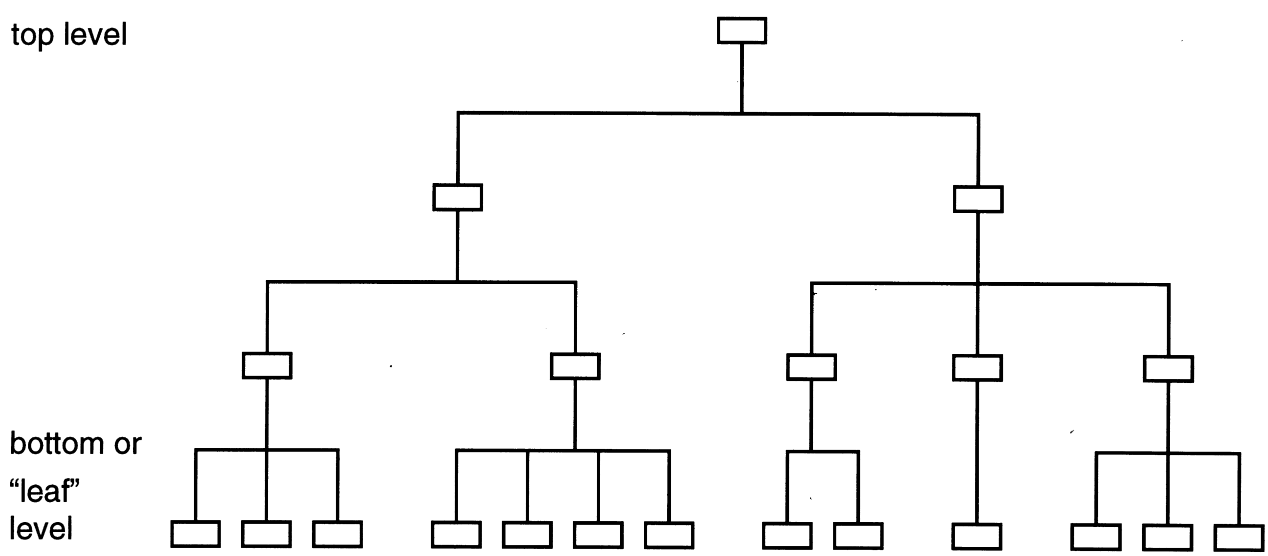

Figure 1: *A sample hierarchy*

2. 2. Number of entities

As a rough rule of thumb, if the results are to be analysed statistically, then the lowest number of entities should be about eight; otherwise, the lowest number is a matter for the questioner to decide.

The maximum number of entities which is conveniently manageable for repeated single-criterion sorts is about twenty or thirty, though it is possible to use significantly more in some circumstances. If the session is going to consist of an "all in one" sort, then a much higher number becomes feasible.

2. 3. Procedure

Once the entities have been chosen, it is necessary to prepare the items to be used in the sessions, and to prepare the instructions for the respondents.

Preparing the items As usual with practicalia, attention to detail can mean the difference between an uneventful session and a disaster.

If objects are being used, then they need to be clean, safe to handle, and solid enough not to fall apart. If the topic area being investigated is unfamiliar to the elicitor, and involves borrowing items, then it is necessary to pay particular attention to the welfare of the items. In such cases it is necessary to explain in fine detail to the lenders exactly what will be involved. Sooner or later something is likely to fall off the table and shatter, or to be taken apart by a curious respondent. A less obvious problem is that it may not be acceptable to use a sticky label to identify the item because of chemical contamination to the item, or because it would gum up the works.

Pictures should be the same size, and as similar as possible with regard to glossiness and other extraneous but possibly distracting features. It is advisable to attach them to a sturdy backing or to cover them in plastic if they are to be extensively used, to keep wear and tear to a minimum.

Cards should likewise all be the same size. We usually use standard small filing cards, with the words word processed onto paper and then stuck onto the cards. This reduces problems with illegible handwriting, and avoids the issue of trying to get filing cards through a borrowed typewriter. Filing cards are cheap and easily available, and it is easy to add more cards to the pack during the session, or even to create a new pack during the session, if need arises, though this does involve using hand-written cards. For our work, this is unlikely to make much difference, but for some applications it may be necessary to standardise.

All the items should be numbered for recording the results. The numbers should be clear and unambiguous (e.g. "6" and "9" may cause confusion with object sorts, if it is not clear which way up the object and number go).

It is also advisable, if more than one different set of items is being used, to label the different sets with e.g. coloured stickers, so that items from one set do not find their way into another.

The instructions The instructions need to make it clear what the respondents are expected to do, using a "toy" example to demonstrate this. The toy example should be from a completely different domain, to reduce the risk of cueing, and should usually be familiar to the respondent. Vehicles and animals are usually suitable domains for this.

