Individual differences in proof structures following multimodal logic teaching

Individual differences in proof structures following multimodal logic teaching

Jon Oberlander and Richard Cox and Padraic Monaghan
Keith Stenning and Richard Tobin
Human Communication Research Centre
University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, Scotland
{J.Oberlander|R.Cox|P.Monaghan|K.Stenning|R.Tobin}@ed.ac.uk

Abstract

We have been studying how students respond to multimodal logic teaching with Hyperproof. Performance measures have already indicated that students' pre-existing cognitive styles have a significant impact on teaching outcome. Furthermore, a substantial corpus of proofs has been gathered via automatic logging of proof development. We report results from analyses of final proof structure, exploiting (i) 'proofograms', a novel method of proof visualisation, and (ii) corpus-linguistic bigram analysis of rule use. Results suggest that students' cognitive styles do indeed influence the structure of their logical discourse, and that the effect may be attributable to the relative skill with which students manipulate graphical abstractions.

Introduction: multimodal logical discourse

Computer-based multimodal tools are giving people the freedom to express themselves in brand new ways. But what do people actually *do* when given these tools? Does everyone end up generating the same forms of multimodal discourse? Do multimodal systems lead to better performance than monomodal systems?

These questions arise in many areas related to human-computer interaction, but they are particularly important in educational applications, since multimodality is believed to be especially helpful to novices (di Sessa 1979, Schwarz and Dreyfus 1993). Hyperproof is a program created by Barwise and Etchemendy (1994) for teaching first-order logic (see Figure 1). Inspired by a situation-theoretic approach to heterogeneous reasoning, it uses multimodal (graphical and sentential) methods, allowing users to transfer information to and fro, between modalities (see Figure 2).

We have been carrying out a series of experiments on Hyperproof, to help evaluate its effects on students learning logic. The study has established that there are important individual differences in the way students respond to logic taught multimodally (Cox, Stenning and Oberlander 1994; Stenning, Cox and Oberlander 1995). In the course of this larger study, we have built up a substantial corpus of proofs. These 'hyperproofs' are an unusual form of discourse, for two main reasons. Firstly, they are primarily used for *self*-communication: a student arranges proof steps and rules in an external representation as an aid to their individual problem-solving activities. Secondly, hyperproofs are, of course, *multimodal* discourse: they involve both language and graphics, and are therefore in some ways more complex than text or speech.

We believe that the corpus of hyperproofs can provide a detailed insight into the paths which students follow in their

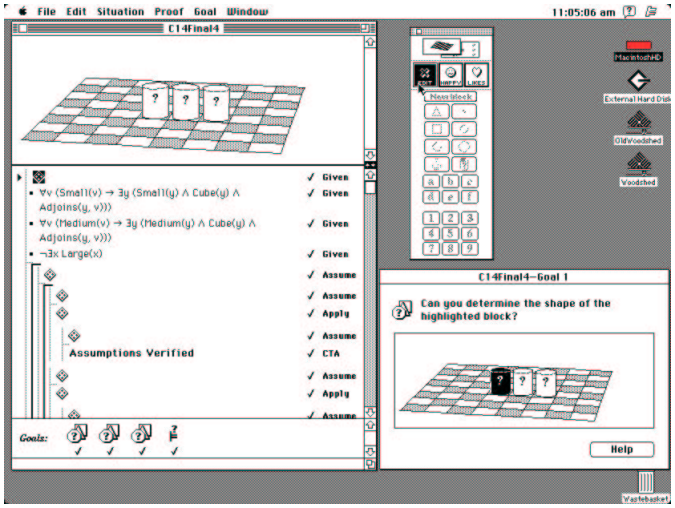

Figure 1: The Hyperproof Interface. The main window (top left) is divided into an upper graphical pane, and a lower calculus pane. The tool palette is floating next to the main window, and other windows can pop up to reveal a set of goals which have been posed.

