British Journal of Psychology (1987), **78**, 99–112 *Printed in Great Britain*

Pictorial communication of abstract verbs in relation to human–computer interaction

Yvonne Rogers and **D. J. Oborne**

Graphic symbols, commonly known as icons, are becoming increasingly popular as a medium to present information about computer systems and their command operations. This paper considers the extent to which such pictorial representations can be generated for verb actions like those used at the computer interface and, subsequently, how effective they are in communicating their intended meaning.

An initial experiment showed that the majority of subjects were able to produce drawings for a set of verbs, but found it much easier to do so for high imagery verbs compared with low imagery verbs. Furthermore, drawings produced for the high imagery verbs were considered to be the most representative and also had the highest stereotype strength. The types of representation found in these drawings were generally complex, depicting concrete objects in conjunction with conventional symbols and various line supplements. The results from this study are discussed in relation to various factors which contribute to the production of pictorial images and how these may affect the communicativeness of such images when used as icons at the computer interface.

Recent trends in human–computer interaction have moved towards the use of iconic representations in which information about aspects of the system is presented on the screen in pictorial form. Such interfacing techniques have recently become implemented in a number of office automation systems, for example, the Xerox Corporation's 'STAR' (Smith *et al.*, 1982) and Apple Corporation's 'LISA' system (Williams, 1983). Other, more general purpose, systems such as Unix's 'UNICON' have also recently become icon-driven (Gittins *et al.*, 1984). Typically the information represented in these iconic interfaces is of 'office-like' objects such as in- and out-trays, files, folders and wastepaper bins. Additionally, icons have been used to represent command operations, such as deleting, creating and printing files or sets of data.

One reason for the increased popularity of such symbols may stem from the fact that graphic symbols are often considered as a potentially universal means of communication which can convey certain types of information more directly and with more immediacy than can words. For example, pictograms used for public information and road signs have been interpreted more easily (e.g. Dewar & Swanson, 1972) and more accurately (Walker *et al.*, 1965) than word signs containing the same message; symbol meanings can be learned rapidly (Walker *et al.*, 1965), with minimum confusion among alternatives (Green & Pew, 1978). Furthermore, studies by Haber and his co-workers (Haber, 1970; Haber *et al.*, 1970), which show that the memory capacity for pictures may be unlimited, suggest that information from visual images may also be among the most sensitive form of memorial storage.

In terms of the universality of the communication mode, Hemenway (1982) suggests that the advantages of iconic systems over verbal representations are similar to the benefits of other graphic systems, such as those used for maps. Compared with a set of words, she suggests, graphical symbols can be visually more distinctive from one another and can represent variation within a set of commands more effectively. Rohr & Keppel (1984) further suggest that using icons enables certain types of complex information to be presented in a more holistic and spatially condensed form.

Despite its apparent advantage, however, graphic communication at the computer

interface presents a major problem. This concerns the extent to which icons can effectively represent the types of complex information associated with the functioning of the system. Most software systems generally have a complex structure with a large number of states and operations. The question needs to be asked, therefore, whether it is possible for these types of abstract concepts to be represented meaningfully in a pictorial form and, at the same time, still retain the advantages claimed for other graphic systems.

From the psychological literature it appears that little is known about how visual symbols representing abstract concepts are processed and, moreover, how well they can be understood. Since it becomes more difficult to represent abstract concepts the more abstract they become, there is a strong preference for depictions which attempt to concretize the concept in some way (e.g. Jones, 1983). Paivio (1971) suggests that this process occurs by means of associative chaining. For example, the word religion may evoke an image of a church as an associative reaction. However, the directness and immediacy of the aroused image is likely to depend on the idiosyncratic associative frequency in the perceiver's experience. Thus the word 'liberty' may immediately evoke an image of the 'Statue of Liberty', especially for people living in New York. Furthermore, Paivio suggests that such an imaginal reaction may not in principle be any different from that occurring to a more concrete word such as 'dog'.

The extent to which abstract concepts such as those involved in the functioning of computer systems are able to evoke imaginal referents, therefore, may depend to a large extent on how well they can be concretized. In addition, it would appear that this process is affected by the degree of abstractness underlying the concept and the frequency with which it is experienced. The purpose of this study was initially to determine whether these attributes affected the extent to which pictorial representations could be produced and subsequently comprehended for a set of verb actions such as those typically used as computer commands.

This paper reports two experiments that address this question. The first study was designed to investigate the extent and type of drawings that might be produced for a set of verbs of varying frequency and imagery levels. It was hypothesized that verbs which have high imagery and high frequency associations would be easier to represent visually and subsequently comprehend than verbs associated with low imagery and low frequency values.

A second aim of this experiment was to evaluate the types of drawings generated. Previously, research has shown a strong preference for using concrete objects as referents for abstract nouns (Jones, 1983). These can also be accompanied by conventional abstract symbols such as ticks, crosses and various forms of lines for other concepts such as those employed in military intelligence concepts (Howell & Fuchs, 1968). The use of both concrete objects and other supplements, therefore, may be necessary for a pictorial representation to convey an ongoing action.

In this first experiment, the sample of verb representations was context-free. This means, therefore, that minimum constraints were made on the semantic domain in which the drawings were produced. The second experiment, therefore, was designed to investigate the drawings judged to be the most meaningful with respect to their intended referent's normal definition in the English language and to assess the degree of stereotyping of the drawings for each verb. It was hypothesized that the drawings considered to be most comprehensive would be those depicting concrete objects in combination with abstract symbols. Similarily the verbs which would exhibit the highest degree of stereotyping would be those associated with high imagery and high frequency.

