Visualization and model formulation: An analysis of the sketches of expert modellers

Article *in* Journal of the Operational Research Society · March 2008

DOI: 10.1057/palgrave.jors.2602331

3 authors, including:

Laurie Waisel
Independent Researcher

William Alan Wallace
Rensselaer Polytechnic Institute

Journal of the Operational Research Society (2008) 59, 353–361

www.palgrave-journals.com/jors

Visualization and model formulation: an analysis of the sketches of expert modellers

LB Waisel[1], WA Wallace[2] and TR Willemain[2]*

[1] *Concurrent Technologies, Johnstown, PA, USA; and* [2] *Rensselaer Polytechnic Institute, Troy, NY, USA*

We report on an exploratory study of the use of sketches by experts engaged in the formulation of operational research models. We outline a methodology to identify sketches, link them to text in an associated verbal protocol, classify them, and relate their use to the modelling process. We found that use of sketches is part of the personal modelling style of experts, that sketches are used more when the focus is on model structure or realization than on model context or assessment, that sentential sketches tend to be started sooner than diagrammatic sketches, and that sketching generally begins earlier on more difficult problems. We expect that such descriptive studies of expert behaviour will be useful in developing a science of modelling and in teaching novice modellers.

Journal of the Operational Research Society (2008) **59,** 353–361. doi:10.1057/palgrave.jors.2602331
Published online 22 November 2006

Keywords: OR practice; OR education

Introduction

This paper is the third in a sequence analysing think-aloud protocols to provide an empirical description of the thought process involved in the first hour in the life on a model (Willemain, 1994, 1995). The ultimate objectives of this research are to better understand the process of modelling and move it more towards a science. Eventually, this line of research may lead to the creation of better models by improving the modelling process through the effective incorporation of visualization. Our immediate objectives are to develop a methodology for quantitative analysis of modellers' sketches and make some first observations about the use of sketches in model formulation. We believe that all science is built on a base of quantitative description, and that this study is a step in the direction of a science of modelling.

Visualization and modelling

We define visualization as the use of non-mental images to enhance insight. Until the last few decades, visualizations were limited to drawings on paper. Jones, in a review of the history of the use of visualization to support mathematical modelling, quoted Descartes as having said, 'Imagination or visualization and in particular the use of diagrams, has a crucial part to play in scientific investigation' (Jones, 1996). With the advent of computers, electronic graphics began to be used as visualizations, and the last few decades have witnessed an explosion in the power, sophistication, and variety of computer-generated visualizations. Here, however, we focus on simple hand sketches.

How might the use of visualization impact the process of modelling? We hypothesize that visualization of the process being modelled leads to improved performance in the building and use of the model. This could occur through either enhanced creativity or increased understanding of the model.

We view modelling as a problem-solving activity (MacCrimmon and Taylor, 1976; Smith, 1988) and agree with researchers who say that problem solving is a creative activity (Newell *et al*, 1962; Evans, 1991); therefore, we assume that conclusions about creativity also apply to modelling. A number of researchers have presented anecdotal evidence about the role of imagery in the creative process (Einstein, 1949; Shepard, 1978; Tweney *et al*, 1981; Nersessian, 1992). Scientists, artists, musicians, and architects have claimed that images are crucial to their creative work. Einstein, Maxwell, Kekule, and Watson all reported images, both waking and dreaming, as the catalysts for their pivotal scientific insights (Shepard, 1978).

In addition to the link between mental imagery and creativity, there also is evidence that visualization enhances creativity (Shepard, 1978; Finke and Slayton, 1988; Anderson and Helstrup, 1993) including the suggestion that as the imagery task becomes more complex, there will be increasing need for external support (Anderson and Helstrup, 1993).

Finally, both anecdotal and experimental evidence exist to support the hypothesis that visualization enhances comprehension. It is self-evident that visualization is an effective comprehension aid, given the widespread and long-standing use of visualization as an explanatory device in science and

**Correspondence: TR Willemain, Department of Decision Sciences and Engineering Systems, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, NY 12180, USA.*
E-mail: willet@rpi.edu

mathematics. For example, semantic networks have been used in teaching linear programming modelling (Evans and Camm, 1990). Empirical evidence also indicates the effectiveness of visualization in aiding comprehension (Gange and Lipton, 1984; Turo and Johnson, 1992; Rook and Donnell, 1993; Alabastro *et al*, 1995; Massey and Wallace, 1996). Willemain's survey of expert modellers (Willemain, 1994) showed a strong consensus about the value of sketching in modelling. The following section describes how we pursued that point by analysing sketches made by expert modellers.

Sketches of expert modellers

Modeller exercises

We began our research into the role of visualization in modelling by studying the simplest form of visualization, and one that has been widely used for centuries: hand sketches made by modellers as they work through a modelling problem. Willemain (1994, 1995) used think-aloud protocols to try to better understand the process of problem solving that goes into the activity of mathematical modelling. Twelve expert modellers worked on four model formulation exercises while their verbal descriptions of their problem-solving processes were recorded and later transcribed. Four core modellers worked on all four exercises, and the remaining eight modellers each worked on one of the four exercises. The exercises presented abstract problems related to graduate admissions, alumni donations, television commercials, and global ecology.

