Proof styles in multimodal reasoning

Jon Oberlander Richard Cox Keith Stenning

August 29, 1994

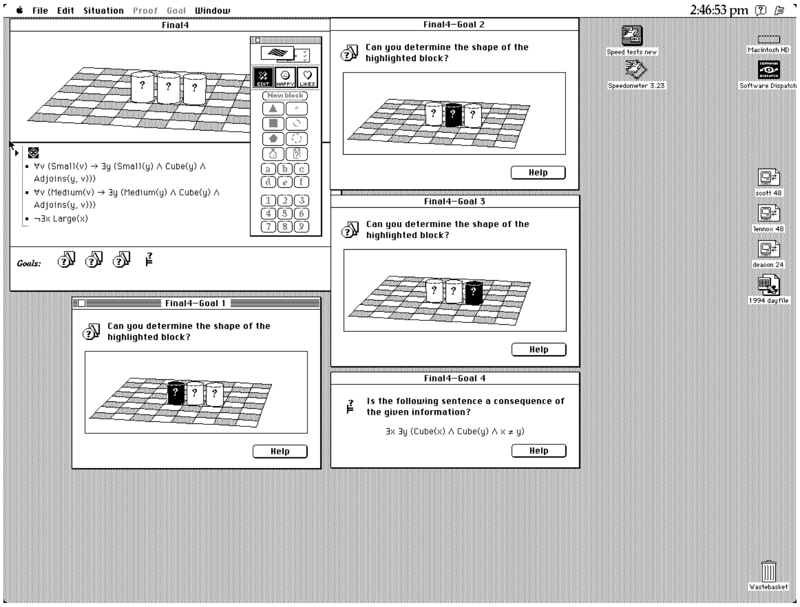
1 Introduction: questions of style

Does multimodal logic teaching create a homogeneous population of human theorem provers? Or do students develop individual styles of proof? If their styles differ, what patterns emerge? Can these patterns be predicted from other information about the students?

Hyperproof is a computer program created by Barwise and Etchemendy for teaching first-order logic. It uses multimodal graphical and sentential methods, and is inspired by a situation-theoretic approach to heterogeneous reasoning (Barwise and Etchemendy 1994). A distinctive feature of Hyperproof is its set of 'graphical' rules, which permit users to transfer information to and fro, between graphical and linguistic modes. We have been carrying out a series of experiments on Hyperproof, to help evaluate its effects on students learning logic.

In earlier work (Stenning and Oberlander 1991, in press), we have emphasised the idea that graphical systems possess a useful property—over-specificity—whereby certain classes of information must be specified. The property is useful because inference with such specific representations can be very simple. We have also urged that actual graphical systems—such as Hyperproof—do allow abstractions to be expressed, and it is this that endows them with a usable level of expressive power. We are therefore interested in determining empirically how students respond to Hyperproof's abstraction mechanisms.

The plan of this paper is as follows: we introduce Hyperproof, and then study two cases of proofs constructed in Hyperproof. The two students addressed the same problem, but we observe a number of significant differences in the way they solved it. We then discuss the experimental regime under which these proofs were gathered; we focus on our finding that there is a robust distinction between subjects who are more or less successful on an independent task, whose solution can involve the use of external representations (such as tables). We then return to properties we noted in the case studies, and show how their patterns of rule use and proof structure reflect systematic differences between the two classes of subjects. We conclude by suggesting how these patterns might be explained by the 'specificity hypothesis' we have developed in earlier work.



- ∀v (Small(v) → ∃y (Small(y) ∧ Cube(y) ∧ Adjoins(y, v)))
- ∀v (Medium(v) → ∃y (Medium(y) ∧ Cube(y) ∧ Adjoins(y, v)))
- ¬∃x Large(x)

Can you determine the shape of the highlighted block?

Can you determine the shape of the highlighted block?

Can you determine the shape of the highlighted block?

Is the following sentence a consequence of the given information?