Experiment 1

The experiment was designed to investigate the extent to which verbs such as those used as computer commands varying in concreteness (indexed by imagery associations) and word frequency could be represented in pictorial form. It was hypothesized that subjects would find it easiest to produce drawings for high imagery and high frequency verbs than for low imagery and low frequency ones. A further aim of the experiment was to determine whether there was a preference for using concrete objects rather than abstract symbols in the drawings, or a combination of both.

To obtain drawings for the word stimuli the sign production method was initially used. This technique, developed and validated by Howell & Fuchs (1968), has frequently been used to obtain pictorial representations of verbal concepts for use in visual communication (see, for example, Karsh & Mudd, 1962; Szlichcinski, 1980; Collins, 1982; Jones, 1983). The rationale behind this technique is that the drawings that people generate when asked to represent a particular referent are the ones they would also be most likely to understand and recognize. Whereas this method has been questioned (e.g. Szlichcinski, 1980) on the grounds that the symbol produced most frequently will not necessarily be the one most easily comprehended, it is argued that it is useful in that it enables context-free attributes of visual representations to be produced easily. Furthermore, the drawings produced can indicate how people expect verb actions to be represented and are related to depictions of associated objects.

Method

Subjects. A total of 93 first year undergraduates studying psychology at the University College of Swansea acted as subjects.

Materials. Four sets of eight 'command-like' verbs were selected from a list of verbs whose imagery values had been obtained in a previous experiment (Rogers & Oborne, 1985). The verbs were chosen by varying the imagery and frequency values. These two factors were orthogonally manipulated, thus producing four categories:

(1) high imagery – high frequency
(2) high imagery – low frequency
(3) low imagery – high frequency
(4) low imagery – low frequency

High imagery verbs were classified as having an associated imagery rating of $< 2{\cdot}5$ on a seven-point scale (1 = high imagery and 7 = low imagery) and low imagery words were classified as being $> 4{\cdot}5$. Word frequency values were taken from Thorndyke & Lorge's (1944) wordbook. Low and high frequency verbs were classified as having a frequency less than 50 and greater than 100 occurrences/million printed words respectively. The stimuli used, with their associated imagery ratings and frequency values, are presented in Table 1.

Each verb was typed in upper case letters at the head of a sheet of paper. To control for the size of each drawing to be produced by the subject, an outlined box measuring 95×55 mm was drawn beneath it. Below this a seven-point difficulty rating scale was presented. Booklets of eight verbs (two taken randomly from each of the four sets) were then constructed. A set of instructions was also provided at the beginning of each booklet.

Procedure. Subjects were instructed to produce drawings that they felt represented the meaning of the set of verbs. They were asked to draw in the box the first thing that came to mind when they read the word and to work as fast as they could. It was also emphasized that their drawings should be made as simple as possible and that the quality was not important. They were told instead that what was important was how they tried to represent the meaning of the word. No other instructions were given.

Table 1. The four verb sets used in Expt 1 with their associated imagery (median) ratings and word frequency

Set/verb	Imagery rating	Frequency	Difficulty rating	Drawings produced (%)	No. of categories	W_{max} (%)
High imagery–high frequency						
OPEN	2·00	> 100	5·26	100	3·25	82·41
WRITE	1·44	> 100	4·86	100	3·88	51·71
CLOSE	2·50	> 100	3·48	100	4·50	47·28
READ	2·33	> 100	5·26	100	3·38	53·80
STOP	2·33	> 100	5·24	100	5·13	48·74
CALL	2·40	> 100	4·85	100	4·38	54·73
DRAW	1·83	> 100	4·48	100	5·13	35·43
CUT	2·00	> 100	5·48	100	3·38	62·46
High imagery–low frequency						
POKE	2·30	12	4·48	100	4·00	51·90
DEMOLISH	1·44	3	4·32	100	4·13	62·74
SKIP	2·19	17	5·21	100	4·00	71·78
FREEZE	1·83	32	4·00	95·8	6·38	40·22
PUNCH	1·55	17	4·91	100	4·88	52·28
CHASE	1·64	48	4·00	95·3	3·89	64·44
SCRIBBLE	1·64	3	5·88	100	3·88	73·35
PEEP	2·28	27	4·32	100	4·38	56·38
Low imagery–high frequency						
END	5·50	> 100	2·90	100	4·75	48·00
GET	5·17	> 100	3·29	95·5	5·75	41·30
REPORT	5·25	> 100	3·43	91·3	5·13	45·35
BEGIN	5·00	> 100	2·69	88	5·38	40·52
REMOVE	5·00	> 100	3·42	96	5·25	52·11
RETURN	5·00	> 100	3·82	100	5·38	45·79
INCREASE	5·00	> 100	4·63	100	4·75	49·47
START	4·83	> 100	3·38	95·5	4·00	75·54
Low imagery–low frequency						
QUIT	5·64	49	2·68	91·7	6·00	38·15
APPEND	6·66	1	1·72	53	5·25	38·17
ASSIGN	6·38	24	2·19	76·2	4·13	52·61
EVALUATE	6·39	1	2·82	95·7	5·25	45·53
COMPILE	5·50	5	2·75	90·9	4·88	41·10
SUBSTITUTE	5·00	29	2·71	96	4·75	46·23
SCAN	5·00	8	3·00	100	4·88	41·24
TRANSFER	5·17	28	3·42	91·3	4·75	53·47

Note. Other metrics shown are the confidence ratings and percentage of drawings produced in Expt 1, the mean number of categories judges used and the stereotypy values (as indexed by W_{max}) derived in Expt 2.

After finishing each drawing subjects were required to use the scale provided to rate the ease of representing the meaning of the word in this way (1 = very easy; 7 = very difficult). No time limit was set but, on average, subjects tended to take 10–15 min to complete the task.

Results and Discussion

The number of drawings produced for the verbs was very high, with 97·8 per cent of subjects producing drawings for the high imagery verbs and 82·7 per cent producing

drawings for the low imagery verbs. The percentage of drawings produced for each verb stimulus is presented in Table 1.