Apply Extracts information from a set of sentential premises; expresses it graphically

Assume Introduces a new assumption into a proof, either graphically or sententially

Observe Extracts information from the situation; expresses it sententially

Inspect Extracts common information from a set of cases; expresses it sententially

Merge Extracts common information from a set of cases; expresses it graphically

Close Declares that a sentence is inconsistent with either another sentence, or the current graphical situation

CTA (Check truth of assumptions) Declares that all sentential and graphical assumptions are true in the current situation

Exhaust Declares that a part of a proof exhausts all the relevant cases

Figure 2: A set of relevant Hyperproof rules.

pursuit of proof goals. In this paper, we therefore first frame some hypotheses concerning the relation between the individual differences in teaching outcome which we found, and the structures to be found in students' proofs. We then outline the relevant aspects of the design of the main study, indicating how it distinguishes two styles of student. We then describe (i) the way 'proofograms' are used to track the way students deal with abstractions; and (ii) the application of bigram and trigram analyses of rule use patterns in the data corpus, demonstrating that the differing styles of student end up producing multimodal proofs of distinctive types.

Hypotheses

The observation that graphical systems require certain classes of information to be specified goes back at least to Bishop Berkeley. Elsewhere, we have termed this property 'specificity', and argued that it is useful because inference with specific representations can be very simple (Stenning and Oberlander 1991, 1995). We have also urged that actual graphical systems do allow abstractions to be expressed, and it is this that endows them with a usable level of expressive power. Thus, Hyperproof maintains a set of abstraction conventions for objects' spatial or visual attributes. As well as concrete depictions of objects, there are 'graphical abstraction symbols', which leave attributes under-specified: the *cylinder*, for instance, depicts objects of unknown size (see Figure 1). A key step, then, in mastering an actual graphical system is to learn which abstractions can be expressed, and how.

As we describe below, our pre-tests independently allowed us to divide subjects into two cognitive style groups, on the basis of their performance on a certain type of problem item. Loosely, one group is 'good with diagrams', and the other less so. The good diagrammers turned out to benefit more from Hyperproof-based teaching than the others. Our belief is that those who benefit most from Hyperproof do so because they are better able to manipulate the graphical abstractions it offers. Call this view the *abstraction ability hypothesis*.

A secondary issue concerns the relation between our binary distinction in cognitive styles and more traditional dimensions of individual difference—such as the 'visualiser–verbaliser' dimension. One hypothesis is that the good diagrammers are simply those subjects who have a preference for the visual modality. Call this view the *visual preference hypothesis*.

In what follows, we aim to show that the first hypothesis is vindicated by the analysis, but that the second is not, and that a rival view might fit the data better.

Method

In the full study, two groups of subjects were compared; one ($n = 22$ at course end) attended a one-quarter duration course taught using the multimodal Hyperproof. A comparison group ($n = 13$ at course end) were taught for the same period, but in the traditional syntactic manner supplemented with exercises using a graphics-disabled version of Hyperproof.

Distinguishing cognitive styles

Subjects were administered two kinds of pre- and post-course paper and pencil test of reasoning. The first of these is most relevant to the current discussion. It tested 'analytical reasoning' ability, with two kinds of item derived from the GRE scale of that name (Duran, Powers and Swinton 1987). One subscale consists of verbal reasoning/argument analysis. The other subscale consists of items often best solved by constructing an external representation of some kind (such as a table or a diagram). We label these subscales as 'indeterminate' and 'determinate', respectively. Scores on the latter subscale were used to classify subjects within both Hyperproof and Syntactic groups into DetHi and DetLo sub-groups. The score reflects subjects' facility for solving a type of item that often is best solved using an external representation; DetHi scored well on these items; DetLo less well. For the moment, we may consider DetHi subjects to be more 'diagrammatic', and DetLo to be less so. Obviously, the relation between diagrammatic ability and the visualiser–verbaliser dimension is an issue to which we return below.

Computer-based protocols

Both the Hyperproof and Syntactic groups contained DetHi and DetLo sub-groups. All subjects sat post-course, computer-based exams, although the questions differed for the two groups, since the Syntactic group had not been taught to use Hyperproof's systems of graphical rules. Student-computer interactions were dynamically logged—this approach might be termed 'computer-based protocol taking'. The logs were time stamped and permitted a full, step-by-step, reconstruction of the time course of the subject's reasoning.