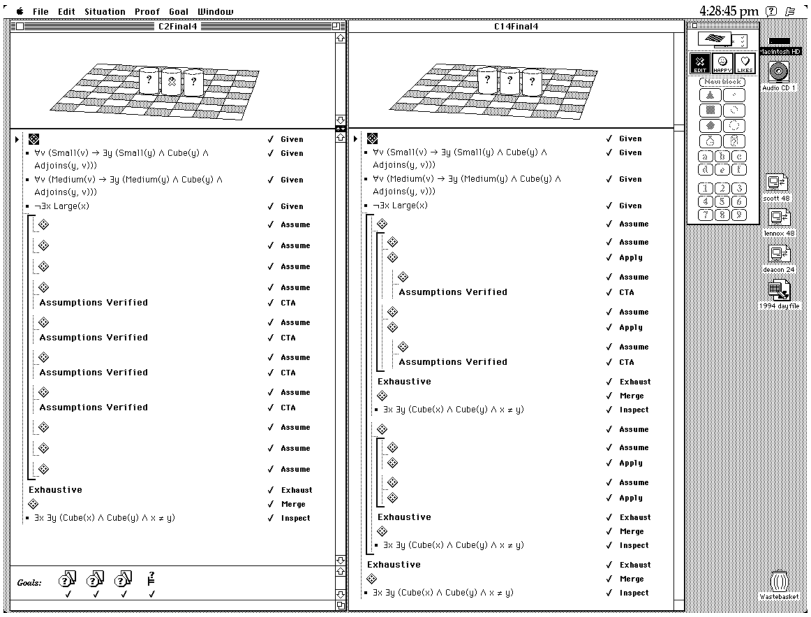
∃x ∃y (Cube(x) ∧ Cube(y) ∧ x ≠ y)

FIGURE 1 The Hyperproof Interface.
The main window (top left) is divided into an upper graphical window, and a lower calculus window. The tool palette is floating on top of the main window, and the other windows reveal a set of goals which have been posed. To achieve them, a proof must be developed, by applying a set of multimodal inference rules to the graphical and calculus premises given.

2 The Hyperproof Interface

As can be seen in Figure 1, the interface contains two main windows: one presents a diagrammatic view of a chess-board world containing geometric objects of various shapes and sizes; the other presents a list of sentences in predicate calculus; control palettes are also available. The main windows are used in the construction and editing of proofs. Several types of goals can be proved, involving the shape, size, location, identity or sentential descriptions of objects; in each case, the goal can involve determining some property of an object, or showing that a property *cannot* be determined from the given information. A number of rules are available for proof construction; some of these are traditional syntactic rules (such as ∧-elimination); others are 'graphical', in the sense that they involve consulting or altering the situation depicted in the diagrammatic window. In addition, a number of rules check properties of a developing proof. Hyperproof should be viewed as a proof-checking environment designed to support human theorem proving using heterogeneous information.