The design of these problems was intended to run the gamut for complexity of the system to be modelled (the problem's context) and for clarity of goals (the problem's purpose) (Jackson and Keys, 1984). The Commercials exercise has a simple system and clear goals and in some respects may be considered the easiest problem. The Trees exercise has a complex system and vague goals, and perhaps may be considered the hardest problem. The Admissions exercise has a simple system and vague goals, while the Alumni exercise has a complex system and clear goals.

As each modeller worked on each exercise, he created a protocol by thinking out loud while formulating his model. These protocols were audio-taped and transcribed. Willemain and his colleagues then divided the transcripts into chunks or episodes, and coded each episode as belonging to one of five modelling topics plus an 'other' category. The modelling topics, which correspond to the generally accepted steps of the modelling process, include Context, Structure, Realization, Assessment, and Implementation. Context, which corresponds to problem formulation, addresses starting facts, assumptions, and information about the client and the client's objectives. Structure, which corresponds to constructing a mathematical model, involves the variables, relationships among the variables, the kind of model to construct, and how data should be analysed. Realization, which corresponds to deriving a solution, includes making the model yield results, deciding how to solve equations, and estimating parameters. Assessment, which corresponds to testing the model, involves evaluation of the model's correctness, feasibility, acceptability, and value to the client. Implementation addresses how to use the model to help the client and how the client can use the model.

Coding the sketches

The modellers had pencil and paper while working the exercises, and most made notes, graphs, sketches, outlines, and diagrams. We refer to this graphical output as 'sketches,' even though some of it is in words. We use the word 'sketch' in the sense of the dictionary definition of 'a rough drawing representing the chief features of an object or scene, often made as a preliminary study' (Merriam-Webster, 1980). While the words spoken by the modellers have been analysed previously (Willemain, 1995), as have their perspectives on modelling (Willemain, 1994), this is the first analysis of their sketches.

Before the sketches could be analysed, they had to be coded. The first step in coding the sketches was to determine the lines of protocol text to which each sketch corresponded. Sketch text (the text associated with a sketch) was not necessarily a continuous chunk of text: sometimes a modeller interrupted a sketch with another idea. Finding the general location of each sketch in the text turned out to be easier than anticipated. There were almost always verbatim clues, that is, one or more words in the sketch that corresponded verbatim to the words uttered by the modeller. In other words, the modellers tended to talk about their sketches with enough specificity that there usually was little to no question about where in the text of the verbal protocol a given sketch was drawn. It was sometimes a little more difficult to determine the precise starting and ending points of the text associated with a given sketch. There also was the question of whether all text related to a given sketch should be considered associated with it, or whether only the text that occurred while the sketch was being drawn should be associated. An example will make this clearer.

The sketches were coded by asking the following questions about each page of sketching:

1. How should this page be divided up into discrete sketches?
2. Which lines in the verbal protocol text should be associated with each sketch? (The protocols were transcribed and printed out in line-numbered documents.)
3. Is the sketch primarily sentential (ie, in a linear natural language formation such as words or equations) or diagrammatic (ie, location of marks in the two-dimensional plane is essential to the sketch's meaning)?

For convenience, each sketch also was given a brief descriptive name, such as 'AHP' or 'Cost versus value' or 'Regionalize'.

A number of heuristics were used to answer these questions; the best way to describe them is through some coding examples. Consider the coding of the first page of sketches drawn by Modeller B on the Admissions problem (see Figure 1).