Here, we discuss only the data from the 22 Hyperproof subjects, all of whom completed the exams. The four questions that these students were set contained two types of item: determinate and indeterminate. Here, determinate problems were taken to be those whose problem statement did not utilise Hyperproof's abstraction conventions. That is: determinate problems contained only concrete depictions of objects in their initially given graphical situation, whereas indeterminate problems—such as that in Figure 1—could contain graphical abstraction symbols in the initial situation.

Results

A proof log captures both the *process* of proof development, and its *product*—the final proof submitted by the subject. Here we discuss the data extracted from the latter.

Proofograms

What evidence is there for the abstraction ability hypothesis? Among the Hyperproof students, do the two sub-groups—DetHi and DetLo—use graphical abstraction symbols in characteristically different ways? To investigate this, we scored each step of each proof on the basis of number of concrete situations compatible with the graphical depiction. We give each graphical symbol in a situation a score: for each visible attribute (size, shape, and location) a symbol scores 1 if that attribute is specified, and 0 otherwise. By totalling the scores for the individual symbols, we can give each situation in a proof a score. For example, in Figure 3, the total concreteness score for the situation shown would be 9, since each object is fully specified; in Figure 5, the score would be just 6, since one object is specified only for location, and another only for size and location. A low score indicates more abstraction; a higher score indicates more concreteness.

We can explore the way concreteness varies through the course of a proof by graphing it against the hierarchical struc-

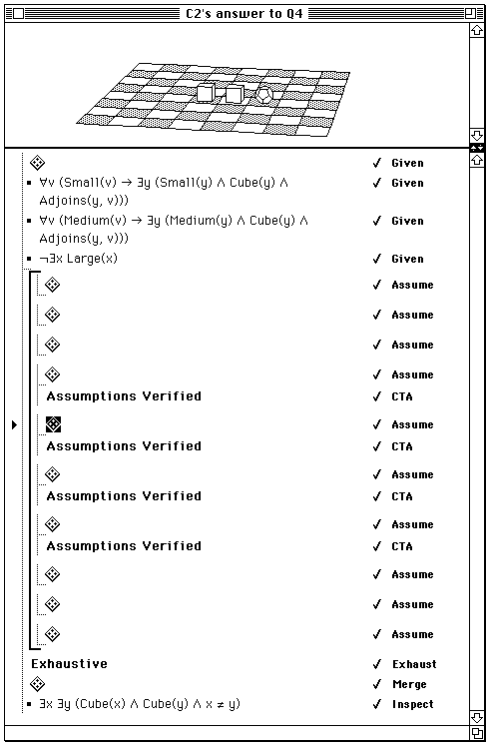

Figure 3: Submitted proof for a DetLo subject (C2) attempting an indeterminate question (Q4). The situation on view is from the 9th step of the proof.

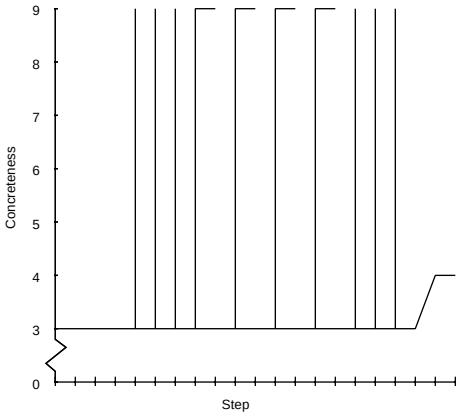

Figure 4: Proofogram for C2 attempting Q4. Proof steps are plotted on the x-axis; the concreteness of the current graphical situation is computed for each step of the proof, and is plotted on the y-axis. Horizontal lines indicate dependency structure; vertical lines indicate uses of Assume; sloping lines indicate uses of Apply or Merge. C2's proofogram is 'spikey', indicating a series of independent, concrete cases.

ture of the proof. We call such graphs 'proofograms'. Figures 3 and 5 show how subjects C2 and C14 tackle an indeterminate exam question; Figures 4 and 6 give their proofograms.