In Question 4, shown here, a student is confronted by a graphical si-



```
C2Final4

◈                                                   ✓ Given
▪ ∀v (Small(v) → ∃y (Small(y) ∧ Cube(y) ∧           ✓ Given
  Adjoins(y, v)))
▪ ∀v (Medium(v) → ∃y (Medium(y) ∧ Cube(y) ∧         ✓ Given
  Adjoins(y, v)))
▪ ¬∃x Large(x)                                      ✓ Given
  ◈                                                 ✓ Assume
  ◈                                                 ✓ Assume
  ◈                                                 ✓ Assume
  ◈                                                 ✓ Assume
  Assumptions Verified                              ✓ CTA
  ◈                                                 ✓ Assume
  Assumptions Verified                              ✓ CTA
  ◈                                                 ✓ Assume
  Assumptions Verified                              ✓ CTA
  ◈                                                 ✓ Assume
  Assumptions Verified                              ✓ CTA
  ◈                                                 ✓ Assume
  ◈                                                 ✓ Assume
  ◈                                                 ✓ Assume
Exhaustive                                          ✓ Exhaust
◈                                                   ✓ Merge
▪ ∃x ∃y (Cube(x) ∧ Cube(y) ∧ x ≠ y)                 ✓ Inspect
```

```
C14Final4

◈                                                   ✓ Given
▪ ∀v (Small(v) → ∃y (Small(y) ∧ Cube(y) ∧           ✓ Given
  Adjoins(y, v)))
▪ ∀v (Medium(v) → ∃y (Medium(y) ∧ Cube(y) ∧         ✓ Given
  Adjoins(y, v)))
▪ ¬∃x Large(x)                                      ✓ Given
  ◈                                                 ✓ Assume
    ◈                                               ✓ Assume
    ◈                                               ✓ Apply
      ◈                                             ✓ Assume
      Assumptions Verified                          ✓ CTA
    ◈                                               ✓ Assume
    ◈                                               ✓ Apply
      ◈                                             ✓ Assume
      Assumptions Verified                          ✓ CTA
  Exhaustive                                        ✓ Exhaust
  ◈                                                 ✓ Merge
  ▪ ∃x ∃y (Cube(x) ∧ Cube(y) ∧ x ≠ y)               ✓ Inspect
  ◈                                                 ✓ Assume
    ◈                                               ✓ Assume
    ◈                                               ✓ Apply
    ◈                                               ✓ Assume
    ◈                                               ✓ Apply
  Exhaustive                                        ✓ Exhaust
  ◈                                                 ✓ Merge
  ▪ ∃x ∃y (Cube(x) ∧ Cube(y) ∧ x ≠ y)               ✓ Inspect
Exhaustive                                          ✓ Exhaust
◈                                                   ✓ Merge
▪ ∃x ∃y (Cube(x) ∧ Cube(y) ∧ x ≠ y)                 ✓ Inspect
```

FIGURE 2 Two different subjects' proofs given in answer to an exam question. The subject on the left (C2) is a 'DetLo' (more verbal in style); the subject on the right (C14) is a 'DetHi' (more diagrammatic in style).

tuation in which little is known—only that there are three objects of unknown shape or size, side by side—and a set of three linguistic premises—comprising two conditionals relating the shape and size of two objects, and a formula telling us that there are no objects of a certain size. When little is fixed in the graphical situation we term a question *indeterminate* in type, and contrast it with those *determinate* questions in which all the relevant information is specified; we return to these notions in Section 4. Here, the student must achieve four proof goals: the first three are shape goals: the students must determine the shapes of the three objects in the world; the last goal is a syntactic goal: the student must determine whether or not a certain formula follows from the graphical and linguistic premises.

3 Case studies

Now consider the two proofs side-by-side in Figure 2. These are responses to Question 4, developed by students using Hyperproof under exam conditions. The students (C2 and C14) were both successful with this exam question: they proved all four goals. But they did so with proofs which look somewhat different. There are at least three considerable differences, involving: (i) structural aspects of the proofs; (ii) patterns of rule use; and (iii) treatment of 'graphical variables'. Let us discuss each of these in turn.

TABLE 1
A set of relevant Hyperproof rules.

RULE	DESCRIPTION
Apply	Extracts information from a set of sentential premises, and expresses it graphically
Assume	Introduces a new assumption into a proof, either graphically or sententially
Inspect	Extracts common information from a set of cases, and expresses it sententially
Merge	Extracts common information from a set of cases, and expresses it graphically
Observe	Extracts information from the situation, and expresses it sententially
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

3.1 Structural aspects of proofs

The most obvious difference between C2's proof and C14's is that is shorter: C2 uses just 21 steps to achieve the four goals, while C14 uses 27.[1] A second obvious difference is that C2's proof is shallower, in that there is just one level of embedding in C2's proof (indicated by the heavy bar enclosing user steps 1–14 of C2's proof). C14, by contrast, uses two such levels, and within these, uses further embedding (cf. the relation between user steps 3 and 4 of C14's proof). However one measures 'proof depth', it is clear just from inspection that C14's proof is more deeply nested than C2's. There is a third global difference between C2 and C14's proofs, although this one is not apparent from direct inspection of Figure 2. C2's proof was constructed more rapidly than C14's: C2 takes 14 minutes, while C14 takes 19.