It is highly advisable always to use the same set of instructions within each set of sessions. Otherwise, there is the risk that the last few respondents will be behaving differently from the first ones, leaving you wondering whether this was due to the instructions or not - for instance, if it turns out that the last respondents used a "don't know" category extensively but the first respondents didn't use this at all.

The instructions should make it clear what the respondents can do, and what you would prefer them not to do. For instance, it is usually advisable to tell respondents that they can use the categories of "other", "not sure", and "not applicable": this identifies areas where a category is being pushed beyond its range of convenience, areas where respondents' knowledge ends, and various other very useful things. It is also usually advisable, for instance, to tell respondents not to lump two sorting criteria together in one sort: e.g. "big and expensive" should be sorted once for "big" and once for "expensive".

There is a sample instruction sheet at the end of this paper: readers are welcome to make use of it themselves, as long as they give appropriate recognition to the source.

2. 4. Conducting the session

Once the respondent has been given the instructions and understands clearly what is involved in the session, the respondent is given the set of items and asked to sort them into groups, with one group for each category, using only one criterion for the sorting. The first sort is the one most likely to cause problems: respondents often do not realise that they will have the chance to use as many different criteria as they want on successive sorts, and proceed as if

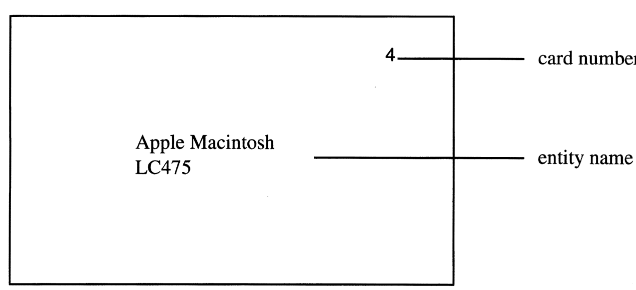

Figure 2: *A sample card*

they were having to cram everything into one sort. Normally, though, sessions proceed smoothly, and respondents usually grasp the basic concept quickly.

We usually encourage the respondents to look at all the items at the start of the session before they do any sorting, so that they are fully aware of the range of items to be sorted.

For practical reasons, it is usually advisable to have a large desk or table on which to spread the items. The desktop or tabletop should be kept completely clear, apart from the items: otherwise, cards and pictures have a habit of vanishing underneath desktop clutter, and in any case, the process requires a reasonable amount of space.

What should then happen is that the respondent should sort the items into separate groups, using a single criterion. For example, if the items are all types of computers, then the criterion may be "power source", and the categories may be "battery powered", "mains powered" and "either mains powered or battery powered". It is perfectly permissible to have "either...or" categories of this sort, as long as they are clear ones. In this example, the respondent could be asked to clarify what the last category involved: if the respondent means that the computers in that category can be used in either of two modes, then the category is acceptable, but if the respondent is simply not sure which type of power applies, then the category should be labelled "not sure" or equivalent instead.

Although in theory it is more efficient to ask respondents to say in advance what criterion they are about to use before each sort, in practice this is inadvisable: respondents often change their mind during the sort, or even during the recording after each sort. This apparently minor point has major implications for anyone thinking of writing their own card sort software, and is worth emphasising.

The first thing to be done after the respondent has stopped sorting is to find out what the criterion was for that sort (e.g. "power type" in the example above). If it is clear that this is a single concept, and that the respondent is not lumping two or more criteria together, then all is well so far; otherwise, it is necessary to explain what is needed, with another example if necessary, and ask the respondent to start again.

Once the criterion is known, the next stage is to go through each of the groups in turn, asking what the category is which corresponds to each group (e.g. one group in the example above would correspond to "battery powered", another to "mains powered", etc). The main source of problems with this stage is the "leftovers", i.e. items which are not included in any of the groups. This may be because the respondent has not realised that it is acceptable to have "don't know" or "not applicable" categories; it may be because the respondent has simply forgotten them. In such cases is simply a question of clarifying what is allowable, and asking the respondent to put the leftovers into the appropriate pile, or to give them the appropriate group name.

Sometimes, however, the leftovers represent cases where the categorisation breaks down, and complex explanations are necessary. This can feel like a real nuisance at the time, but it is actually a valuable source of information, and can make a huge difference to understanding the topic. It is advisable to clarify such cases immediately, in case the rest of the sorting session would be a waste of time. If the issue is clearly a complex one, then it may be advisable either to stop sorting and use the time to clarify the issue with whatever technique seems appropriate, or to book another session on another day, to give you time to think and plan.