Table 1 also shows the mean difficulty ratings for each of the stimuli. A two-way within-subject analysis of variance (ANOVA) performed on the data revealed a large main effect for imagery ($F = 233{\cdot}94$, d.f. = 1, 92, $P < 0{\cdot}001$) and a smaller, although still significant, effect for word frequency ($F = 31{\cdot}23$, d.f. = 1, 92, $P < 0{\cdot}001$). These results suggest, therefore, that producing drawings for high imagery verbs is considered as being a much easier task than for low imagery verbs (see Fig. 1). Similarily, drawings for high frequency verbs were judged as being easier to produce when compared with low frequency verbs, but to a lesser extent. Interestingly, a small interaction was found between the two variables ($F = 8{\cdot}9$, d.f. = 1, 92, $P < 0.005$). An analysis of the treatment means, using the Tukey *post hoc* test, revealed that there was a significant difference ($\bar{d}_T = 0{\cdot}538$, $P < 0{\cdot}01$) between the low imagery means but not between the high imagery means. These results suggest, therefore, that word frequency has a slight modifying effect on low imagery verbs, which is not apparent for high imagery verbs.

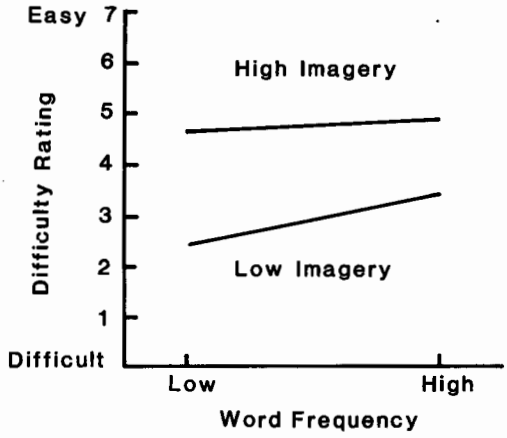

Figure 1. Difficulty ratings (mean for 93 subjects) for generating pictorial representations as a function of imagery and word frequency.

The drawings themselves were initially analysed by the experimenter in terms of whether they represented concrete objects, abstract symbols, or a combination of both. For a representation to be considered as concrete it had to depict obvious objects such as a pair of scissors or a person. Representations which were judged as being totally abstract were those which consisted of mathematical formulae, geometric figures, nonsense shapes or various forms of lines. Of the 706 drawings produced, 417 (59 per cent) were judged as depicting concrete objects and abstract symbols, 177 (25 per cent) as depicting just concrete objects, whilst only 112 (16 per cent) were considered as totally abstract.

The number and percentage of the various types of drawings for the different conditions are presented in Table 2. Interestingly, the frequency of the drawings was found to be significantly different across conditions ($\chi^2 = 83{\cdot}07$, d.f. = 6, $P < 0{\cdot}001$). The main difference found in the results is between the distribution of drawings consisting of exclusively concrete representations and those depicting both concrete and abstract representations. Whereas in the high imagery–high frequency (HI–HF) condition over 60 per cent of the drawings were considered as totally concrete and 30 per cent as both concrete and abstract, the opposite was true for the low imagery–low frequency (LI–LF) condition. Moreover, the proportion of abstract drawings in this condition was greater, meaning that there were fewer concrete drawings. Thus there was a strong tendency for subjects to concretize their representations using only objects for HI–HF verbs. To

Table 2. Percentage of drawings (actual number in parentheses) judged as being concrete (C), abstract (A) or both (CA) for each verb set

Verb set	C	A	CA
HI–HF	61 (113)	8·9(17)	30·1 (56)
HI–LF	31 (57)	20·1(37)	48·9 (90)
LI–HF	37 (66)	15·2(30)	47·8 (85)
LI–LF	15·4(25)	21·1(35)	63·4(105)

represent the meaning of the LI–LF verbs and to a lesser extent the verbs in the other conditions, however, it appears that both concrete objects and abstract symbols were required.

In the drawings which were judged as depicting both concrete objects and abstract symbols, various types of symbols were used – although arrows and dotted, dashed and other lines were the most common. The function of these universals (i.e. arrows, crosses, etc.) and supplements (i.e. line marks and dots, etc.) appeared to be to act as indicators of a change of state, ongoing motion or an action that is about to be, or has been, carried out. Thus concrete objects associated with a particular verb were often considered as not being sufficient in themselves to represent the meaning of the verb and additional line marks were required. Examples of verbs where arrows were often used to imply a change in state were 'to compile' and 'to substitute'. A typical representation of 'to compile' was of an arrow arising from several objects to a larger object. A common object used for this was a book. The implication from this particular type of drawing, therefore, is that the contents of several books are being used to make one book (see Fig. 2*a*). Likewise for 'to substitute': arrows between two footballers indicate a change of state; the direction of the arrow indicates that one person is taking the place of another (Fig. 2*b*).

Examples where arrows were used to indicate actual motion of objects, on the other hand, were in the drawings for the verbs 'to chase' and 'to return'. In these cases arrows were drawn in the path of a person appearing to run after another person or walking towards a house, respectively (Figs 2*c* and 2*d*). Lines which were either dotted or dashed were also often used to indicate motion such as the trajectory of a ball. Another function of these types of supplement was to portray ongoing 'invisible' action such as for the representation of the verb 'to scan' where several drawings used dashed lines to connect the eyes of people with objects on the horizon (Fig. 2*e*).

An alternative way to suggest a change in state, used by several subjects, was to present a series of two or more representations. For example, several drawings of the verb 'to increase' were represented by a series of objects or geometrical figures which were drawn increasing in size or number (Fig. 2f).