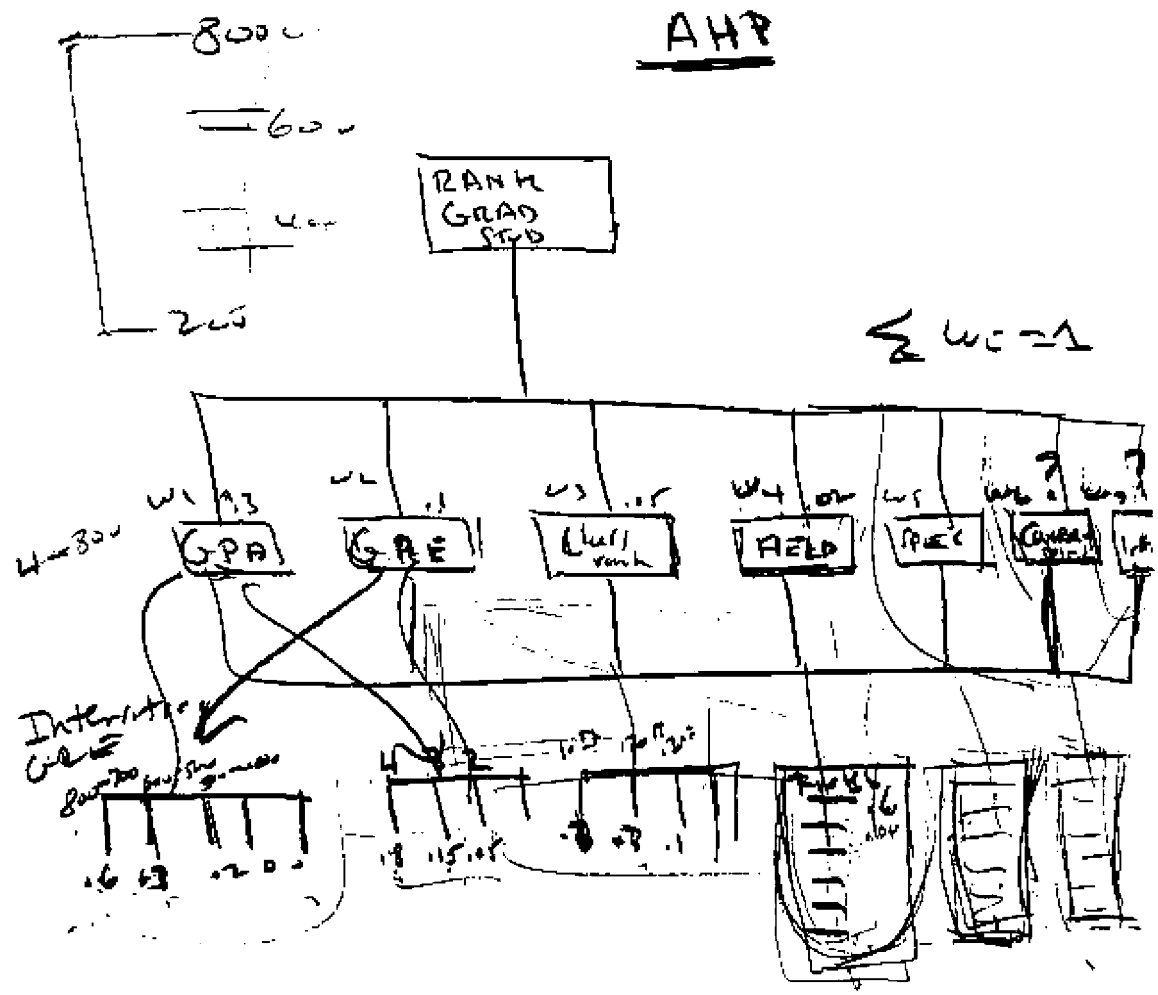

Figure 1 'AHP' sketch by modeller B working the Admissions exercise.

The first question is how many sketches are present. Physical continuity was an important determinant of where a sketch began and ended. Using this criterion, it appears that there are no more than two sketches on this page. One potential sketch is the large chart, and the other is the sketchy drawing in the upper left in which the numbers 200 and 800 are connected by a line. Are both of these drawings part of the same overall sketch, or are they separate? In addition to physical continuity or discontinuity, the other important criterion for division into separate sketches is that the potential sketches express different ideas, as opposed to being simultaneous illustrations of the same idea. Accordingly, Figure 1, which contains two discontinuous sketches that express different ideas, was found to consist of two sketches. The first one, the large flow chart, spanned lines 66–145 and then skipped over to 157 and finished on line 240. The second sketch, the small sketch in the upper left, took up the intermediate space, lines 146–156. How were these determinations made? On line 66 was a bracketed notation from the experimenter that at this point in the protocol, the modeller began writing on page one (of the sketches for that modeller and exercise). This notation, along with the text of line 66, which states, 'Well, look-, looking at the A-, AHP approach here', clearly signals the beginning of the chart labeled 'AHP' in the sketch. Line 241 contains the bracketed notation that the modeller began to write page two, thus signalling the end of the sketch on page one. The starting and ending lines of the chart sketch have now been clearly established. Next, how do we know that lines 146–156 belong to the small sketch in the upper left? Let's examine the transcript:

Line 146: *And so here's, sorta, maybe some judgmental call, or you can even do this by*
Line 147: *pairwise comparisons, OK? And so, if someone has an 8-, if someone has an 800,*
Line 148: *I would go ahead and, ah, and again this is, this is done on a normalized scale*
Line 149: *to 1. If, if someone has an 8-, 800, I would, I, I might want to, you know -*

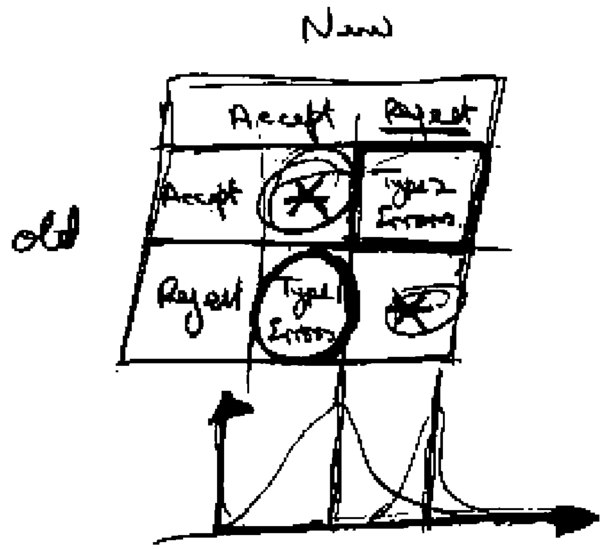

Figure 2 'Evaluation model' sketch by modeller D working the Admissions exercise.