Once the criterion and the categories have been established, it is time to record the items in each group. The reason for asking these questions after the sort rather than before it is that people very often change their minds during a sort, or make mistakes; it is much simpler for everyone involved to ascertain the criterion and categories afterwards.

2. 5. Recording the session

The main recording for a sorting session is paper-based. However, we also advise use of a tape recorder (for respondents' comments if problems occur). It is also worth considering using a Polaroid-type camera (for a quick backup record of groupings). If using a camera, it is advisable to check beforehand that the photographs can catch enough detail to allow all the entities to be easily identified.

Recording the items in the group can easily go wrong, for various reasons: people surprisingly often change their minds during the recording, for instance, and move items from one group to another. The method we use is to write:

1: respondent number, date, facet used, and any session codes, then for each sort:

- the sort number, and the criterion for that sort;
- the group/category names;
- the code numbers of the items in each group together, with a mark to distinguish the end of one group from the start of another, and a comma between each number (so that "1,2" cannot be confused with "12", for instance).

(In principle, it should be at least as simple to list each category on a new line, with the corresponding card numbers beside it; the main disadvantage of this method is the risk of confusing the criterion with the first category.)

We then count the numbers to make sure that all the items are accounted for. If time permits, it is also advisable to account for each number individually (in case there are two "6"s and no "9"s, for instance). It is surprisingly easy to make recording mistakes, and the extra attention can save a lot of woe. We use squared paper for increased tidiness and correspondingly reduced risk of error, but this is

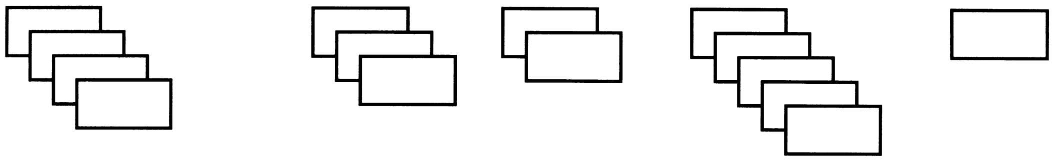

Figure 3: *A sample set of cards after sorting*

not essential. A recorded sort would therefore look something like Figure 4:

Sort number 1:
Criterion: power type

mains powered	battery powered	battery or mains
1,3,5	2,8	4,6,7

Figure 4: *Record of a sort*

Using code numbers rather than names saves a lot of recording time, and can reduce the risk of cueing respondents towards a particular type of response. With object sorts the objects often only have code numbers rather than meaningful names anyway.

One point which needs to be stressed is that the only questions which the questioner should ask are ones involving clarification. It is highly inadvisable to comment on the respondent's categorisation by e.g. telling them that they are wrong, or asking them if they seriously mean that a particular set of items can be grouped together. The point of the session is to find out what the respondents' categorisation is, not the questioner's.

Once the sort has been performed and recorded, the items are returned to the respondent for the next sort, using a different criterion. After the first sort, this usually proceeds smoothly.

Respondents usually start to run out of ideas for criteria after a while (somewhere between five and ten sorts, or up to twenty or more if dealing with experts). It is worth recording the point at which this happens, since it may reflect a change from explicit knowledge to semi-tacit or tacit knowledge of some sort.

If the questioner wants to elicit as many categories as possible, then one way of doing this is to choose two items at random, and ask the respondent to say what the differences are between them (dyadic elicitation), or to choose three items at random and ask what two have in common which the third does not (triadic elicitation). If any of the differences are criteria which have not yet been used, then these can be used as the basis for the next sort.

Eventually the respondent will run out of criteria and categories. Respondents are quite often apologetic about this, and it is both courteous and advisable to reassure them: after all, nobody could be expected to keep going forever. Another thing worth pointing out is that you want to find out what categories people actually use, not to find more categories than anyone else.

At the end of the session, it is advisable to check that all the paperwork is clearly identified and labelled.

Supplying criteria and categories It is often useful to supply the criteria and the categories to the respondents, or to supply some and elicit others. If supplying some and eliciting others, then it is advisable to think about whether to supply before eliciting or vice versa. There are arguments in favour of both. Supplying first has the advantage of tidiness: it is easier to compare results across respondents, since "sort 1" for one respondent will correspond with "sort 1" for all the others. This may sound like a trivial point, but it is horribly easy to get numbers mixed up, and produce transcription errors. On the other hand, eliciting first makes it easier to see which categories the respondents would have used on their own initiative, and would therefore be more realistic. Whatever the decision, it is highly advisable to include in each supplied sort the categories of "other" and "not applicable", in case the supplied categories do not correspond to the respondent's own categories. If anything turns up in the "other" category then it should be investigated during the session.

The actual procedure for supplying criteria and categories is much as one would expect. It is a case of telling the respondent what will happen, then for each sort telling them which criterion and which categories to use. It might be necessary to write labels for each category to remind respondents which categories to use, if there is a risk of respondents forgetting.