For the more abstract verbs such as 'to evaluate', think balloons and question marks above matchstick people's heads were often used in conjunction with a 'judgement' scene such as two differently priced objects (Fig. 2*g*). The function of the abstract symbols in these cases was not so much to indicate actual motion but to link physical objects with a particular thought process, such as decision making. These types of abstract representations have, in fact, become generally accepted as universal symbols and have become familiar to us through our exposure to cartoon and comic drawings.

The inclusion of supplements and universals with concrete representations, therefore, seems to suggest that, apart from the HI–HF verbs, drawings of actions generally require some indication of movement from one state to another. Simple representations of objects

drawings for the low imagery verbs. The percentage of drawings produced for each verb stimulus is presented in Table 1.

Table 1 also shows the mean difficulty ratings for each of the stimuli. A two-way within-subject analysis of variance (ANOVA) performed on the data revealed a large main effect for imagery ($F = 233{\cdot}94$, d.f. = 1, 92, $P < 0{\cdot}001$) and a smaller, although still significant, effect for word frequency ($F = 31{\cdot}23$, d.f. = 1, 92, $P < 0{\cdot}001$). These results suggest, therefore, that producing drawings for high imagery verbs is considered as being a much easier task than for low imagery verbs (see Fig. 1). Similarily, drawings for high frequency verbs were judged as being easier to produce when compared with low frequency verbs, but to a lesser extent. Interestingly, a small interaction was found between the two variables ($F = 8{\cdot}9$, d.f. = 1, 92, $P < 0.005$). An analysis of the treatment means, using the Tukey *post hoc* test, revealed that there was a significant difference ($\bar{d}_T = 0{\cdot}538$, $P < 0{\cdot}01$) between the low imagery means but not between the high imagery means. These results suggest, therefore, that word frequency has a slight modifying effect on low imagery verbs, which is not apparent for high imagery verbs.

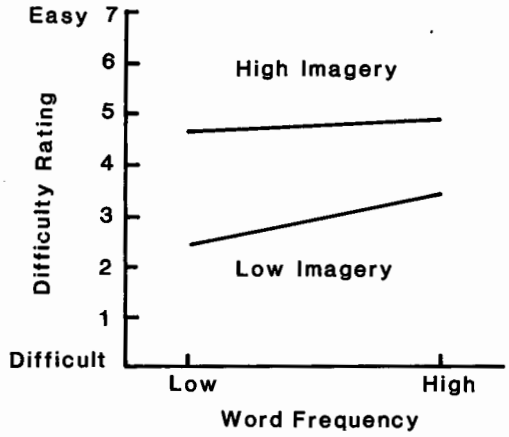

Figure 1. Difficulty ratings (mean for 93 subjects) for generating pictorial representations as a function of imagery and word frequency.

The drawings themselves were initially analysed by the experimenter in terms of whether they represented concrete objects, abstract symbols, or a combination of both. For a representation to be considered as concrete it had to depict obvious objects such as a pair of scissors or a person. Representations which were judged as being totally abstract were those which consisted of mathematical formulae, geometric figures, nonsense shapes or various forms of lines. Of the 706 drawings produced, 417 (59 per cent) were judged as depicting concrete objects and abstract symbols, 177 (25 per cent) as depicting just concrete objects, whilst only 112 (16 per cent) were considered as totally abstract.

The number and percentage of the various types of drawings for the different conditions are presented in Table 2. Interestingly, the frequency of the drawings was found to be significantly different across conditions ($\chi^2 = 83{\cdot}07$, d.f. = 6, $P < 0{\cdot}001$). The main difference found in the results is between the distribution of drawings consisting of exclusively concrete representations and those depicting both concrete and abstract representations. Whereas in the high imagery–high frequency (HI–HF) condition over 60 per cent of the drawings were considered as totally concrete and 30 per cent as both concrete and abstract, the opposite was true for the low imagery–low frequency (LI–LF) condition. Moreover, the proportion of abstract drawings in this condition was greater, meaning that there were fewer concrete drawings. Thus there was a strong tendency for subjects to concretize their representations using only objects for HI–HF verbs. To