Line 150: *I'll just, again, make up num-, numbers, ah, here. We're gonna go from 200 to*
Line 151: *800, and so we have to do a little research on this one, OK? So if I have 200*
Line 152: *to 800, OK? And so I might want to, I've, I've, I've got 600 points to divide*
Line 153: *up. But if someone is an 800, he might be much, much more valuable to me than*
Line 154: *a 600, so just because there's, there's 200 points difference here and, and 600*
Line 155: *to 400, the, eh, someone at 400 might be, no-, nobody I really, really want.*
Line 156: *So I gotta worry what weight I want to give, up over here.*

The lines immediately preceding this chunk of text were about the section of the flow chart labelled 'GRE' and the need to 'divide the [GRE] structure from 200 to 800 into a weight itself'. Lines 146–156 describe how this could be done, and the sketch illustrates the words precisely. This example was easy, since both its discontinuity and its easily separable ideas pointed to declaring Figure 1 as two sketches.

Figure 2 (also from the Admissions exercise) contains an example of a sketch whose meaning criterion is in conflict with the physical continuity feature. At first glance, it appears that the chart and the graph may be two separate sketches; although the table and the graph are very close together, they do not quite touch. Close examination of both, however, makes it clear that the graph and the table present the same ideas. In the chart, the Type I errors (third row, second column) are the students who were accepted under the new admissions system but rejected under the old. The Type II errors (second row, third column) are the students who were accepted under the old system but rejected under the new. In this context, the graph below the chart is instantly recognizable as a graph of the distribution of Type I and Type II errors. How does this example differ from the previous one? First, the two sketches express the same idea, rather than one sketch elaborating on something in the other. Second, there is no separate text that addresses the graph alone. If the graph represented a thought process distinct from the chart, the modeller would have said something specifically about it. The graph is subsumed in the discussion about the Type I and Type II error, however, and therefore the chart and the graph are considered parts of a single sketch.

The terms 'sentential' and 'diagrammatic' were derived from Larkin and Simon's seminal paper (Larkin and Simon, 1987), although we do not use the terms exactly as they do. A sketch was coded 'sentential' if its meaning is conveyed primarily by words or equations and thus could be translated to ASCII code without significant loss of meaning. A sketch was coded 'diagrammatic' if its meaning is conveyed primarily by a drawing.

'First shot', seen in Figure 3, is an example of a sentential sketch. The meaning behind this sketch is that the modelling process should begin with a 'first shot' that takes a quick look at the elements listed in the sketch in the order specified. The essence of the sketch's meaning (as determined by taking the sketch in the context of the verbal protocol) is given by the words in the sketch. This sketch has diagrammatic elements, such as the arrows and the small chart in the upper right, but those elements are secondary in importance; thus, this is classified as a sentential sketch.

Examples of diagrammatic sketches include 'Graph 1' by modeller C working the Commercials exercise (see Figure 4) and 'Rank' by modeller B, also doing the Commercials exercise (see Figure 5). 'Rank' illustrates the idea of ranking the proposals for commercials according to a subjective measure of quality and then allocating portions of the budget to produce each of the commercials. Although the meaning of this sketch could be expressed in ASCII, the actual sketch conveys the primacy of the ranking idea, which is central to this sketch. In 'Graph 1', the focus of the sketch, as the name suggests, is the graph. The textual annotations add meaning but clearly are secondary in importance.

A subtle point: Is the purpose of sketching to clarify ideas and stimulate creativity for oneself or to communicate ideas to others? Since the modellers were told to focus on their own thoughts and process of problem-solving, the assumption is that in this setting, the primary purpose of sketching was to assist the modeller's problem-solving activities. Sometimes, however, the experimenter's presence was an excuse for the modellers to explain to him and also to themselves what they were doing. (This artefact illustrates a limitation of examining modeller behaviour *in vitro* rather than *in vivo*).