It is advisable to label clearly in each session the point at which supplied criteria stopped and elicited ones began. (This is partly because it is easy to forget or lose the list of supplied criteria, and partly because the session may sometimes elicit the first supplied criterion at the end of the elicited ones, which would otherwise lead to confusion.)

2. 6. *Common problems*

Subjects often lump two or more criteria and/or categories together, especially in the first few sorts. This is usually easy to cure by politely checking whether this is indeed what has happened, and then asking the subject to sort on one criterion this time, and the other on the next sort.

It seems reasonable to assume that the number of items will be the same for all respondents who were in the same session, but this is sometimes not the case. Very occasionally a respondent will say that one of the items is actually a class in its own right, and needs to be subdivided to make sense — for instance, that there were two utterly unrelated computers called “The Dragon”, or two versions of the same computer, one of which was a disaster and the other of which was a classic. In either of these cases, the response in the session needs to be the same: write new cards to reflect this information, take it at face value at the time, run the session, and do more homework afterwards before doing any further sessions. If the respondent is right, and there is the risk of serious confusion between two utterly different entities, then it is time to start again, or to throw away any information affected by this mistake. If, on the other hand, the respondent is simply being too pedantic, then the simplest solution is to humour the respondent at the time, and decide afterwards whether to throw away their results, or to use them with appropriate caution (they might be informative).

3: Analysis and further steps

As usual, the type of analysis is dependent on the purpose of the sessions. Proceeding from the simplest to the most complex, the main forms of analysis are as follows.

3. 1. Counting the criteria

The number of criteria used is informative.

If all the respondents only use two or three criteria, then there is unlikely to be much categorisation knowledge there to elicit from the respondents. This might be because they do not happen to know much, or may be because they simply don’t care much about the area; it might also be because the knowledge involved is not categorisation knowledge. If the number of criteria is consistently low when there is reason to expect otherwise, then it is worth checking the procedures used in the sorting, in case the procedures have mis-directed the respondents into answering the wrong questions.

If all the respondents use large numbers of criteria, then there is considerable knowledge involved, for the respondents’ population at least. If there is a range in the number of criteria used, then this means that there is a corresponding range of knowledge within the respondents’ population.

3. 2. Type of criteria

The types of criteria used are informative, and often surprising.

The criteria used may be observable ones (size, colour, etc.) or unobservable ones (cost, prestige, etc.). They may be “objective”, such as size and colour, or subjective, such as “ones I like a lot”. They may be intrinsic, i.e. an integral part of the entity itself, such as “made of plastic”; they may be extrinsic, i.e. attributes attached to the entity but not part of it, such as the fact that a particular computer was made in Japan.

Where criteria are unobservable, subjective, or extrinsic, it is possible to investigate whether there are any ways of inferring information about the criterion from other criteria. Laddering is a particularly suitable method for this. For instance, the attribute “expensive” may be deduced from the manufacturer’s name, or from the use of materials known to be expensive, or from the type of design, or from the presence of extra product features known to be expensive. A manufacturer engaged in product research would want to know about such cues, especially ones such as design style which could imply expensiveness in a product without necessarily involving extra manufacturing cost.

3. 3. Commonality of criteria

The next thing to look at is the amount of overlap between respondents in the criteria used (“commonality”). Sometimes most of the respondents will use much the same criteria as each other. Other times, although the respondents use about the same number of criteria, the criteria used will be different from respondent to respondent. Generally speaking, high commonality means a simpler life for the questioner, and low commonality means more work.

If, for example, the sorting is being used to identify issues to include in a questionnaire, high commonality in the sorts results means that the criteria with high commonality can be included in the questionnaire with reasonable confidence in their validity and usefulness. If, on the other hand, there is low commonality, then this means that a questionnaire will probably be doomed to failure, since any given question is likely to be relevant only to a low proportion of respondents.

3. 4. Distribution of commonality

Where commonality exists, and especially where there is also a wide range in the number of criteria used, it is instructive to look at the distribution of commonality.

If the same criteria are used by everyone, including the respondents who only use a few criteria, then this implies that those criteria are the main ones, or the “public domain” ones. If there is expertise involved, then there may well be criteria which are restricted to the expert respondents.

If, on the other hand, the commonality is scattered without any obvious pattern, then this may imply different sets of tastes, or different cultures. For example, computer games enthusiasts may have a well-developed set of criteria for evaluating machines in relation to graphics and speed, whereas system managers may have an equally well developed different set of criteria relating to performance or manufacture. In this case, it would be a good idea to look for evidence of such groupings, and to consider treating the various groups separately.