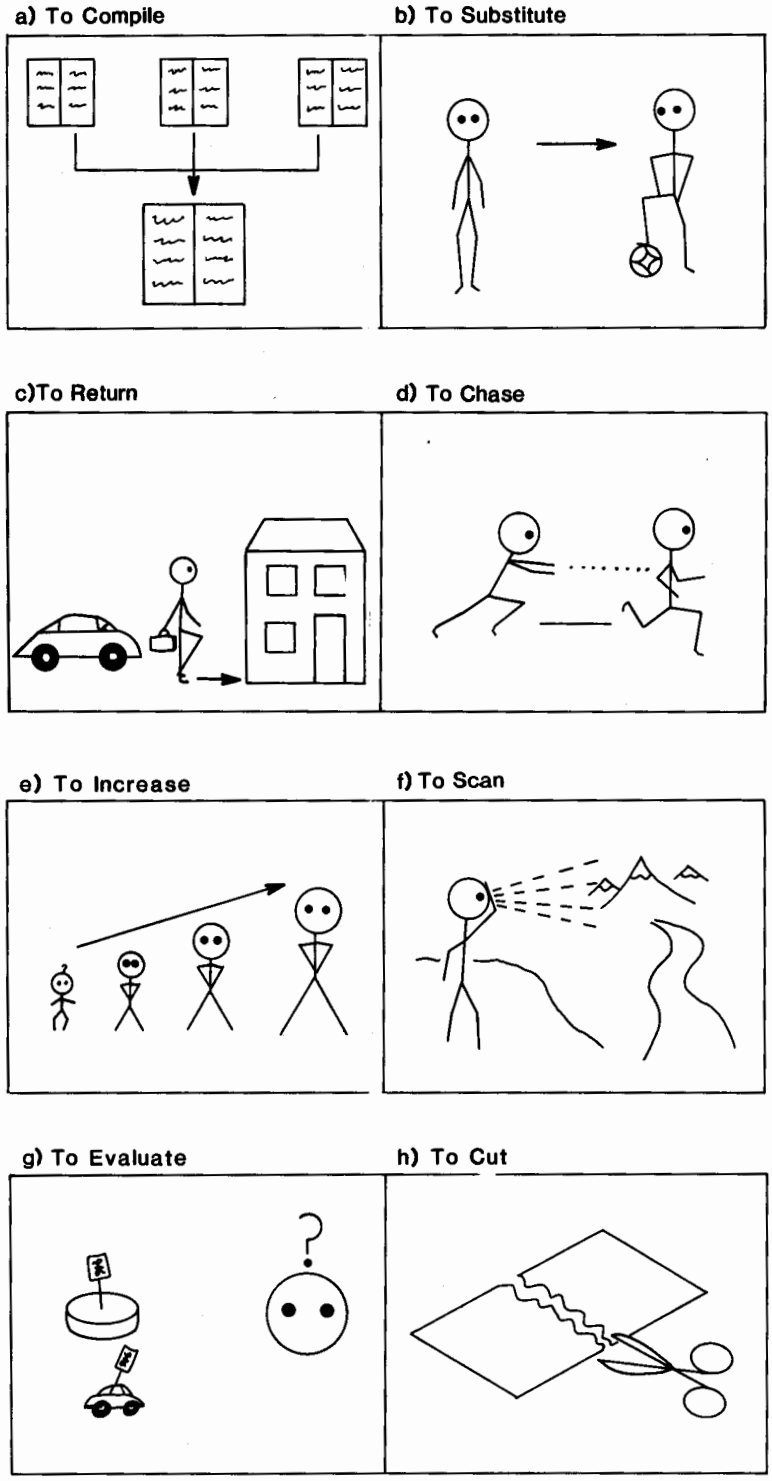

Figure 2. Examples of some of the drawings produced for the verbs.

associated with the verb may not be sufficient to depict adequately the meaning of a verb. To determine whether this is true for the comprehension of these representations, a further experiment was carried out to assess people's judgement of the ideas being expressed and what they considered as the most representative.

Experiment 2

The aim of this experiment was to investigate the extent to which the drawings produced in Expt 1 were judged as representing the meaning of the verb they portrayed. The drawings were analysed in terms of the degree of stereotyping exhibited. This is a measure which indicates the proportion of drawings judged as being the most appropriate, relative to the total number of drawings generated for each verb. The criteria used to determine the type of categorization judges were to use to evaluate the drawings were based on those used by Jones (1983).

The extent to which imagery and word frequency influences the degree of stereotyping was again evaluated. It was hypothesized, on the basis of the results of Expt 1, that high imagery verbs would show higher stereotyping than low imagery words and that word frequency would have a similar but smaller effect.

Method

Subjects. Eight students (four male, four female) aged between 19 and 36 acted as subjects.

Procedure. The drawings produced in Expt 1 were cut from the booklets and sorted into piles according to the verb they represented. The piles were placed into 32 separate envelopes. Each judge was then presented with each envelope separately and told the name of the verb which the contents were supposed to represent. They were then instructed to sort the drawings into piles so that each pile contained drawings that expressed the same idea. It was emphasized that the judges should sort the drawings according to the similar ideas being expressed, rather than just similar objects being represented. An example of the verb 'to help' was given, which illustrated how the concept may be represented by a variety of pictures showing people helping others in a task. It was suggested that in this example the idea being expressed in the drawings was the same, but that the tasks and number of people may vary. Alternatively, another set of drawings might represent a different idea of 'to help' such as 'to provide financial support'.

No time limit was given and the judges were told they could sort the drawings into as few or as many piles as they wanted. In addition they were given the option of placing drawings which they considered as meaningless, and totally inappropriate as a representation of the verb, into an 'irrelevant' pile. After sorting the drawings for each verb, the judges were asked to express verbally the idea which they thought each pile depicted. Finally, they were asked to rank the piles of drawings in order of how well they represented the meaning of the verb.

The order in which the envelopes were presented to each judge was varied. In addition, the drawings for each verb were well shuffled each time they were placed in the envelope.

Results

The number of categories used by each subject when sorting the drawings for a particular verb ranged between one and nine. The modal number of categories for each verb was four.

The category data were analysed using a three-way ANOVA treating imagery and word frequency as fixed factors and verb stimuli as a nested random factor (see Table 3). Imagery was found to be highly significant ($F = 18{\cdot}87$, d.f. = 1, 217, $P < 0{\cdot}001$) but no effect was found for word frequency ($F = 0{\cdot}86$, d.f. = 1, 217). A significant effect was found, however, for verb stimuli ($F = 2{\cdot}30$, d.f. = 28, 217, $P < 0{\cdot}001$) indicating that the results generalize across both subjects and stimuli.

The mean number of categories for each verb are also presented in Table 1. To evaluate the contribution of individual stimuli to the overall imagery effect a *post hoc* analysis was

performed on the means within each treatment. Using the Tukey test ($d_T = 1{\cdot}94$, $P < 0{\cdot}05$) it was found that only one stimulus mean was significantly greater than two or more of the other means within a treatment. This was for the verb 'to freeze', whose mean number of categories was 6·38. Although this stimulus is a high imagery verb it would appear that generally there was a tendency for judges to use more categories for low imagery verbs than for high imagery verbs.