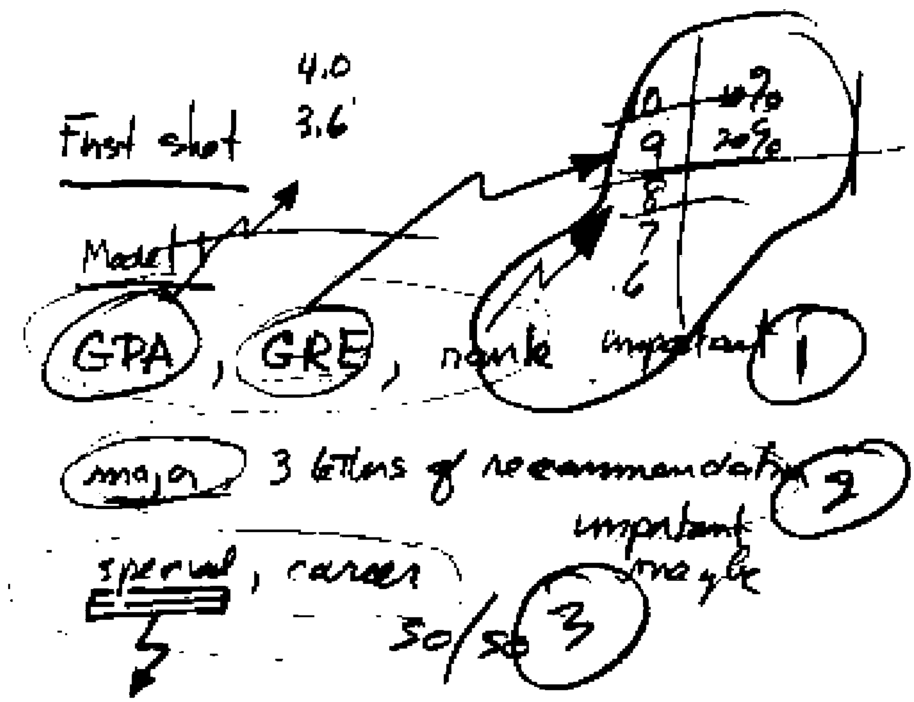

Figure 3 ‘First shot’ sketch by modeller E working the Admissions exercise.

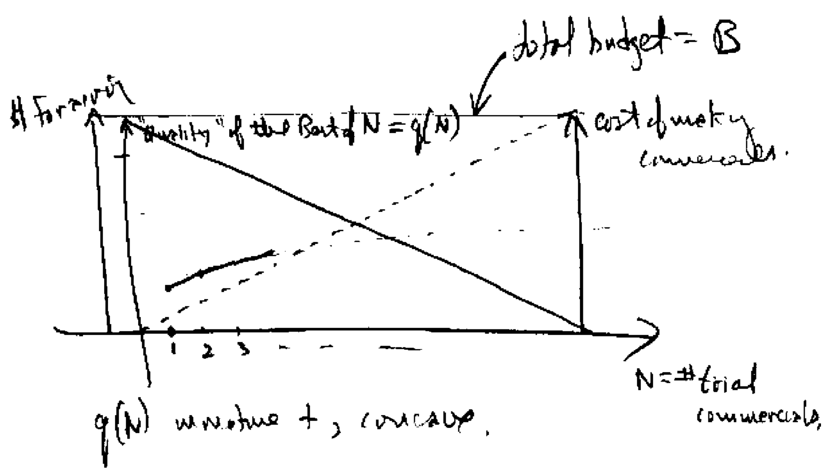

Figure 4 ‘Graph 1’ sketch by modeller C working the Commercials exercise.

Analysis and results

The results of the sketch coding were placed in a series of spreadsheets. Each line of each spreadsheet corresponded to a line from the verbal protocol analysis. The lines that did not correspond to any sketches contained only the line number and a topic designation (context, structure, assessment, realization, implementation, or other). Each text line that was associated with a sketch contained the sketch’s name, its designation as sentential or diagrammatic, and the ranges of text lines that were associated with that sketch. Each line of text was associated with at most one sketch. A separate spreadsheet was created for each problem-solving session (eg, modeller A solving the Admissions problem, modeller A solving the Alumni problem, and so on). These data were used to create summary statistics for each modeller and exercise, which were then used as the basis for analysis. Only the four core modellers (A–D) were used in the analysis, since those were the modellers who worked on all four exercises. The analysis focused on discovering (1) what relationship, if any, the sketches had with the modelling topics and (2) what the sketches could reveal about the modelling process and the role of visualization.

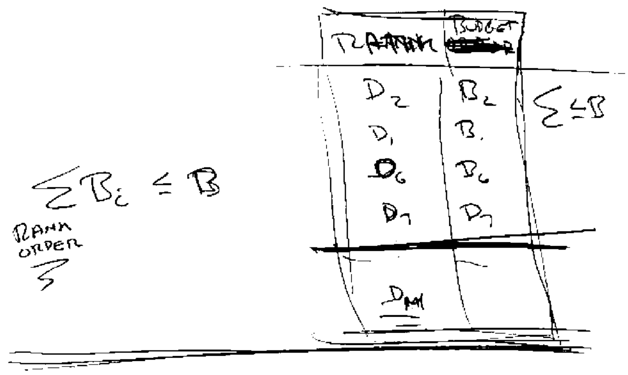

Figure 5 'Rank' sketch by modeller B working the Commercials exercise.

Sketch proportion

A three-way fixed-effects analysis of variance (ANOVA) model was used to investigate the relationship between the modelling topics and the sketches. For this analysis, the data for the topic categories Implementation and Other were dropped, since those categories comprised a very small proportion of the data and were least relevant to the model formulation process. After dropping the extra modellers and topic categories, data on all four core modellers doing all four exercises were used to perform the ANOVA. The independent variables were modeller, exercise, and topic.