So, to summarise, C2's proof in Question 4 is shorter, shallower, and faster.

3.2 Patterns of rule use

Looking at the proofs in more detail, we find that there are differences in the way that rules have been deployed. Hyperproof's relevant graphical rules are summarised in Table 1. Consider first the pattern that characterises C2's proof: the subproof from steps 1–14 begins and ends with three invocations of Assume; the central part of it possesses a repetitive pattern of Assume, CTA, Assume, CTA, Assume, CTA.

Both C2 and C14 end their proofs with the same final pattern: they show that they have a exhausted a set of cases, with Exhaust, then show

[1] These figures include premises.

what is common to all of them in a graphical situation, with Merge, and then draw out the linguistic conclusion from that situation, via Inspect.

But the body of C14's proof differs from C2's. It is true that C14 has two instances of an Assume–CTA pattern; but these are essentially parallel structures, each following an instance of an Assume–Apply pattern. Further on in C14's proof, we also find a further pair of Assume and Apply patterns. And while C2 uses the Exhaust–Merge–Inspect pattern just once, at the end of their proof, C14 uses it three times in total, ascending twice from subsubproofs, and once from the larger subproof.

It is clear that these differing patterns fit together with the relative depth of the two proofs: C2 uses repetitive patterns in a shallow proof; C14 avoids the repetition by creating a deeper proof, which in turn requires a particular pattern of rules to recur, as information is drawn together from exhaustive sets of subcases.

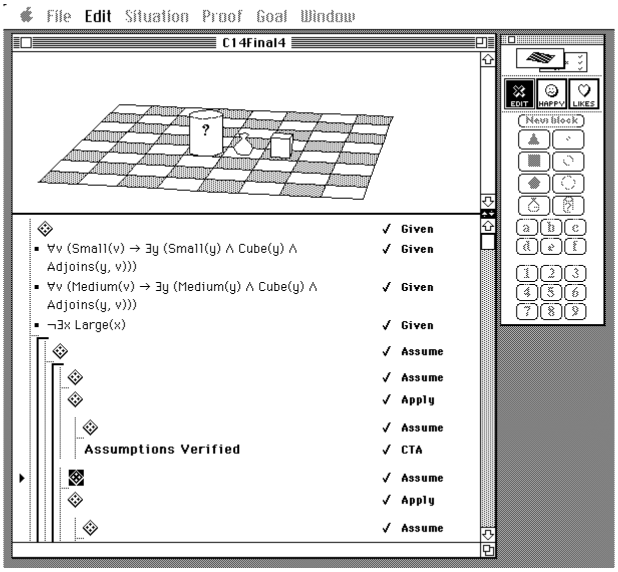
In sum, C2 uses Assume–CTA more frequently and repetitively than C14. C14 uses Assume–Apply, and Exhaust–Merge–Inspect more frequently.

3.3 Treatment of graphical variables

A further difference emerges when one examines the graphical situations which correspond to the diamond-shaped situation icons in the body of the proof. Obviously, C2 uses fewer graphical situations than C14, with 11 as opposed to 15. However, the graphical situations themselves are interestingly different.

Hyperproof's graphical window contains two sorts of symbols, which we may think of as *concrete* and *abstract*. Consider the three symbols that appear in the fifth situation of C14's proof (Figure 3). The righthand symbol is small and cubic, and obviously enough, it depicts a small cube. The central symbol is a small paper bag; however, it doesn't depict a small bag, but rather an object of known (small) size and location, but of unknown shape. It is an abstraction device, in that a picture containing it abstracts over three possible situations, corresponding to the three possible shapes the object could be (cube, tetrahedron or dodecahedron). The lefthand symbol is a cylinder sporting a question mark; it doesn't depict any sort of cylinder, however, but an object of unknown size or shape. Like the paper bag, a cylinder allows us to abstract over several situations. A question-marked cylinder in fact abstracts over nine situations in total. Although not shown here, symbols can also be removed from the checkerboard, and placed by its side, in order to abstract over many possible situations, corresponding to the possible locations the depicted object could occupy. We may therefore contrast concrete symbols (like the small cube) with abstract symbols (like the small bag, or the cylinder). The latter function as graphical variables, more or less.