3. 5. *The categories*

After the criteria have been examined, it is time to look at the categories. Again, the simple numbers involved can be informative, in the same way as the criteria. Are there large numbers involved, or relatively few? Again, the commonality is informative: how many respondents use the same categories as each other at some point?

In addition, it is instructive to look at the distribution of categories within criteria. Not all respondents will use the same categories within the same criterion: sometimes this is because they have forgotten one, but sometimes it reflects different categorisation. If there are frequent differences of this sort, then this is an issue to be investigated in later sessions, if possible.

Another useful source of information is the "ragbag" categories, such as "don't know", "other" and "not applicable". The first of these tells you how much uncertainty there is in the population of respondents. The second and third can both mean that there are more complex issues which need to be investigated. Sometimes they only mean that there are unimportant categories which are more conveniently lumped together under "other", or that a criterion has a limited range of convenience, and that the "not applicable"s fall outside it. Sometimes, though, these categories mean that the version of categorisation being used is an over-simplified one, and that there is another layer of knowledge to be elicited.

It can also be interesting to look at the categories themselves. Are they very abstract ones, or very concrete, or a mixture of both? How subjective are they? How technical are they? What are the implications of this for the questioning? For instance, if the sessions are investigating people's perceptions of computers, with the aim of improving design briefs, then criteria such as "amount of memory" or "footprint size" imply that the respondents already take account of objective criteria; criteria such as "response time" or "how easily upgraded" are more abstract, but still objective, whereas "user-friendliness" or "smartness" mean that there is more work to be done on finding out just what goes to make up "cosiness" in the respondent's world view.

The main source of interest with the items, though, is their distribution: how much agreement is there between respondents about which items go in which categories?

Where there is agreement, the situation is straightforward. Where there is disagreement, this needs to be investigated if possible. Sometimes disagreement will be the result of human error and variation: if there are a lot of "don't know"s in the sessions, then this implies a higher likelihood of respondents disagreeing about where items go. Sometimes, though, disagreement reflects differences in definitions, or disagreements about fact, and this can be worth investigating.

The "drying up" point can also be instructive: if the same criteria occur before the drying up point for most or all respondents, and a different set of criteria occur after the drying up point for all respondents, then this implies some systematic change which needs to be investigated - for example, it may reflect the respondents' tapping in to tacit knowledge.

Significant absences Finally, one should look for significant absences: were there any criteria which did not appear, and which would have been expected to appear? If so, what might this imply? Interpreting significant absences requires good domain knowledge, and takes practice, but it is an invaluable skill once mastered.

3. 6. *Statistics*

One possible reason for the humble status of sorts in the academic community relative to rep grids is that relatively few types of statistics can be easily applied to sorts. The statistically-minded, though, might like to try applying nearest neighbour analysis, which allows the user to build up diagrams showing how closely the various items are related to each other on the basis of the sorts. This can be useful for identifying clusters of items which are similar in many ways, and, conversely, for identifying outliers, which have little in common with the other items, and perhaps should not have been included in the set used for the session.

There is more on this in the "further issues" section of this paper.

3. 7. *Analysis: conclusion*

At the end of the sorting sessions, the questioner should know the following things.

(1) How much the respondents can categorise in this area (i.e. the number of criteria and categories).
(2) What type of criteria and categories are being used by the respondents, and whether these imply further questioning, perhaps using other techniques.
(3) Which criteria and categories are being used, and which of these criteria and categories apply to each entity.
(4) How much the respondents agree or disagree with each other, and what the implications are.

4. Case study

4. 1. *Introduction*

For this case study, we used sorts to elicit the categories which a software expert uses to categorise computers. The example was chosen to be applicable both to KA and to RA.

For this case study, we decided to use sorts as an exploratory technique, without any preliminary elicitation using other techniques. We expected that we would miss some criteria and categories as a result, and used this study only to provide some starting points.

We decided that card sorts would not be practical because the respondent would probably not know a wide enough and varied enough range of computers by name to be able to sort by names alone. Object sorts were impractical, on grounds of cost and space. This left picture sorts as the best choice.

For the study, we used 11 pictures of computers from advertisement pages in a computer magazine. We cut out all the illustrations from two consecutive pages which featured illustrations of identical size and with plain backgrounds.

This meant that the pictures were already standardised, and were without distracting background clutter. We deliberately removed the accompanying text, which gave the name and cost, etc., for each computer. We did this because otherwise the respondent would have been prompted by the verbal information, and we wanted to elicit her own categorisation.

We glued the pictures onto 5″ x 3″ file cards, and numbered each card in the top right hand corner by hand (for speed of preparation). We used random numbering for the cards.