The logit of the started sketch proportion was used as the dependent variable. 'Started count' proportions were used instead of raw proportions to avoid having zeroes in the table (Mosteller and Tukey, 1977). The logit transformation was used to make the data distribution more like the normal in order to improve the validity of statistical inference. A high sketch proportion means that a large amount of the protocol text is associated with sketches. The sketch proportion is equal to the number of lines of text in all sketches associated with topic X divided by the total number of lines of text coded as topic X, for that modeller working that exercise. The calculation for started counts is $p = (\#\text{ sketch lines} + 1/6)/(\#\text{ text lines} + 1/3)$. So, for example, modeller A doing the Trees exercise had a total of 137 lines of text that were associated with sketches and also were coded as Context, Structure, Assessment, or Realization. Of those 137 lines of sketch text, 54 were coded as Structure. The total number of lines of text coded as Structure for modeller A doing Trees was 316. So, the started sketch proportion for A doing Trees is $(54 + 1/6)/(316 + 1/3) = 0.17089$. The SAS procedure GLM was used to carry out the ANOVA.

Univariate statistics for sketch proportion data by topic and by exercise are shown in Table 1. The differences in means are clearly visible. ANOVA results are shown in Table 2. Normal probability plots of residuals and scatterplots of residuals *versus* fitted values confirmed the assumptions underlying ANOVA inference. Significant effects were found for modeller and topic. The significant effect for modeller indicates individual differences in the amount of sketching that experts do when modelling. The fact that there was no significant effect for exercise suggests that the modellers' behaviour was not problem-specific but rather a matter of personal style.

Table 1 Summary statistics for sketch proportion data

	Mean	*Standard deviation*	*Standard error of the mean*
By topic			
C = Context	0.20	0.25	0.06
A = Assessment	0.28	0.23	0.06
S = Structure	0.45	0.26	0.06
R = Realization	0.49	0.34	0.08
By exercise			
Admissions	0.33	0.29	0.07
Alumni	0.36	0.30	0.08
Commercials	0.35	0.30	0.07
Trees	0.39	0.31	0.08

A test for multiple comparisons of means showed that there were significant differences between the topics Context and Realization, with more sketching associated with Realization. Why? Perhaps because building and solving a model are the specific tasks at hand, whereas learning about problem context and assessing the model as it is built are meta-modelling tasks.

Sketch starting times

It seemed that an analysis of the starting times of the sketches might shed some light on what role the sketches played in the model formulation process. For example, it might be the case that diagrammatic sketches tended to start earlier than sentential sketches, or *vice versa*. A normalized starting time for each sketch was calculated by dividing the protocol line number at which the sketch started by the total number of lines in the protocol. Since all of the protocol sessions lasted for 1 h, a starting time of 0.5 should mean that the sketch was begun about 30 min into the session. Table 3 shows the univariate statistics for start times.

A three-way fixed-effects ANOVA was used to investigate the effects of modeller, sketch type, and exercise on the logits of the starting times. The logit transformation was again used to improve the normality of the data; analysis of residuals showed that this transformation succeeded in justifying the usual ANOVA inference procedures. Table 4 shows the results of the ANOVA. Both exercise and sketch type were found to have significant effects on the sketch start times. The fact that there was no significant modeller effect implies that the observed behaviour is general across modellers. A comparison of means by sketch type showed that the sentential sketches started significantly earlier than the diagrammatic sketches.

Why did sentential sketches start significantly earlier than diagrammatic sketches? These results suggest that expert modellers tend to start out solving a problem by expressing it in words and equations, and only later, after the initial

Table 2 ANOVA results for logit of sketch proportion data

General linear models procedure
Dependent variable: Started logit of sketch proportion

Source	*d.f.*	*Sum of squares*	*Mean square*	*F value*	*Pr > F*
Model	9	210.26157270	23.36239697	7.46	0.0001
Error	54	169.07309612	3.13098326		
Corrected total	63	379.33466881			
	R-square	*C.V.*	*Root MSE*		*SKPROP mean*
	0.554290	−157.6110	1.7694585		−1.1226748
Source	*d.f.*	*Type I SS*	*Mean square*	*F value*	*Pr > F*
Topic	3	35.93716337	11.97905446	3.83	0.0148
Exercise	3	14.41420297	4.80473432	1.53	0.2160
Modeller	3	159.91020635	53.30340212	17.02	0.0001
Source	*d.f.*	*Type III SS*	*Mean square*	*F value*	*Pr > F*
Topic	3	35.93716337	11.97905446	3.83	0.0148
Exercise	3	14.41420297	4.80473432	1.53	0.2160
Modeller	3	159.91020635	53.30340212	17.02	0.0001

representations have been established, do they turn to diagrams. This result may be related to the fact that the problems were given in words. Rook and Donnell (1993) found that experimental subjects gave their best performance when a graphical interface was paired with a graphic mental model of the problem and a textual interface was paired with a textual mental model of the problem. This may indicate a preference for working in the same mode in which the problem is presented.

