Journal of Experimental & Theoretical Artificial Intelligence
Vol. 18, No. 2, June 2006, 249–266

Heterogeneous reasoning in learning to model

KEITH STENNING*† and MELISSA GRESALFI‡

†Human Communication Research Centre, Division of Informatics, University of Edinburgh, 2 Buccleuch Place, Edinburgh EH8 9LW, UK
‡Cognition and Learning Laboratory, CERAS 105, Stanford, CA 94305-3084, USA

(*Received in final form 26 July 2005*)

Conceptual learning in mathematics and science involves learning to coordinate multiple representation systems into smoothly functioning heterogeneous reasoning systems composed of sub-languages, graphics, mathematical representations, etc. In these heterogeneous systems information can be transformed from one representation to another by inference rules, and learning coordination is learning how and when to apply these rules. The study of heterogeneous representations in learning has had the benefit of focusing attention on the reality of representation in the 'wild'. We propose that the concept of heterogeneity of representation should be extended from multimodal (e.g. diagrammatic plus language) systems to multiply interpreted systems, even when those systems are apparently homogeneously linguistic. We proceed by analysing, from the perspective of the heterogeneity of reasoning, three learning incidents which happened in groups of students engaged in learning the mathematics and biology involved in modelling biological populations. We observe both learning successes and failures that cannot be understood without understanding the integrations of heterogeneous systems of representation involved and the inference rules and operations required to get from one to another. The purpose of presenting real incidents in some of their undomesticated detail is that they show what phenomena a homogeneous theory of reasoning would really have to account for. We argue that this type of rich naturalistic data makes implausible the instrumentality of any reconstruction in terms of a pre-existing fully interpreted homogeneous interlingua.

Keywords: Reasoning; Heterogeneous representation; Modelling; Learning; Real-world problems; Mathematics; Science; Conceptual learning

*Corresponding author. Email: K.Stenning@ed.ac.uk

Journal of Experimental and Theoretical Artificial Intelligence
ISSN 0952–813X print/ISSN 1362–3079 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/09528130600557838

1. Introduction

The coordination of multiple representations (e.g. diagrammatic and sentential systems) is frequently instrumental in conceptual learning (Barwise and Etchemendy 1994, Stenning *et al.* 1995, van Someren *et al.* 1999, Rinella *et al.* 2001, Stenning 2002). When diagrams and language are both involved, it is easy to see at least surface heterogeneity of representation, even if theoretical issues about its depth may remain. Perhaps all these representational surfaces are 'translated' into some homogeneous interlingua? We believe that one of the best antidotes to this kind of thinking is to appreciate the need for invoking heterogeneity of interpretation even in situations where apparently homogeneously interpreted language may appear to be the only representation in sight. Once one appreciates the heterogeneity of linguistic interpretation, the heterogeneity of diagrammatic and linguistic modalities appears to be merely another species. After all, it is the possibility of heterogeneity of interpretation which signals that care must be taken to adopt inference regimes appropriate to the reigning interpretation.

Even in deceptively simple laboratory reasoning tasks where it may appear that only homogeneous linguistic representations are in play, careful theoretical analysis and empirical observation reveals the richness and variability of the interpretation processes in which naive undergraduate subjects engage. One set of studies shows how apparently simple experimental materials in Wason's selection task (Wason 1968) in fact introduce contradictions in the subjects' initial interpretational attempts and thereby invoke highly variable efforts to induce consistent overall interpretations, yielding data that have traditionally been interpreted as revealing irrationality (Stenning and van Lambalgen 2001, Stenning and van Lambalgen 2004). Another heterogeneity of interpretation arises in systematic individual differences in quantifier interpretations which play out in subsequent syllogistic reasoning (Newstead 1995, Stenning *et al.* 1996, Stenning and Cox submitted). A group of studies shows how different logics (with different logical connectives and different concepts of validity) are required to model the different things that human beings do with natural language, even within a discourse—another form of heterogeneity (Stenning and van Lambalgen 2005). For example, credulous interpretation of a speaker's discourse requires a non-monotonic logic with a default interpretation of conditionals, whereas this logic cannot represent a blunt contradiction between the views of two participants.

We will call these kinds of heterogeneity of reasoning 'interpretational heterogeneity' to distinguish them from heterogeneity of modalities such as diagrams and languages. We believe that interpretational heterogeneity is more fundamental even if less visible than heterogeneity of modality. Heterogeneity of interpretation is what forces reasoning procedures to respect contextual variability.

However, these are reasoning studies in artificial laboratory settings, and one common reaction is to complain that this sort of nonsense does not go on in natural discourse. The purpose of this paper is to use some example episodes from project-based group learning to illustrate how the concepts of heterogeneous reasoning present themselves in the classroom setting in less formal contexts than Hyperproof. This investigation emphasizes how local the semantic interpretation of representations is in context. Words change meaning frequently and systematically, and the information they carry is moved into and out of other representation systems, both linguistic and non-linguistic. The investigation also provokes examination of the relation between heterogeneity and localness of interpretation (Moravscik 1998). With diagrams, it is usually quite evident to users that the diagram has a local interpretation and that the naive user needs to learn this local interpretation, even though there are regular features of such diagrammatic systems from use to use. With natural language, we are often so practised at making the contextual interpretation of its local semantics that it is easy to fail to realize that this is what we do.