The visual differences between proofograms are quite striking: one group is 'spikey'—as in Figure 4; and the other is 'layered'—as in Figure 6. The differences are most pronounced on the 2 indeterminate exam questions. The visual grouping of proofograms suggests the existence of a 'staging phenomenon': DetHi introduce concreteness *by stages*, whereas DetLo introduce it more immediately. In terms of proof structure, DetHi tend to produce structured sets of cases, with superordinate cases involving graphical abstraction; DetLo tend to produce sets of cases without such overt superordinate structure.

To assess whether this apparent patterning was reliable, the 88 proofograms (4 exam questions for each of the 22 Hyperproof subjects) were printed. The proofograms were randomly ordered, and two prototypes (one spikey, one layered) were selected as category exemplars. Two independent raters then assigned each proofogram to either the 'spikey' or 'layered' category, under a forced choice regime. There was a high degree of inter-rater agreement, with a discrepancy on only 2 of the 88 proofograms. A third observer was employed to resolve the two categorisation disagreements.

To test for existence of the staging phenomenon, the concordance between subject type (DetLo/Hi) and proofogram style was analysed. For each of the 4 exam questions, 2 $\times$ 2 tables were produced, showing the number of items in each cell (DetLo/spikey; DetLo/layered; DetHi/spikey;

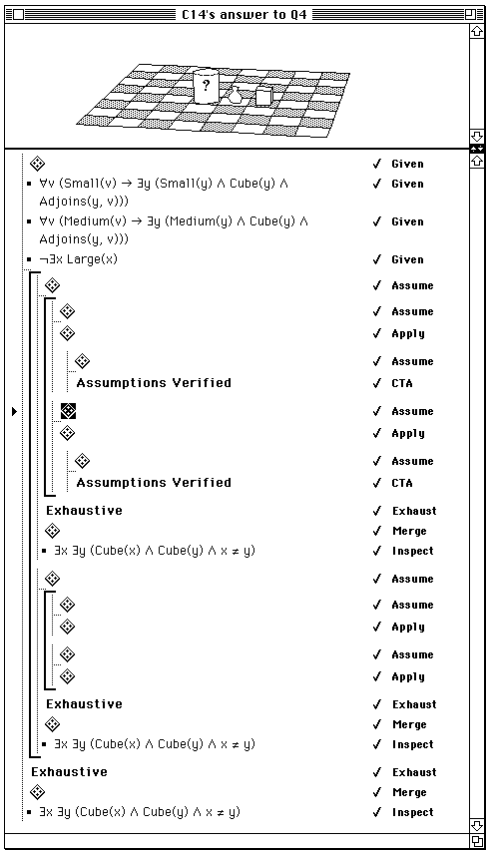

Figure 5: Submitted proof for a DetHi subject (C14) attempting an indeterminate question (Q4). The situation on view is from the 9th step of the proof.

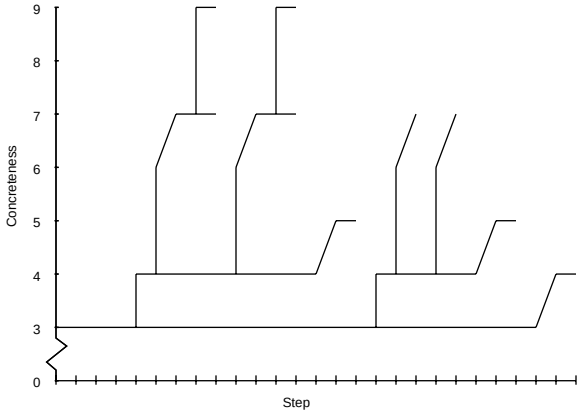

Figure 6: Proofogram for C14 attempting Q4. C14's proofogram is 'layered', indicating parallel sub-case structures with abstract superordinate cases.