It should be obvious that some variables are more abstract than others: the question-marked cylinder specifies less information than the small paper



```
  ◈                                              ✓ Given
▪ ∀v (Small(v) → ∃y (Small(y) ∧ Cube(y) ∧        ✓ Given
  Adjoins(y, v)))
▪ ∀v (Medium(v) → ∃y (Medium(y) ∧ Cube(y) ∧      ✓ Given
  Adjoins(y, v)))
▪ ¬∃x Large(x)                                   ✓ Given
  ◈                                              ✓ Assume
    ◈                                            ✓ Assume
    ◈                                            ✓ Apply
      ◈                                          ✓ Assume
      Assumptions Verified                       ✓ CTA
▸   ◈                                            ✓ Assume
    ◈                                            ✓ Apply
      ◈                                          ✓ Assume
```

FIGURE 3 Use of graphical abstraction symbols.
The situation being viewed is the fifth in the course of the proof,
and contains three symbols of varying degrees of abstraction.
The lefthand symbol denotes an object of unknown size or shape;
the middle symbol denotes an object of known size but unknown shape;
and the righthand symbol denotes an object of known size and shape.

bag. Graphical situations containing many highly abstract symbols will, in turn, seem globally more abstract than those which contain symbols which abstract less, or not all. This being so, C2's and C14's proofs can be categorised in terms of the 'abstractness' of the graphical situations they contain. Scanning through the situations, it seems as if C14's are generally more abstract; each contains the same number of graphical symbols as C14's; but in many cases, the symbols are abstract ones, and often highly abstract ones (such as the example in Figure 3).

This first impression of 'extra abstractness' in the case of C14 can be made slightly more precise, by attaching an *index of determinacy* to the graphical situations in the proof. In essence, we can give each graphical symbol in a situation a score: for each attribute (size, shape, location, and label), a symbol scores 1 if that attribute is specified, and 0 otherwise. By totalling the scores, we can give the situation an index: modulo the number of symbols, the higher the score, the more determinate the situation; the lower the score, the more abstract. So, in Figure 3, the lefthand symbol scores 1, the middle symbol 2, and the righthand symbol 3. As a result, the situation as a whole scores 6. Now, if we score each of the situations in each of C2's and C14's proofs, this is the pattern of scores which emerges:

C2 9 9 9 9 9 9 9 9 9 9 4

C14 4 6 7 9 6 7 9 5 4 6 7 7 5 4

Taking the means of these figures, we see that C2's situations have a mean determinacy of 8.55; C14's have a mean of 6.13. There is thus a substantial quantitative gap, which vindicates the initial impression that C14 is entertaining more abstract situations.

3.4 Case studies: summary

So, in response to Question 4, C2 constructs a shorter, shallower, faster proof, containing Assume–CTA rule cycles, and entertaining situations with relatively little abstractness. C14 constructs a longer, deeper, slower proof, with fewer Assume–CTA cycles, but more Exhaust–Merge–Inspect cycles, and entertaining situations with rather more abstractness.

These cases raise at least three types of question: (i) Are the differences between C2 and C14 accidental, or do they represent differing cognitive styles that we may attribute to groups of individuals? (ii) We have stated that Question 4 is of an indeterminate type; do the kinds of differences we have pointed to also occur on more determinate questions? (iii) Are there underlying reasons linking together the properties in, say, C14's proof?