A similar effect to that shown for the category data was also found in the distribution of drawings judged to be irrelevant. Whereas the mean number of drawings discarded for the HI–HF and HL–LF conditions was only 9·25 and 12·38, respectively, it was 22·5 and 28·13 for the LI–HF and LI–LF conditions. Furthermore, a large main effect was found for imagery ($F = 45{\cdot}02$, d.f. = 1, 7, $P < 0{\cdot}001$), indicating that far more low imagery verbs were considered as irrelevant compared with high imagery verbs. A main effect was also found for word frequency ($F = 14{\cdot}81$, d.f. = 1, 7, $P < 0{\cdot}001$) showing that, likewise, more low frequency verbs were considered as irrelevant than high frequency verbs.

To determine whether these effects were also present in the degree of stereotyping for the different sets of verbs, the frequency of drawings in each category was assessed. Using the method described by Jones (1983) [based on Howell & Fuchs', 1968, technique] the stereotype weight (W) was calculated. This is a measure of the number of drawings in each category expressed as a proportion of the total number of drawings produced (excluding the discards) for a particular verb. To determine the strength of the population stereotype (W_{max}) the W value for the category containing the largest number of drawings is then taken. This value, then, indicates the extent of agreement among the subjects producing the drawings of the most frequent interpretation of the concept. Table 1 shows W_{max} for the four sets of verbs. As can be seen, generally the strongest stereotypes (as indexed by the highest percentages) were for those categories in the two high imagery sets. As expected, however, this relationship was found to depend on the number of categories selected for each verb ($r = -0{\cdot}77$, $P < 0{\cdot}001$). This is because the greater the total number of categories selected, the smaller will be the pool of remaining drawings that can contribute to W_{max}. Thus the size of W_{max} tended to decrease in relation to the increasing number of categories.

A three-way ANOVA, again treating imagery and word frequency as fixed factors and verb stimuli as a nested random factor, was performed on these data (see Table 4). As expected, imagery was found to be highly significant ($F = 33{\cdot}59$, d.f. = 1, 217, $P < 0{\cdot}001$) and word frequency to be insignificant ($F = 0{\cdot}04$, d.f. = 1, 217). A main effect was found, however, for verb stimuli ($F = 5{\cdot}24$, d.f. = 28, 217, $P < 0{\cdot}001$), as was an interaction between imagery and word frequency ($F = 8{\cdot}52$, d.f. = 1, 217, $P < 0{\cdot}001$). A profile of the mean treatments is presented in Fig. 3. Contrary to the pattern of results in Expt 1, these means suggest that frequency has an effect but that this is confined to high imagery words.

Table 3. Three-way ANOVA of the number of categories used to judge the drawings elicited for the verbs as a function of verb imagery, frequency and verb stimuli

Source	SS	d.f.	MS	F
Imagery	30·94	1	30·94	18·87**
Frequency	1·41	1	1·41	0·86
Stimuli	105·42	28	3·77	2·30**
Imagery × frequency	3·29	1	3·29	2·00
Subjects	160·53	7	22·93	
Residual	355·85	217	1·64	

** $P < 0{\cdot}001$.

Table 4. Three-way ANOVA of the population stereotype categories (W_{max}) obtained in Expt 2

Source	SS	d.f.	MS	F
Imagery	6013·23	1	6013·23	33·59**
Frequency	6·35	1	6·35	0·04
Stimuli	26287·54	28	938·84	5·24**
Imagery × frequency	1525·10	1	1525·10	8·52**
Subjects	13581·16	7	1940·17	
Residual	388947·25	217	179·01	

** $P < 0{\cdot}001$.

A *post hoc* analysis, using the Tukey test, however, revealed that there was in fact no significant difference between the treatment means ($\bar{d}_T = 20{\cdot}817$, $P < 0{\cdot}01$). The factor causing this interaction, therefore, was probably due to the substantial ranges of W_{max} within each treatment. In particular, the high W_{max} value of 75·54 per cent for the verb 'to start' (a low imagery verb) and the low W_{max} values of 40·22 and 35·43 per cent for 'to freeze' and 'to draw', respectively (both high imagery verbs) suggest that these individual stimuli may have biased the results. Using the same critical Tukey score, a pairwise comparison of the individual means within each treatment showed that they were, in fact, significantly different from two or more of the other means within their respective treatment. The results from these analyses again provide strong evidence to suggest that imagery is an important factor in determining the extent to which verbs can be meaningfully represented in a pictorial form.

A significant negative correlation between imagery and W_{max} ($r = -0{\cdot}61$, $P < 0{\cdot}001$), indicating that stereotype strength tends to increase as the verb becomes more concrete, supports the above interpretation. To determine whether the population stereotype (W_{max}) was, in fact, the most appropriate representation, the relationship between W_{max} and W_i (the category of drawings which each judge considered as most representative) was examined. Interestingly, in only just over half (58 per cent) of all judgements were the categories with the greatest number of pictures judged as the best representation. Of these, 64 per cent were for high imagery verbs and 51 per cent for low imagery verbs. Contrary to normal expectations about population stereotypes, therefore, this suggests that the judges' decision in ranking the categories of ideas according to which best represented the meaning of the particular verb was not necessarily influenced simply by the number of drawings they had assigned to each category. A further inspection of the data also revealed that, for a few of the judgements, the category which was considered as best representing the verb contained only one (4·6 per cent of the total) or two (6·6 per cent of the total) drawings.

Finally, the stereotype strength and type of verbal descriptions given for the ideas being expressed were also analysed. The word length of the verbal labels varied between subjects, although there was consistency within subjects. Some subjects tended to give very terse labels for their ideas whilst others were more specific. For example the description 'weighty object knocking something down' for the verb 'to demolish' was a typical label given by one judge, whereas another simply gave one-word labels such as 'destroy'. On the other hand, the content of the verbal labels for the various ideas being expressed for each concept was often quite similar. Nearly 75 per cent of the W_i categories in the high imagery set and 50 per cent in the low imagery set were given equivalent verbal labels by three or more judges. This would seem to suggest that there is a general consensus among the judges to express the same ideas as the most meaningful, particularly in the high imagery set.