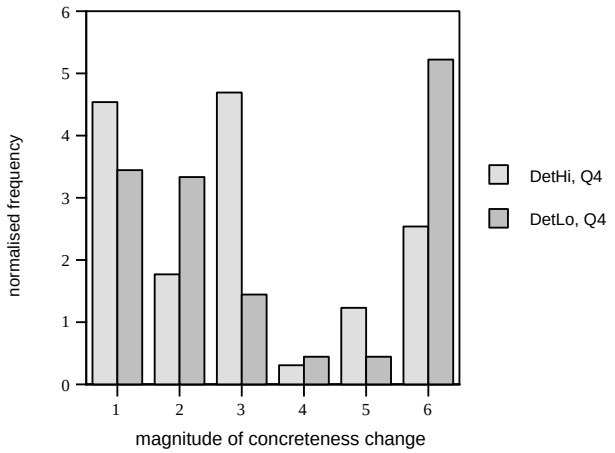

Figure 7: Frequencies with which DetHi and DetLo subject groups employ changes in proof concreteness of varying magnitude, when attempting Question 4. The frequency is normalised to take into account the differing size of the subject groups (DetHi $n = 13$, DetLo $n = 9$).

DetHi/layered). A nonparametric measure of association (ϕ coefficient) was calculated for each table. The results indicated that the hypothesised association only held on indeterminate questions (on question 2, $\phi = .43^{*}$; on question 4, $\phi = .28$). On questions 1 and 3 (determinate questions), both raters assigned all proofograms to the spikey category.

It seems, then, that on indeterminate questions, DetHi subjects do differ from DetLo subjects, in that they are more prone to develop layered proofs, introducing concreteness by stages. Evidence for the staging phenomenon therefore provides support for the abstraction ability hypothesis: the two groups are certainly using abstractions in different ways.

It would, of course, be convenient to be able to encapsulate the graphical attributes of the proofogram in numerical form. Our first attempts to do so have involved computing mean change in concreteness per proof step per proof. However, the change only exceeds unity on the most indeterminate exam problem—question 4. In fact, we can consider the frequency with which subjects employ changes in concreteness of varying magnitude. Figure 7 graphs the differing behaviour of the subject groups on question 4. This reinforces the idea that DetHi subjects tend to make small changes in concreteness, whereas DetLo subjects make larger changes.

Corpus analysis

Of Hyperproof's rules, only Assume, Apply and Merge increase concreteness. We therefore examined the kind of patterns in which they occur through proof-corpus analysis. The proofogram results already indicate that DetHi and DetLo differ in the way they handle concreteness. Since Assume is by far the most frequent means of adding concreteness (see, for instance, Table 3 below), the corpus analysis distinguishes between uses of the rule which introduce totally concrete graphical situations, and those which leave some abstractness in the graphic. The term Fullassume denotes the former type of use, and assume denotes the latter.

Using techniques developed originally for the analysis of linguistic corpora, we have carried out bigram and trigram analyses of rule use, utilising Dunning's (1993) 'Log–Likelihood Test', which can be applied to relatively small corpora. The test is designed to "highlight particular *A*'s and *B*'s that are highly associated in text" (p.71). Ranking the bigrams according to this test provides a good *profile* of the individual's, or the group's, rule use in the corpus. We can then compare the profiles for the sub-groups on the two question types, assessing the significance of a given bigram by using the χ^2 test on the log-likelihood value.

Tables 1 and 2 illustrate the nature of the resulting profiles, and show the most important parts of the bigram profiles for DetHi and DetLo on indeterminate and determinate questions, respectively. Taking the profiles for the two groups, we can consider differences both between-groups and within-groups; the former are the most interesting.

On indeterminate questions, we find that the bigrams assume Apply, Merge Inspect, CTA Observe, assume Close, Given assume, and assume Fullassume are significant in DetHi proofs, but not in DetLo ones. Conversely, only the bigram Inspect Merge is significant in DetLo proofs, but not in DetHi ones. The profiles are weakly but significantly correlated ($r = 0.167^{*}$).[1] When taking into account only those bigrams that are significantly associated in the profiles, the correlation is higher, but not significant ($r = 0.315, ns$).