In particular, learning mathematics and science concepts involves learning to coordinate multiple formalisms (numerical, graphical, algebraic, terminological), but it also involves learning how to apply formalisms in contexts. It is all too possible for students to succeed at the first and to fail at the second—to learn the internal operation of some formalism without learning how to apply it to new problems. Barwise and Etchemendy (1994) have used the term 'heterogeneous reasoning' for reasoning using multiple coordinated representations, and have applied heterogeneity of representation in order to improve students' grasp of the application of formalisms in computer environments such as Tarski's World and Hyperproof. Another related curriculum response to this problem of teaching the application of formalisms to real-world problems has been a move toward project-based approaches which teach formalisms in close relation to their context of application, in particular teaching scientific concepts along with the mathematics that goes with them (Goldman *et al.* 1995).

What interest does any of this have for the researcher concerned with machine intelligence? The spirit of our enterprise is an existence proof of the heterogeneity of representation systems and their interpretations in the 'wild'. Our own concern is to reveal how much reasoning has to go on to establish local interpretations, and how many such local interpretations are involved in reconstructing even brief interludes of learning discourse. Although examination of isolated episodes cannot establish statistical trends, nor finally settle whether homogeneous interlinguas can be proposed, we believe that at the current state of development, rich natural examples can provide important plausibility for claims of heterogeneity, and show just how much work those who claim homogeneity will have to do.

A further aim is to coordinate this semantic approach with others which focus on discourse practices and students' recruitment of material from their diverse experiential worlds. Stenning *et al.* (2002) analyse these and other incidents from the same curriculum intervention from the additional points of view of the group dynamics and the adoption of social roles. The purpose of this aim is to show that a representational approach which does justice to the richness and subtlety of real-life interpretations is not as incompatible as is generally thought with approaches more usually associated with 'distributed cognition'. The advantage of pursuing several parallel analyses of the same data for cognitive theory may share something with the advantages of the project-based approach for the students. Applying several kinds of the theory to the same episodes turns up new questions about how the theories relate to each other, and thus may induce conceptual learning and improved ability to apply the theories in novel circumstances.

1.1 *Heterogeneous reasoning*

Theories of human reasoning have begun to pay more attention to how representational systems are selected or constructed, and the variety of systems that may be used in solving a single problem, rather than conceiving of reasoning as a system-internal activity. Barwise and Etchemendy (1994) have called this use of multiple coupled systems of representation **heterogeneous** reasoning, and have developed several computer environments for teaching it. For example, Hyperproof presents a graphical window containing diagrams of a blocks-world inhabited by regular solids on a chequerboard, and a sentential window containing first-order logic sentences. The proof rules of the heterogeneous system incorporate the inference rules of the conventional sentential calculus1, augmented by rules for moving information between diagram and sentences, in both directions. For example, the user can *observe* a feature of the diagram as the basis for inferring a sentence, or may *apply* information from a sentence by inferring (and constructing) a new feature of the diagram. **Observe** and **apply** are two (of about half a dozen) of the heterogeneous inference rules which coordinate the diagrammatic and sentential representations into a heterogeneous system. Figure 1 shows an example application of the rule **apply** which applies information from the highlighted sentences to infer a label in the diagram.

Fundamental semantic distinctions between how diagrammatic and sentential representation systems express abstraction have been shown to play an essential part