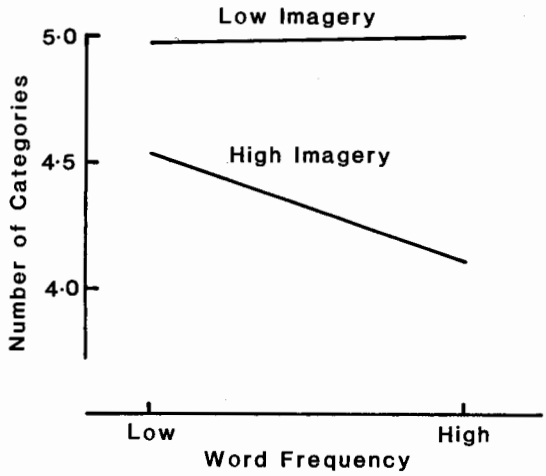

Figure 3. Mean number of categories selected across subjects as a function of imagery and word frequency.

Discussion

The results from this experiment clearly illustrate that a verb's associated imagery rating has an effect on how comprehensive and meaningful will be its pictorial representation. In all analyses performed on the data, performance was always superior with the high imagery verbs. In particular, they were judged as producing the majority of most appropriate drawings and the least number of irrelevant drawings. In addition, significantly fewer categories were used by judges to sort the high imagery drawings into piles of similar ideas. Taken together, these results suggest that the ideas being expressed were more universal for high imagery verbs. On the other hand, for low imagery verbs the ideas being expressed were more diverse.

The stereotype strengths of high imagery verbs and the number of times that they were judged to be the most representative were also found to be greater compared with low imagery verbs. This implies, therefore, that the relation between the frequency of producing a particular representation and its relative comprehensibility is far more evident for high imagery verbs. Interestingly, though, there were a few instances in all sets of verbs where the category selected as representing the best idea contained only one or two examples. This suggests, that, irrespective of imagery associations, the best idea was sometimes judged as being unique or more idiosyncratic than the one which the general consensus of drawings represented. Jones (1983) also noted this, but much more markedly, in her study (19·1 per cent of total judgements selected a category which had only one example as representing the best idea). She suggested that this may occur because one creative person might be able to produce a drawing which is able to encompass the essential meaning of certain types of abstract concept more efficiently than that which most people would produce. Alternatively, a lack of drawing skills may prevent people from producing the drawings that they would prefer (Szlichcinski, 1980*b*). It may be, therefore, that people share similar ideas as to how such concepts are best represented in a pictorial form but are unable to depict them effectively. These arguments, therefore, are not contrary to the general rationale of population stereotyping. Furthermore, since only a few examples of this effect occurred in this experiment, producing little agreement amongst judges, it would seem that the population stereotype is still the most appropriate measure of the meaningfulness of these types of concepts.

The types of drawings considered as being the most representative for both imagery and frequency sets were, interestingly, the ones which were highly pictorial (in the sense of

depicting a concrete object or person), often with a high degree of complexity. For example, the most representative drawing for the verb 'to cut' was of a pair of scissors alongside two pieces of jagged paper (see Fig. 2*h*) rather than simply by analogy of the cutting implement. Furthermore, they generally included supplements and universal symbols, i.e. lines, arrows, question marks, think balloons, etc. These findings seem to suggest, as earlier hypothesized, that the most comprehensive and meaningful visual representations of verbs appear to be those which attempt to portray ongoing motion or action that is about to or has been carried out, by use of concrete objects associated with the verb plus conventionally accepted abstract symbols.

General discussion and conclusions

One of the main findings from these two experiments was that meaningful drawings could be produced for a range of abstract verbs and that the majority of them were considered to be highly pictorial. Moreover, the results showed that, irrespective of imagery or frequency association, there was a strong tendency for subjects to concretize the concepts in some way. Similar findings reported by Jones (1983) for abstract nouns and Howell & Fuchs (1968) for abstract military concepts suggests, therefore, that the process of concretization is necessary for evoking images for all classes of abstract concepts.

The finding that imagery and, to a lesser extent, frequency systematically affect the ease with which these images are produced and their relative comprehensibility, however, indicates that the actual process is somewhat complex. For instance, whilst it appears to be a relatively direct process for high imagery verbs where common associations often seem to exist, this is not the case for low imagery verbs. Instead, images elicited for these types of verbs appear to be less forthcoming and depend much more on personal and idiosyncratic associations.

Another factor which appears to play an important role in the concretization process is the extent to which common perceptual experiences can be visually represented. For example, many of the drawings that were produced frequently portrayed one or more human figures performing an action. This was especially the case for high imagery verbs which had a large active component, e.g. 'to run'. Interestingly, a similar type of strategy was frequently adopted for the more passive types of low imagery verbs, e.g. 'to scan', but which also included objects associated with the action. This enabled the image to be more specific and in doing so helped to reduce the number of possible meanings that could be inferred.

An implication of having to use associated objects is that such a strategy might restrict the generalization of this class of abstract data to the context of the computer interface. If the most stereotypic images associated with abstract verbs are of human beings, it might prove difficult to develop a set of icons which are sufficiently distinct from one another and which can differentiate the various operations for a particular application. On the other hand, it might actually be possible to use effectively the various consistent features associated with the human figure to represent the relations among a set of commands. A common theme could be developed which would provide the potential for a whole set of icons to be created that are able to share similar features. A similar idea has been suggested by Houston (1983) who gives an example in which food processing is used as an analogy for information processing. As part of this analogy, Houston suggests that each of the rings on a cooker offers 'a delicious way of cooking one's data' – such as word processing or statistical packages.