On determinate questions, the bigrams assume Apply, CTA Observe and Close Fullassume are significant in DetHi proofs, but not in DetLo ones. Conversely, as with the indeterminate questions, the only bigram significant in DetLo proofs, but not in DetHi ones, is Inspect Merge. Here, the two subject group's profiles are significantly correlated ($r = 0.537^{**}$). The correlation between significantly-associated bigrams is even stronger and still highly significant ($r = 0.918^{**}$).

This finding accords with the proofograms' indication that it is indeterminate questions which best discriminate the two subject groups. Recall that these are the questions in which the

[1] Correlations reported here are non-parametric (Spearman's ρ). Significance at the $p < .05$ level is denoted by *; significance at the $p < .001$ level by **.

Table 1: Bigram profiles for subjects' indeterminate questions: (a) DetHi; (b) DetLo. The first column indicates Dunning's 'log-likelihood'; the higher the number, the more 'natural' the association. $k(AB)$ is a count of the number of times the bigram AB occurs, $k(A \sim B)$ is a count of the number of times A is followed by a rule other than B, and so on. For reasons of space, we show only those bigrams that are significantly associated ($p < .05$).

DETHI

-2logλ	k(AB)	k(A$\sim$B)	k($\sim$AB)	k($\sim$A$\sim$B)	A	B
145.43	57	22	29	355	Fullassume	CTA
123.52	26	13	5	419	Exhaust	Merge
78.72	33	85	3	342	assume	Apply
69.47	17	11	12	423	Merge	Inspect
53.63	39	43	40	341	CTA	Fullassume
36.46	2	77	116	268	Fullassume	assume
26.16	14	68	7	374	CTA	Observe
26.01	12	27	17	407	Exhaust	Inspect
25.06	15	103	4	341	assume	Close
19.56	17	9	101	336	Given	assume
19.55	6	112	73	272	assume	Fullassume

DETLO

-2logλ	k(AB)	k(A$\sim$B)	k($\sim$AB)	k($\sim$A$\sim$B)	A	B
134.90	67	33	15	221	Fullassume	CTA
78.70	19	13	5	299	Exhaust	Inspect
36.75	10	8	11	307	Inspect	Merge
34.48	1	99	54	182	Fullassume	assume
33.09	45	35	55	201	CTA	Fullassume
27.97	11	21	10	294	Exhaust	Merge

Table 2: Bigram profiles for subjects' determinate questions: (a) DetHi; (b) DetLo. For reasons of space, we again show only those bigrams that are significantly associated ($p < .05$).

DETHI

-2logλ	k(AB)	k(A$\sim$B)	k($\sim$AB)	k($\sim$A$\sim$B)	A	B
112.37	24	11	5	366	Exhaust	Merge
72.82	48	73	14	271	Fullassume	CTA
65.00	15	13	6	372	Merge	Inspect
36.39	35	86	17	268	Fullassume	Close
26.79	10	16	15	365	Given	Apply
26.16	16	36	19	335	Close	Exhaust
23.08	11	27	14	354	assume	Apply
20.19	32	25	89	260	CTA	Fullassume
13.65	10	47	12	337	CTA	Observe
10.78	26	26	95	259	Close	Fullassume

DETLO

-2logλ	k(AB)	k(A$\sim$B)	k($\sim$AB)	k($\sim$A$\sim$B)	A	B
51.49	13	8	5	207	Exhaust	Merge
50.03	28	47	4	154	Fullassume	CTA
41.35	10	8	4	211	Given	Apply
38.05	9	6	5	213	Merge	Inspect
24.93	23	52	9	149	Fullassume	Close
19.20	11	21	11	190	Close	Exhaust
13.99	4	2	14	213	Inspect	Merge
12.59	18	11	57	147	CTA	Fullassume

initial graphical situation is abstract, so that all concreteness must be introduced explicitly by the subjects.

Discussion

The proofogram and corpus analyses provide evidence that subjects differ in the way that they use the graphical abstraction conventions of Hyperproof. On questions where the subject must construct the concrete graphic, it seems that DetHi subjects exhibit staging behaviour, and build their graphics incrementally, whereas DetLo subjects are prone to construct their concrete graphics in one go. The abstraction ability hypothesis thus seems plausible, since the 'stagers' are exactly those whom our main study showed benefit most from teaching with Hyperproof (Stenning, Cox and Oberlander, 1995).