4. 2. Results

The respondent grasped the principle of the picture sort immediately, and did not attempt to sort on more than one criterion at a time in the first sort. After the second sort, the respondent asked whether the sorting should be related to a specific task, or should be in terms of any criteria which came to mind. This is interesting, because it shows that respondents can and do explicitly use different "views" or facets for their categorisations. In this study, for instance, the respondent was checking whether the criteria should be functional, or should include criteria such as aesthetic ones. Where facets are explicit, there is little point in using factor analysis in an attempt to induce indirectly the underlying factors, since the respondents are already directly supplying their own explicit description of underlying factors.

The session produced five sorts, which are described in more detail below.

The respondent provided five criteria before "drying up". It was clear from her comments afterwards that she would normally use considerably more criteria, but was not able to do so in this study, for reasons discussed below.

We did not attempt to elicit more criteria, since we already had enough for our purposes. If we had wanted to elicit more, we would have used dyadic or triadic elicitation (asking the respondent to say what was the main single difference between two or three randomly selected cards). We used only a paper record of the session, which was adequate for the sorts themselves, but would not have been adequate if we had decided to ask detailed questions about some of the comments which the respondent made, if this had been a real preliminary session.

The criteria and categories were as follows: (for brevity we have simply listed criteria and categories, and have not shown which card numbers went with which categories.)

Sort 1: type
Categories:

- PCs
- Laptops or small computers which you could carry on a train
- Small portables (bigger than laptops)

Sort 2: storage space
Categories:

- Hard disks
- Floppies
- Not sure

Sort 3: input device
Categories:

- Rollerball mouse attached to side
- No visible mouse- maybe just key input
- Mouse linked to hard disk
- Not sure

Sort 4: what is displayed on the screen
Categories:

- Spreadsheets/bar graphs of data
- Graphical
- Program managers
- Don't know

Sort 5: colour (for what it's worth!) (sic)
Categories:

- Black/dark grey
- Beige/creamy colours

In discussion after the session, the respondent spontaneously mentioned that she had considered using screen colour (monochrome versus full colour) as a criterion, but had decided against it because that would have led to a sort with only one item in the "monochrome" category and the other ten items in the "full colour" category. She said that even though there had not been any instruction to use categories containing roughly equal numbers of items, she had felt reluctant to use very unequal sized categories.

4.3. Analysis of case study

The criteria and categories used were all linked to tangible characteristics of the computers themselves. If card sorts of known entities rather than picture sorts of unfamiliar entities had been used, then the respondent might well have used intangible characteristics such as cost and reliability. Such intangibles are likely to be systematically missed in sorts involving entities which are not known to the respondent.

The respondent mentioned one practical, non-computing, criterion, i.e. whether or not the computer could be carried on a train. In comments during the session, she also referred to the criterion of whether or not the computer could fit inside a briefcase. This sort of practical constraint is useful for product researchers and requirements engineers, since it helps them to establish the design "envelope" within which a successful product should fit.

One interesting criterion was "colour (for what it's worth!)". This referred to the colour of the casing of the machine, not the colour of the screen display. It is significant that the respondent felt it necessary to make a joke of the criterion, but equally significant that she mentioned it at all. This implies that there might be scope for wider use of colour in computer casings, and market researchers who encountered a finding such as this would want to follow it up.

4.4. Discussion of case study

Although the session went smoothly, it was clear from the respondent's comments during and after the session that we were missing a significant number of criteria.

During the session, the subject repeatedly peered at the pictures, trying to make out fine detail which was not properly visible because of the small size of the pictures. After the session, the respondent spontaneously said that if she was choosing a computer she would want to know about other criteria as well as those elicited in the session; she named physical size, storage space, memory, number of disk drives, cost, and compatibility as relevant criteria. None of these could be inferred from the pictures.

It was therefore clear that picture sorting alone would not provide adequate coverage for this domain, and that other techniques would need to be used as well in order to provide full coverage. This is scarcely surprising, since no single technique is able to access all the types of memory and knowledge which may be needed in a domain (Maiden and Rugg, 1996).

The picture sorts did, however, elicit two issues which merited following up, namely practical non-computing criteria (such as fitting into a briefcase) and casing colour. In addition, the criteria which were elicited were acceptable to the respondent - she did not judge them to be a distortion of her knowledge. The session therefore appears to have produced a valid, but only partial, insight into this domain.

The conclusion from these findings would be that (for this respondent at least) physical size, storage space, memory, number of disk drives, cost, and compatibility were relevant factors not inferable from the pictures alone, and therefore not elicited using sorts. In addition, it would be advisable to investigate further the issues of colour and of physical features such as fitting into a briefcase, possibly using techniques such as laddering.

Most of the discussion above deals with problems encountered, so it is worth reiterating the positive aspects of the case study. The session proceeded smoothly, although the technique was completely new to the respondent, and produced a large volume of information in a regular, formalised representation which did not cause significant distortion of that information, and which did not require subsequent transcription or coding. These are significant advantages.