The data for start times by exercise show that sketching for the Trees exercise started earlier than Admissions, which started earlier than Alumni, which started earlier than Commercials. This pattern bears an interesting relation to the characteristics of the exercises. Recall that Trees was the exercise with a complex system and vague goals, and that the Commercials exercise had a simple system and clear goals. Thus, Trees could be considered the hardest exercise and Commercials the easiest. Alumni (clear goals, complex system) and Admissions (vague goals, simple system) would fall somewhere in the middle.

If we believe that sketching aids the modelling process, then it seems reasonable that the starting times of sketches may be related to the difficulty of the modelling task. In other words, the harder the problem, the earlier the sketching begins. The results support this hypothesis, since sketches started earliest for Trees, the hardest problem, and latest for Commercials, the easiest problem, with the sketch start times for the two problems of intermediate difficulty (Admissions and Alumni) falling in between.

So, the results of this analysis show that the amount of sketching being done depends on individual differences and the current modelling topic. Some modellers sketch more than others and more sketching takes when the focus is on structure and realization than on context and assessment. The results also demonstrate that the timing of sketches depends on the problem and problem type and also on the type of sketch being done. Sketching starts earlier for more difficult problems than for easier ones, and sentential sketches are begun earlier than diagrammatic sketches.

Table 3 Summary statistics for start time data

	Mean	*Standard deviation*	*Standard error of the mean*
By type			
Diagrammatic	0.51	0.26	0.03
Sentential	0.37	0.25	0.03
By exercise			
Admissions	0.37	0.23	0.05
Alumni	0.40	0.26	0.05
Commercials	0.56	0.27	0.04
Trees	0.32	0.22	0.04

Univariate statistics for start time data.

Summary and conclusions

Sketches play a major role in the formulation of models by experts. They extend the modeller's memory and thereby support development of more comprehensive models. They provide a way to sharpen mental models by making them explicit. They communicate ideas to other people involved with the model, allowing development of shared mental models. If we are to develop a science of modelling, in hopes of improving both education and practice, then it will be useful to understand how sketches help experts make good models.

Table 4 ANOVA results for logit of start time data

General linear models procedure *Dependent variable: Logit of start time*					
Source	*d.f.*	*Sum of squares*	*Mean square*	*F value*	*Pr > F*
Model	7		8.50108487	4.60	0.0001
Error	125	231.21110733	1.84968886		
Corrected total	132	290.71870141			
	R-square	*C.V.*	*Root MSE*		*SKPROP mean*
	0.204691	−295.2012	1.3600327		−1.1226748
Source	*d.f.*	*Type I SS*	*Mean square*	*F value*	*Pr > F*
Exercise	3	29.17626349	9.72542116	5.26	0.0019
Modeller	3	6.57985152	2.19328384	1.19	0.3180
Type	1	23.75147907	23.75147907	12.84	0.0005
Source	*d.f.*	*Type III SS*	*Mean square*	*F value*	*Pr > F*
Topic	3	31.66246175	10.55415392	5.71	0.0011
Exercise	3	3.80533517	1.26844506	0.69	0.5624
Modeller	1	23.75147907	23.75147907	12.84	0.0005

We conducted an exploratory study of the sketches created by four expert modellers working four problems in model formulation. We classified the sketches as sentential or diagrammatic, noted their positions within think-aloud protocols, and associated them with episodes of one of four fundamental modelling tasks or 'topics': understanding the context of the problem, structuring the model, assessing the value of the model, and realizing results from the model.

We developed quantitative measures describing the use of sketches in modelling. One such measure is the 'sketch proportion', defined as the proportion of the protocol associated with each episode that is also associated with a sketch. This measure reflects the centrality of sketches within each episode. ANOVA on suitably transformed data found that modeller and topic had statistically significant effects, but exercise did not. In other words, reliance on sketches is part of individual modelling style, sketches are more important to some aspects of modelling than others, and the salience of sketches is insensitive to the problem context. Sketches were more central to the actual building of the model, as distinct from understanding the problem and evaluating the model. It is interesting to speculate whether this finding would be reversed among British modellers, given their affinity for soft systems methodologies including use of 'rich pictures' in problem formulation (Rosenhead, 1989).

Another quantitative measure we developed is the 'normalized starting time' of a sketch, defined as the proportion of the way into the protocol where the sketch first was drawn. ANOVA on suitably transformed data found that there were no significant modeller effects, but type of sketch and exercise were highly significant. Sentential sketches appeared before diagrammatic (though this result might be influenced by the manner in which the exercises were presented to the experts). The differences across exercises corresponded to the perceived complexity of the exercise: greater complexity was associated with an earlier start to sketching. Both the lack of a modeller effect and the association with complexity point to the universal utility of sketches as an aid to model formulation.