Snow (1987) and colleagues, in their studies of aptitude-treatment interactions (ATI), characterise such within-person adaptations and flexibilities as an important source of individual differences in complex skill performance. Snow (1987) reports that subjects' ability to 'strategy-shift' is particularly detectable on tests of complex spatial visualization such as the paper-folding test (Test VZ2 in Ekstrom, French and Harmon, 1976). Such tests involve mental manipulations that are very similar to those required for skilled use of a multimodal system such as Hyperproof.

However, to characterise the difference between subjects solely in terms of visuo-spatial ability differences—or in terms of cognitive style differences along a 'visualiser–verbaliser' dimension—may be too hasty. To be sure, the visual preference hypothesis has some initial plausibility: if the DetHi are 'good with diagrams', perhaps they are simply the visualisers, and have a preference for the visual modality. However, the evidence from the corpus goes against the hypothesis.

First, consider the way that use of assume and Fullassume varies between the DetHi and DetLo groups, as shown in Table 3. DetHi make more use of assume than DetLo, while the latter make more use of Fullassume than the former. The bigram assume Fullassume is found to be significant in DetHi indeterminate proofs, but not in DetLo proofs. However, these facts do not support the hypothesis that DetHi prefer visual over verbal. On the contrary, DetLo subjects' favouring of Fullassume over assume confirms that they are not 'stagers': in a sense, it is *they* who exhibit a preference for the graphical modality. By contrast, DetHi subjects' use of assume indicates gradual addition of information to the graphical window pane, either by assumption, or by transfer from the sentential pane (via Apply). The difference seems to be that the DetHi group *operate over* the graphical situations, frequently using a graphic as input to further stages of proof construction. The DetLo, on the other hand, seem just to *output* graphics, without subsequently using them.

Table 3: Occurrences of Assume, Apply, and Observe.

RULE	frequency per group		frequency per subject	
	DetHi	DetLo	DetHi	DetLo
assume	143	62	11	6.9
Fullassume	222	186	17	20.7
Apply	61	24	4.7	2.7
Observe	27	9	2.1	1

In addition, Table 3 indicates that DetHi subjects make more use of rules that transfer information between the modalities. DetHi make the bigram assume Apply a significant component of all their proofs, and use it more frequently than DetLo: 44 times to just 7. By contrast, DetLo exhibit a tendency to invoke Apply as the first rule in their proofs (giving rise to the bigram Given Apply). Subsequent interaction between the modalities is thereby reduced, with case construction being performed only within the graphical window.

Thus, DetHi subjects do *not* show a simple preference for the visual–graphical modality. Rather, what distinguishes the DetHi subjects is their greater tendency to *translate* between graphical and sentential modalities in *both* directions.

Perhaps, as Monaghan (1995) has suggested, the individual differences between subjects might be better captured by two or more cognitive style dimensions. There may, for example, be an interaction between individuals' *processing* style and the preferences that they may exhibit for information *representation*. One promising candidate for the second factor is the field-dependence and field-independence dimension. Field-independent individuals have been found to prefer more formal methods of instruction, to rely more upon internal frames of reference, and to perform better on tasks that require cognitive re-structuring. They also seem better able than field-dependents to represent concepts analytically rather than taking on ideas as presented (Jonassen and Grabowski, 1993; Witkin and Goodenough, 1981). So, DetHi individuals might well be more field independent—but we should not immediately conclude that the differences are purely representational, as opposed to operational.

Another possibility is that DetHi subjects may just have higher levels of expertise. Research in the physics domain (such as Chi, Feltovich and Glaser, 1981; Larkin, McDermott, Simon and Simon, 1980) has shown that expertise is characterised by greater domain knowledge, with an ability to classify problems according to deep structure and physical principles, whereas novices tend to classify problems on the basis of surface features. Experts also tend to spend more time than novices on analysing and understanding problems, but produce faster solutions. Working forward is typical of experts, whereas novices tend to work backwards.