Our own experience of using the sorting techniques has been that they are well worth using. They offer simplicity and speed, while allowing respondents to use their own categorisation; the format used makes it easy to compare results across respondents. Our experience of teaching these techniques to other people is that the techniques are easily learned, and that they tend to be used repeatedly, rather than being abandoned after the initial learning session. We can therefore recommend them as a practical technique both for knowledge acquisition and for requirements acquisition.

5. Further issues: category theory and semantics

Sorting involves categorisation, and there is a sophisticated literature on categorisation, taxonomy and semantics, all of which are potentially relevant.

One issue which is particularly relevant involves fuzzy sets (Zadeh, 1965) and prototype theory (Rosch and Lloyd, 1978). Sorting implies a clear-cut, yes/no decision about the categories into which something is sorted, whereas life is often more complicated, and categories may grade into each other. Even as simple a category as "aquatic animal" versus "land-living animal" exhibits this gradation, with a spectrum from whales, which are fully aquatic, through seals, which are mainly aquatic but spend time on land, via otters, which are equally at home in either environment, through to sloths, which are strongly land-lovers (but can, apparently, swim if need arises). This is the subject area of fuzzy set theory and prototype theory, which respectively provide mathematical and psychological methodologies for handling this issue.

Fuzzy sets work by mathematically specifying the degree to which an item belongs to a set, ranging from full membership to no membership. This overlaps with, but is logically distinct from, prototype theory. Prototype theory defines a set in terms of a number of criteria for set membership; the more of these criteria which apply to a particular entity, the more prototypical that entity is of the set. The usual example of this is birds, which prototypically have feathers, make nests, live in trees, sing, etc.; robins display high prototypicality, whereas penguins display low prototypicality. Prototype theory therefore always involves a number of criteria in set membership, whereas fuzzy sets may involve several criteria or only one.

Where gradations of this sort are likely to be an issue, other techniques such as repertory grids may be more appropriate than the sorting techniques. However, it is also possible to use repeated, single criterion sorts to tackle this, by addressing the individual constructs used to define the categories. In the example above, the signal that fuzzy sets might be involved would probably come from respondents hesitating about which category to put which entities in; some probing would (with luck) reveal the underlying problem, and identify the component constructs.

It is clear that the sorting techniques are highly compatible with prototype theory, though fuzzy sets may be more problematic. Identification of prototypical characteristics could be approached by e.g. pooling results from a number of experts, and then using the frequency with which a characteristic was mentioned across experts as an indication of that characteristic's role in prototypicality for the item in question.

6. Conclusion

The sorting techniques are an invaluable part of the knowledge engineer's or requirements engineer's tool kit. They are simple to use, and combine flexibility of use with a highly formalised representation formalism, bridging the gap between qualitative and quantitative techniques. In addition, they are easily computerised. They therefore offer the same attractions as the more familiar repertory grid approach, which they complement well.

References

BANNISTER, D. and F. FRANSELLA (1980) *Inquiring Man*, Harmondsworth: Penguin.

BARR, A. and E.A. FEIGENBAUM (1982) *The Handbook of Artificial Intelligence*, London: Pitman.

BOOSE, J.H., D.B. SHEMA and J.M. BRADSHAW (1989) Recent progress in AQUINAS: A knowledge acquisition workbench, *Knowledge Acquisition*, **1**, 185–214.

CHI, M.T.H., R. GLASER and M.J. FARR (eds.) (1988) *The nature of expertise*,London: Lawrence Erlbaum Associates.

CORBRIDGE, C., G. RUGG, N. P. MAJOR, N.R. SHADBOLT and A.M. BURTON (1994) Laddering: Technique and tool use in knowledge acquisition, *Knowledge Acquisition*, **6**, 315–341.

CULLEN, J. AND A. BRYMAN (1988) The knowledge-acquisition bottleneck: Time for reassessment, *Expert Systems*, **5**, 216–225.

ELLIS, C. (ed.) (1989) *Expert Knowledge And Explanation: The Knowledge-Language Interface*, Chichester: Ellis Horwood.

GAMMACK, J.G. (1987) Different techniques and different aspects of declarative knowledge, in Kidd, A.L. (ed.), *Knowledge Acquisition for Expert Systems; A Practical Handbook*, New York: Plenum Press.

GOFFMAN, E. (1955) *The Presentation Of Self In Everyday Life*, New York: Doubleday.

KAHNEMAN, D., P. SLOVIC and A. TVERSKY (1982) *Judgement Under Uncertainty: Heuristics And Biases*, Cambridge: Cambridge University Press.