To answer questions (i) and (ii), we must first describe our experimental setup, and show how we found there to be two distinctive groups among the Hyperproof subjects; we can then indicate in section 5 how the proof style differences we noted in our cases studies appear to be systematically related to these differences in cognitive style. To answer question (iii), we discuss in section 6 how the properties we have drawn attention to may be explained given our cognitive theory.

4 Hyperproof Experiments: Method

Two groups of subjects were compared; one group ($n = 22$) attended a one-quarter duration course taught using the heterogeneous reasoning approach of Hyperproof. A comparison group ($n = 13$) were also taught for one quarter but in the traditional syntactic manner supplemented with exercises using a graphics-disabled version of Hyperproof (to control for the motivational and other effects of computer-based activities). A fuller description of the method and procedure is provided in Cox et al. 1994.

All subjects were administered two kinds of pre and post-course paper and pencil test of reasoning. The first test was of 'analytical reasoning' and contained two kinds of item derived from the GRE-type of scale of that name (see for example, Duran et al. 1987). We refer to this test as the 'GRE' test. The first GRE subscale consists of verbal reasoning/argument analysis. The other GRE subscale consisted 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.

The second paper and pencil test we term 'Blocks world'. This test requires reasoning about blocks-world situations like those used in Hyperproof, but is couched in natural language rather than first order logic.

Both groups also sat post-course, computer-based Hyperproof exams. The exam questions differed for the two groups, however, since the syntactic group had not been taught to use Hyperproof's systems of graphical rules. The four questions set the Hyperproof group, though, contained two types of item: determinate and indeterminate. Figures 2 and 3 illustrate Question 4, one of the two indeterminate questions. 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. The results reported here are based on analyses of those protocols.

Scores on the determinate subscale of the GRE test were used to classify subjects within both Hyperproof and syntactic groups into DetLo and DetHi sub-groups. In other words, the score reflects subjects' facility for solving a type of item that often is best solved using an external representation. DetHi and DetLo subjects in the Hyperproof and syntactic groups responded differently to traditionally versus heterogeneously taught courses; those results are reported in Cox et al. 1994).

Here, however, we are concerned with comparing proof-style differences on the exam questions between DetLo and DetHi subjects' within the Hyperproof group only.

5 Systematic proof style differences

We can now return to questions (i) and (ii): are the differences in proof style between C2 and C14 accidental?; and are they as dramatic on determinate questions as on indeterminate ones?

The answer to question (i) should by now be apparent. C2 is a fairly typical DetLo subject; C14 is a fairly typical DetHi subject. When we take together the performances of all the Hyperproof subjects on the four exam questions, we can uncover some significant results relating to proof parameters, rule usage and use of graphical variables. These results go some way towards showing that the differences between DetLo and DetHi subjects reach down into their styles of proof. Regarding question (ii), it also seems that the differences in proof style are more pronounced in indeterminate exam questions (2 and 4) than in the determinate questions (1 and 3).

5.1 Structural aspects of proofs

Preliminary analyses were performed on several parameters of these examination proofs. Each proof-log was coded for score (number of proof goals

TABLE 2
Mean parameters of exam proofs:
DetHi/Lo subjects within Hyperproof group.
Score is number of goals achieved (out of 3 in Q1–3; out of 4 in Q4);
Time is in minutes; Steps is steps in overall proof; Depth range is 0–3.

			Score	Time	Steps	Depth	n
DetLo	Det	Q1	3.0	8.42	14.7	0.22	9
		Q3	2.7	18.10	17.2	0.22	9
	Indet	Q2	3.0	11.60	18.6	0.11	9
		Q4	3.5	19.70	27.8	0.56	9
DetHi	Det	Q1	3.0	6.64	17.6	0.23	13
		Q3	3.0	15.14	21.3	0.23	13
	Indet	Q2	2.8	15.23	16.0	0.23	13
		Q4	3.8	20.10	28.6	0.85	13

validated), time (time spent on proof), number of proof steps and the proof depth (the depth of nested subproofs the subjects used in their solution). Table 2 shows the mean proof parameters for DetHi and DetLo subjects within the Hyperproof class. There thus seems to be a tendency for DetHi subjects to produce longer, more accurate, and more nested proofs than their DetLo counterparts.