Analysis of this set of sketches has provided a starting point for further quantitative investigation of the role of visualization in modelling. Such practical questions as how to develop a 'modelling support system' to make best use of visualization still need to be examined. This is particularly important in order to support the vast numbers of inexpert modellers now empowered by spreadsheet technology. It is our observation from many years in the classroom that US student modellers drastically underuse visualization and need to be instructed in its use as an aid to model formulation.

Perhaps the greatest value of this work is that it serves as a kind of existence proof: one can study visualization in modelling by running controlled experiments, coding the work products of the modellers, then performing statistical analyses of the results. Bringing this paradigm of scientific experimentation into the world of modelling is an essential step in the development of a modelling science. Combined with more qualitative and subjective approaches, it has the potential to improve the quality of model-based decision making.

Acknowledgements—This work was supported by US National Science Foundation Grants SES-9012094, SBS-9730465 and SES-9730465. Sarah Riddell provided helpful research assistance. Versions of this research were presented at OR42 in Swansea, at the Universities of Lancaster and Warwick, and at IFORS 2002 in Edinburgh. The third author acknowledges support provided by the University of Lancaster Department of Management Sciences during a sabbatical visit.

References

Alabastro M *et al* (1995). The use of visual modelling in designing a manufacturing process for advanced composite structures. *IEEE Trans Eng Mngt* **42**: 233–242.

Anderson RE and Helstrup T (1993). Multiple perspectives on discovery and creativity in mind and on paper. In: Roskos-Ewoldsen B, Intons-Peterson MJ, Anderson RE and Vroon PA (eds). *Imagery, Creativity, and Discovery: A Cognitive Perspective*. Elsevier Science Publishers BV: Amsterdam. pp 223–254.

Einstein A (1949). Autobiographical notes. In: Schilpp PA (ed). *Albert Einstein, Philosopher-Scientist Volume 7*. Library of Living Philosophers: Evanston: IL.

Evans JR (1991). *Creative Thinking In the Decision and Management Sciences*. SouthWestern Publishing Company: Cincinnati.

Evans JR and Camm JD (1990). Using pictorial representations in teaching linear programming modelling. *IIE Trans* **22**: 191–195.

Finke RA and Slayton K (1988). Explorations of creative visual synthesis in mental imagery. *Mem Cognition* **16**: 252–257.

Gange C and Lipton A (1984). Word-free setup instructions: Stepping into the world of complex products. *Tech Commun* **31**: 17–19.

Jackson MC and Keys P (1984). Towards a system of system methodologies. *J Opl Res Soc* **35**: 473–486.

Jones CV (1996). *Visualization and Optimization*. Kluwer Academic Publishers: Boston.

Larkin JH and Simon HA (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Sci* **11**: 65–99.

MacCrimmon KR and Taylor RN (1976). Decision making and problem solving. In: Dunnette MD (ed). *Handbook of Industrial and Organizational Psychology*. Rand McNally: Chicago. pp 1397–1453.

Massey AP and Wallace WA (1996). Understanding and facilitating group problem structuring and formulation: Mental representations, interaction, and representation aids. *Decis Support Syst: Special Issue on Model Formulation* **17**: 253–274.

Merriam-Webster. *Webster's New Collegiate Dictionary*. G&C Merriam Co.: Springfield, MA.

Mosteller F and Tukey JW (1977). *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley: Reading, MA.

Nersessian NJ (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In: Giere RN (ed). Cognitive Models of Science, Vol. XV. Minnesota Studies in the Philosophy of Science, University of Minnesota Press: Minneapolis: MN. pp 3–44.

Newell A, Shaw JC and Simon HA (1962). The processes of creative thinking. In: Gruber HE, Terrell G and Wertheimer M (eds). *Contemporary Approaches to Creative Thinking: A Symposium held at the University of Colorado*. Atherton Press: New York. pp 63–119.

Rook FW and Donnell ML (1993). Human cognition and the expert system interface: Mental models and inference explanations. *IEEE Trans Syst Man Cybernet* **23**: 1649–1661.

Rosenhead J (ed) (1989). *Rational Analysis for a Problematic World* John Wiley & Sons: Chichester, UK.

Shepard RN (1978). Externalization of mental images and the act of creation. In: Randhawa BS and Coffman WE (eds). *Visual Learning, Thinking, and Communication*. Academic Press: New York. pp 133–189.

Smith GF (1988). Towards a heuristic theory of problem structuring. *Mngt Sci* **34**: 1489–1506.

Turo D and Johnson B (1992). Improving the visualization of hierarchies with treemaps: Design issues and experimentation. *IEEE Visualization '92 Conference*, Boston, MA, pp 124-131.

Tweney RD, Doherty ME and Mynatt CR (1981). *On Scientific Thinking*. Columbia University Press: New York.

Willemain TR (1994). Insights on modelling from a dozen experts. *Opns Res* **42**: 213–222.

Willemain TR (1995). Model formulation: What experts think about and when. *Opns Res* **43**: 916–932.

Received December 2005;
accepted July 2006 after three revisions