KELLY, G.A. (1955) *The Psychology Of Personal Constructs*, New York: W.W. Norton.

MAJOR, N.P. (1991) CATO — an automated card sort tool, in M. Linster and B. GAINES (EDS.), *Proceedings of EKAW-91, GMD-Studien Nr. 211*, September 1992.

MAIDEN, N.A.M. and G. RUGG, G. (1996) ACRE: a framework for acquisition of requirements, *Software Engineering Journal*, May, 183–192.

MCGEORGE, P. and G. RUGG, (1992) The uses of "contrived" knowledge elicitation techniques, *Expert Systems*, **9**(3).

OSGOOD, C.E., G.J.SUCI and P.H.TANNENBAUM (1957) *The Measurement of Meaning*, Illinois: University of Illinois Press.

ROSCH, E. and B.B. LLOYD, (eds.) (1978) *Cognition and Categorisation*, Hillside, New Jersey: Lawrence Erlbaum Associates.

RUGG, G., C. CORBRIDGE, N.P. MAJOR, A.M. BURTON and N.R. SHADBOLT (1992) A comparison of sorting techniques in knowledge elicitation, *Knowledge Acquisition*, **4**(3), 279–291.

RUGG, G. and P. MCGEORGE (1995) Laddering, *Expert Systems*, **12**(4), 279–291.

RUGG, G. and N.R. SHADBOLT (1991) On the limitations of repertory grid technique in knowledge acquisition, *Proceedings of the 6th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop, Banff, Canada*, **2**, 22–1 to 22–17.

SEGER, C.A. (1994) Implicit learning, *Psychological Bulletin*, **115**(2), 163–196.

SHADBOLT, N.R. and B.J. WIELINGA, B.J. (1990) Knowledge based knowledge acquisition: The next generation of support tools, in B.J. Wielinga, J. Boose, B. Gaines, G. Schreiber and M.W. van Someren, (eds.), *Current Trends in Knowledge Acquisition*, Amsterdam: IOS Press, 313–338.

SHAW, M.L.G. (1980) Recent Advances in Personal Construct Technology, London: Academic Press.

SHAW, M.L.G. and B.R. GAINES (1988) A methodology for recognising consensus, correspondence, conflict and contrast in a knowledge acquisition system, *Proceedings of Workshop on Knowledge Acquisition for Knowledge-Based Systems, Nov 7–11, 1988, Banff, Canada*.

STEPHENSON, W. (1953) *The Study Of Behavior: Q-Technique and its Methodology* Chicago: University of Chicago Press.

VICKERY, B.C. (1960) *Faceted Classification: A Guide to the Construction and Use of Special Schemes*, London: Aslib.

YORK, D.M. (1983) *The Repertory Grid: A Critical Appraisal*, unpublished PhD thesis, University of Nottingham, UK.

ZADEH, L. (1965) Fuzzy sets, *Information and Control*, **8**, 338–353.

Appendix 1. Sample instructions

You will be given some cards to sort. Each card will have the name of an object written on it.

We would like you to sort the cards into groups, using one criterion at a time. When you have finished sorting, please tell us what the criterion was for that sort, and what the groups were into which you sorted the cards, so that we can record this. Once this has been done, we would like you to sort the cards again, using a different criterion, and then to keep on sorting them until you have run out of criteria.

For example, if the task was sorting different types of car, your first criterion might be “place of manufacture” and the groups might be “American”, “British”, “French”, etc.; the second criterion might be “cost”, with the groups being “expensive”, “medium” and “cheap”.

You are welcome to use any criteria you like, and any groups you like, including “don’t know”, “not sure” and “not applicable”. The main thing is to use only one criterion in each sort - please don’t lump two or more in together. If you’re not sure about something, just ask.

You may have noticed that the cards are numbered: this is for convenience when recording the results. The numbering is random, so please don’t use that as a criterion for sorting!

If you have any comments or questions, then please say, and we will sort them out.

Thank you for your help.

The authors

Gordon Rugg

Gordon Rugg’s first degree was in French and Linguistics, which was followed by a PhD in Psychology. His work since then has included formal comparison of knowledge acquisition techniques, and researching requirements acquisition in safety-critical domains. His current work includes requirements acquisition, risk research, on-line information retrieval, social factors affecting the acceptance of new technologies, and knowledge management.

Peter McGeorge

Peter McGeorge is a lecturer in the department of Psychology, University of Aberdeen. His academic qualifications include a BSc in Behavioural Science and a PhD in Psychology, both gained at the University of Nottingham. His interest in expert systems relates mainly to the development of effective techniques of knowledge elicitation. His work has included studies of medical expertise using a range of knowledge elicitation methods. He is particularly interested in the nature of expertise and the role of implicit knowledge in cognition.