To determine whether visual imagery can be effectively mapped on to a set of command operations for a specific application, it would be necessary also to carry out the task in a computer context where the verbs are specified as commands and the subjects are familiar

with the concepts as command verbs. It would also be interesting to see whether the images produced in this context were as concrete and complex as many of the drawings judged to be the most meaningful in this study.

It may be that the essential meaning of these types of abstract concepts in either a free or specified context can only be represented in a complex and highly pictorial form. Moreover, such representations may actually be a less efficient means of communication for human computer interaction than a verbal representation. On the other hand, much of the information represented in this type of pictorial form may be redundant. This suggests that there could be a trade-off between complexity and compactness. In their research on military signs, Howell & Fuchs (1968) proposed that symbols could be greatly reduced in complexity without affecting their specific meaningfulness. Some kind of 'weeding out' of irrelevant features may therefore be possible. This would have to be achieved, however, without affecting the global meaning of the representation. In addition, the possiblity of increased demands made on cognitive resources would need to be considered.

Another prevalent finding from the two experiments was the widespread use of supplements and universals, whose function appeared to be as indicators of ongoing action. In particular, the arrow and pseudo-naturalistic streaking lines indicating speed or movement were common features. The history of these conventions, which can be found in any comic strip, remains largely unexplored (Gombrich, 1982). Perhaps one way of reducing the complexity/compactness problem would be to investigate further how these symbols could be more fully exploited in iconic displays, since they are generally very simplistic and universally understood. It would be interesting also to see whether these types of symbols, in comparison with the previously mentioned human figure analogy, would be an effective form of communication for human–computer interaction. In particular, it would be interesting to determine whether they would be efficient if they were used exclusively rather than in combination with other representations.

In summary, the findings from these two preliminary experiments suggest that visual representations can be generated for verb stimuli using the sign production method. Furthermore, the most meaningful and comprehensive representations were those which depicted concrete objects in conjunction with universal abstract symbols and supplements. The types of verbs which were found to be most difficult to represent visually and subsequently to communicate their intended meaning were those associated with low imagery. Perhaps these types of command verbs may be more appropriately left in a verbal form. On the other hand, command verbs associated with high imagery appear to have a much greater potential for being represented in iconic form at the computer interface.

References

Collins, B. & Lerner, N. (1982). Assessment of fire safety symbols. *Human Factors*, **24**, 25–84.

Dewar, R. E. & Swanson, H. A. (1972). Recognition of traffic control signs. *Highway Research Record*, **414**, 16–23.

Gittins, D. T., Winder, R. L. & Bez, H. E. (1984). UNICON: An icon driven end-user interface to UNIX. *International Journal of Man–Machine Studies*, **21**, 451–461.

Gombrich, E. H. (1982). *The Image and the Eye: Further Studies in the Psychology of Representation*. Oxford: Phaidon.

Green, P. & Pew, R. (1978). Evaluating pictographic symbols: An automative application. *Human Factors*, **20**, 103–114.

Haber, R. N. (1970). How we remember what we see. *Scientific American*, **222**, 104–115.

Haber, R. N., Standing, L. & Conezio, J. (1970). Perception and memory for pictures: Single trial learning of 2500 visual stimuli. *Psychonomic Science*, **19**, 73–74.

Hemenway, K. (1982). Psychological issues in the use of icons in command menus. In *Proceedings on Human Factors in Computer Systems*. Gaithersburg, MD: ACM.

Houston, T. (1983). The allegory of software. *BYTE*, **8**(12), 210–214.

Howell, W. C. & Fuchs, A. F. (1968). Population stereotypy in code design. *Organizational Behaviour and Human Performance*, **3**, 310–339.

Jones, S. (1983). Stereotypy in pictograms of abstract concepts. *Ergonomics*, **26**, 605–611.
Karsh, R. & Mudd, S. A. (1962). Design of a picture language to identify vehicle controls. III: A comparative evaluation of selected picture-symbol designs. Aberdeen Proving Ground, MD, US Army Human Engineering Laboratory, Technical Memorandum 15–62, August.
Paivio, A. (1971). *Imagery and Verbal Processes*. Eastbourne: Holt, Rinehart & Winston.
Rogers, Y. & Oborne, D. (1985). Some psychological attributes of potential computer command names. *Behaviour and Information Technology*, **4**(4), 349–365.
Rohr, G. & Keppel, E. (1984). Iconic interfaces: Where to use and how to construct? In H. W. Hendrick & O. Brown (eds), *Human Factors in Organizational Design and Management*.
Smith, D. C., Irby, C., Kimball, R. & Verplank, B. (1982). Designing the Star User interface. *BYTE*, **7**(4), 242–282.
Szlichcinski, K. P. (1980*a*). The production and comprehension of pictorial instructions. Unpublished PhD thesis, University of London.
Szlichcinski, K. P. (1980*b*). The syntax of pictorial instructions. In P. A. Kolers, M. E. Wrolstad & H. Bouma (eds), *Processing of Visible Language*. New York: Plenum.
Thorndyke, E. L. & Lorge, I. (1944). *The Teacher's Word Book of 30,000 Words*. New York: Teachers College, Columbia University.
Walker, R. E., Nicolay, R. C. & Stearns, C. R. (1965). Comparative accuracy of recognizing American and international road signs. *Journal of Applied Psychology*, **49**, 322–325.
Williams, G. (1983). The LISA computer system. *BYTE*, **8**(2), 33–50.

Received 23 October 1985; revised version received 24 June 1986

Requests for reprints should be addressed to Yvonne Rogers, Department of Psychology, University College of Swansea, Singleton Park, Swansea, West Glamorgan, UK.
D. J. Oborne is also at the above address.