Comparisons between DetLo and DetHi subjects were not statistically reliable, due to wide variation between subjects within the groups. However, taken together, the *pattern* of proof parameters shown in Table 2 suggests superior proof development strategies on the part of Hyperproof DetHi subjects.

In the case studies, we observed that C2's proof in Question 4 was shorter, shallower, and faster. The only uncharacteristic fact about this proof, then, is that a DetLo subject constructed their proof more rapidly than a DetHi; C2, in fact, was one of the fastest DetLo subjects on this question, while C14 was one of the slower DetHi subjects.

5.2 Patterns of rule use

For the analyses, rule use frequencies for the two indeterminate questions were added and frequencies for the two determinate questions were added.

A two-factor ANOVA for subjects (DetHi, DetLo) and item determinacy (determinate, indeterminate) was conducted separately for each of the rules. The results of these analyses revealed that all subjects used the following rules significantly more frequently[2] in developing proofs for the two indeterminate questions than for the two determinate questions: Assume, Apply and CTA. The Close rule was used significantly more on the *determinate* than on indeterminate questions. A two-way interaction was significant in one of the analyses: the Apply rule was used more on deter-

[2] As evidenced by significant main effect for determinacy factor.

minate questions by DetLo subjects than by DetHi subjects. Conversely, on indeterminate questions, DetHi subjects used it more frequently than DetLo subjects.

Cluster analyses of the rule use patterns of DetLo and DetHi subjects was also used as an initial exploratory method. These reveal *correlations* between rule uses and suggest the following observations. First, in general, DetLo subjects seem to make CTA a more central part of their rule repertoire than do DetHi subjects, who exploit Exhaust more centrally. Secondly, DetLo subjects seem to have a more stable set of relationships between their rules; the only rule which seems substantially less central for them on indeterminate questions is Close. Thirdly, the more flexible DetHi subjects may use CTA on indeterminate questions more frequently than on determinate questions, but the rule does not correlate closely with their other central rules. By contrast, Apply, and Inspect do seem central, on indeterminate (but not determinate) questions. Finally, like DetLo subjects, DetHi subjects use Close less frequently and centrally on indeterminate questions; however, it remains well correlated with Observe, which one might therefore conclude is also less central a weapon on indeterminate questions.

We observed that C2 used Assume–CTA more frequently and repetitively than C14. C14 used Assume–Apply, and Exhaust–Merge–Inspect more frequently. We can now see that these differences are indeed characteristic of the groups as a whole. We can also see that DetHi subjects, such as C14, have more flexible strategies, and appear to resemble their DetLo colleagues more on determinate questions than on indeterminate questions, like Question 4.

5.3 Treatment of graphical variables

We used determinacy indices to show that C14 is entertaining more abstract situations than C2, on Question 4. Using the indeterminacy index scoring method described in Section 2.3, we can derive scores for all the DetLo and DetHi subjects. So far, we have derived these scores for one of the determinate questions (Question 1) and one of the indeterminate questions (Question 4).

Considering Question 1, all subjects in both the DetLo and DetHi subgroups proved all three proof goals. The index of determinacy scores for DetHi and DetLo subjects proofs did not differ significantly. The mean index of determinacy score for DetLo subjects was $17.6, SD = .31, n = 9$. For the DetHi subjects, the mean was $17.7, SD = .25, n = 13$.

Considering Question 4 (indeterminate), two subjects (one DetLo, one DetHi) did not succeed in proving all of the proof goals. Considering only the subjects who did succeed in proving the proof goals, a one-tailed t-test between DetLo and DetHi subjects index of determinacy scores reveals a significant effect ($t = 1.91, df = 18, p < .05$). The mean index of determinacy score for DetLo was $7.98, SD = .92, n = 8$ and for DetHi it was $7.13, SD = .98, n = 12$. The lower mean index of determinacy score for DetHi indicates more use of abstraction in the steps of the proof.