An account of the differences in terms of expertise seems implausible, however, for at least two reasons. First, subjects in both groups received equal exposure to Hyperproof and it is difficult to see how the DetHi could have acquired more domain knowledge than the DetLo. Secondly, the expertise literature would predict that DetHi produce faster solutions. However, the two groups did not differ significantly in terms of solution times on any of the four Hyperproof exam questions.

Traditional psychometric approaches to the measurement of cognitive and learning styles contribute detailed and useful characterisation of human behaviour, but at the level of description and taxonomy. Micro-analyses of process, of rule usage patterns, is a methodology that promises to extend such accounts and is one that we expect to pursue. The next phase will involve the building of computational models, testing the theoretically important parameter of abstraction ability. This approach should make a useful contribution to the development of a cognitive characterisation of just what it means in computational terms to be a 'verbal' or 'visual' thinker.

Acknowledgements

The support of the Economic and Social Research Council for HCRC is gratefully acknowledged. The work was supported by UK Joint Councils Initiative in Cognitive Science and HCI, through grant G9018050 (Signal); and by NATO Collaborative research grant 910954 (Cognitive Evaluation of Hyperproof). The first author is supported by an EPSRC Advanced Fellowship. Special thanks to John Etchemendy, Tom Burke and Mark Greaves at Stanford, and Chris Brew at Edinburgh; our thanks also to our three anonymous referees, for their helfpul comments.

References

Barwise, J. and Etchemendy, J. (1994). *Hyperproof*. CSLI Lecture Notes. Chicago: Chicago University Press.

Chi, M. T. H., Feltovich, P. J. and Glaser, R. (1981). Categorisation and representation of physics problems by experts and novices. *Cognitive Science*, **5**, 121–152.

Cox, R., Stenning, K. and Oberlander, J. (1994). Graphical effects in learning logic: reasoning, representation and individual differences. In *Proceedings of the 16th Annual Meeting of the Cognitive Science Society*, pp237–242, Atlanta, Georgia, August.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* **19**, 61–74.

Duran, R., Powers, D. and Swinton, S. (1987). Construct Validity of the GRE Analytical Test: A Resource Document. ETS Research Report 87–11. Princeton, NJ: Educational Testing Service.

Ekstrom, R. B., French, J. W. and Harmon, H. H. (1976). *Manual for Kit of Factor Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service (ETS).

Jonassen, D. H. and Grabowski, B. L. (1993). *Handbook of individual differences, learning and instruction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Larkin, J. H., McDermott, J., Simon, D. and Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, **208**, 1335–1342.

Monaghan, P. (1995). A corpus-based analysis of individual differences in proof-style. MSc Thesis, Centre for Cognitive Science, University of Edinburgh.

Schwarz, B. and Dreyfus, T. (1993). Measuring integration of information in multirepresentational software. *Interactive Learning Environments*, **3**, 177–198.

di Sessa, A. A. (1979). On 'learnable' representations of knowledge: A meaning for the computational metaphor. In Lochhead, J. and Clement, J. (Eds.) *Cognitive Process Instruction*. Philadelphia, PA: The Franklin Institute Press.

Snow, R. E. (1987). Aptitude complexes. In Snow, R. E. and Farr, M. J. (Eds.) *Aptitude, learning, and instruction, Volume 3: Conative and affective process analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Stenning, K., Cox, R. and Oberlander, J. (1995). Contrasting the cognitive effects of graphical and sentential logic teaching: reasoning, representation and individual differences. *Language and Cognitive Processes*, **10**, 333–354.

Stenning, K. and Oberlander, J. (1991). Reasoning with Words, Pictures and Calculi: computation versus justification. In Barwise, J., Gawron, J. M., Plotkin, G. and Tutiya, S. (Eds.) *Situation Theory and Its Applications*, Volume 2, pp607–621. Chicago: Chicago University Press.

Stenning, K. and Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science*, **19**, 97–140.

Witkin, H. A. and Goodenough, D. R. (1981). *Cognitive styles: Essence and origins: Field dependence and field independence*. New York, NY: International Universities Press.