Thus some support is provided for the prediction, based on specificity theory, that DetHi subjects are more skilled in the deployment of graphical abstraction conventions during reasoning.

6 Conclusions

So, C2 and C14 represent two differing cognitive styles; and their differing proofs are characteristic of these styles. They differ in length, depth, patterns of rules used, and quantity of graphical abstraction. But what ties these differences together?

Indeterminate problems, such as Question 4, demand that subjects entertain multiple cases during the course of the proof. There are basically two ways of breaking into cases: one can exhaustively enumerate all the different cases, in a flat list-like structure (C2). Or one can impose a hierarchical structure on the cases, with sister subcases being derived from the same mother case by adding extra information (C14). The first strategy makes for shallower proofs, with repetitive patterns of rule use (Assume–CTA); the latter makes for deeper proofs, with characteristic 'case opening' (Assume–Apply) and 'case closing' (Exhaust–Merge–Inspect) sequences. Deeper proofs actually require more steps, because the intermediate levels in the hierarchical structure of the proof are made explicit. Deeper proofs with subcase structure also require abstract situations to act as superordinate cases; hence, there will be more graphical abstraction in the proofs of subjects who generate proofs with this sort of nested structure.

As we stated in the introduction, our theoretical work has emphasised the idea that graphical systems possess a useful property—over-specificity—whereby certain classes of information must be specified. The property seen as useful because inference with such specific representations can be very simple. We have also urged that real systems, like Hyperproof, do in fact allow abstractions to be expressed, and it is this that endows them with a usable level of expressive power.

From the case studies we have discussed, and from the empirical results which lie behind them, we can see that it is not enough for an inferential system simply to possess usable graphical abstractions. They must be *available* to the users of that system. DetHi subjects can exploit Hyperproof's graphical variables to create elegant proofs on indeterminate problems; DetLo subjects appear to lack the required competence with graphical variables, and so they attack indeterminate and determinate problems alike with roughly the same strategy.

The educational implications of this is are far from clear. Should all students be taught to use graphical reasoning methods, or should students be encouraged to follow their existing representational modality preferences? Much depends upon whether these prior cognitive styles are themselves responsive to educational intervention. Perhaps students should be encouraged to broaden their representational repertoires, before they encounter any formal logic teaching. For the time being, however, it seems that heterogeneous reasoning is bound to produce heterogeneous outcomes.

7 Acknowledgements

The support of the Economic and Social Research Council for HCRC is gratefully acknowledged. The work was supported by NATO Collaborative research grant 910954, and UK Joint Councils Initiative grant G9018050. The first author is supported by an EPSRC Advanced Fellowship. Special thanks to John Etchemendy, Tom Burke, and Mark Greaves.

References

Barwise, Jon, and John Etchemendy. 1994. *Hyperproof.* CSLI Lecture Notes Number ?? Stanford: CSLI Publications.

Cox, Richard, Keith Stenning, and Jon Oberlander. 1994. Graphical effects in learning logic: reasoning, representation and individual differences. In *Proceedings of the 16th Annual Conference of the Cognitive Science Society*, ??–??

Duran, Richard, Donald Powers, and Spencer Swinton. 1987. Construct Validity of the GRE Analytical Test: A Resource Document. Research Report 87-11. Princeton, New Jersey: Educational Testing Service, April 1987.

Stenning, Keith, and Jon Oberlander. 1991. Reasoning with Words, Pictures and Calculi: computation versus justification. In *Situation Theory and Its Applications, II*, ed. Jon Barwise, Jean Mark Gawron, Gordon Plotkin, and Syun Tutiya. 607–621. CSLI Lecture Notes Number 22. Stanford: CSLI Publications.

Stenning, Keith, and Jon Oberlander. in press. A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science* ??:??–?? Also Research Report HCRC/RP-20, Human Communication Research Centre, University of Edinburgh, April 1992.