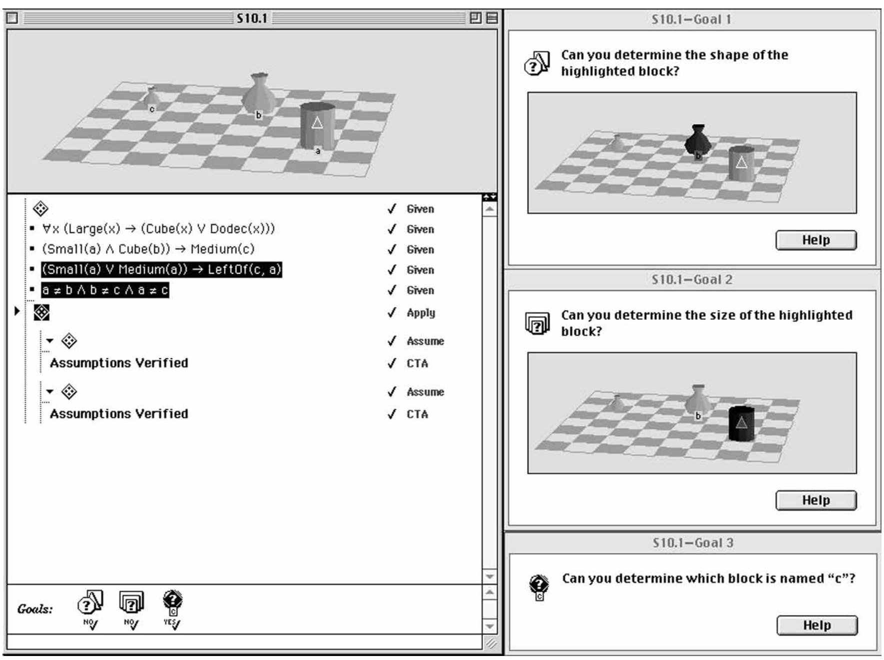


Figure 1. An example of a Hyperproof problem. The graphical givens of the problem are in the upper-left window and the sentential givens on the bottom left. The goals are set at the right. The rule **apply** is invoked using information in the highlighted sentences to draw the inference that the cylinder icon is labelled a.

in analysing the learning that occurs as students master the construction of proofs in Hyperproof's heterogeneous environment, allowing the learning to be characterized as the learning of strategies of representation selection and use (Oberlander *et al.* 1999, Stenning 2002). Whether students benefit from the diagrammatic facilities of Hyperproof is determined to a great extent by their facility at grasping helpful strategies for using Hyperproof's expressions of graphical abstraction.

Hyperproof reveals an important property of representational systems in use. The semantics, both of its sentences and its diagrams, are partially interpreted, i.e. the system has some of its meaning fixed while other parts are defined in episodes of reasoning. This contrasts with the usual introductory presentation of first-order logic as an entirely uninterpreted language. In practice, even predicate calculus is usually understood as partially interpreted in that its logical 'constants' are taken to have a fixed meaning except where 'non-standard' interpretations are entertained. In Hyperproof, even the predicate and relation symbols of the first-order calculus have to be given a partial local interpretation in terms of spatial relations because the predicates and relations have to coincide with those of the diagrams. Nevertheless, elements such as the individual constants a, b, . . . take their interpretation anew in each proof.

Thinking about the correspondence of partly diagrammatic systems of representation like Hyperproof reveals the need for coordinating diagrammatic and sentential representation systems, but leads to the further realization that in situations of real language use, the apparently homogeneous languages utilized are in fact often heterogeneous in the fundamental sense that many schemes of interpretation are in play at once. (We will use 'sentential' throughout to describe linguistic representations such as calculi and natural languages, not to discriminate propositional from first-order logic.) Even when natural language is the only modality, the reasoning systems in operation must be thought of as heterogeneous because the apparently single language can only be understood as being composed of overlapping language fragments, each constituting a distinct (if closely related) system of representation. We are, of course, accustomed to examples of blatant equivocation: 'Things with wings fly. This building has wings. Therefore, this building flies'. What we are after here are much more subtle cases where we have to accommodate shifts of interpretation, often without noticing that we have accommodated anything at all.

To illustrate this point, this paper takes some classroom interactions of a group of students learning to model biological populations in terms of mathematical functions and analyses the multiple partially interpreted representation systems which are in play. The students' representational resources and activities include at least the following: worksheet filling, graph drawing, computer operation, calculator use, group speech and gesture, reference material, and teacher interventions. The importance of taking real rich material for exemplification, in preference to constructed examples, is that truth is much stranger than fiction. It would be hard to invent the kinds of complexities observed here, and easy not to take them seriously if one did invent them. Episodes of conceptual learning are a good source of material for theoretical analysis because the misunderstandings that arise when the user's conceptual system is under strain are clear evidence of the interpretational adjustments that otherwise remain beneath the surface.

The educational issue in focus is learning about modelling, and particularly learning about the process of formalization and interpretation. A recurring theme is the struggle to coordinate formalism internal operations (calculation) and formalism external correspondences (semantics). We will analyse both successes and failures of coordination.

To summarize, the question that has dominated debate about heterogeneous representations is whether the best account of reasoning with them is one in which superficial differences are removed by translation into an interlingua on which uniform reasoning can take place, or whether the differences between components of the heterogeneity have to be accommodated by having rules of inference which operate on combinations of information, without there being any homogeneous representation system. Our proposal is first to broaden the notion of heterogeneity to include systems which have homogeneous surfaces (say, completely linguistic, or completely diagrammatic) but which encompass contrasting interpretations of overlapping fragments. If we can establish that this extension is necessary to analyse rich naturally occurring episodes of conceptual learning, that goes some way to providing evidence that heterogeneous systems have to be taken as more than superficially heterogeneous. By hypothesis, contrasting interpretations of the same surface forms require us to accommodate our inference patterns. We also believe that our naturalistic observations of rich systems are suggestive about why interpretational heterogeneity is ineradicable.

2. The educational setting

The data we have analysed come from an 8th grade Middle-school Mathematics through Applications (MMAP) classroom in the San Francisco Bay Area (Goldman *et al*. 1995). The purpose of MMAP is to have students use mathematics to address real-world problems, often with the assistance of computer applications. In the approximately 30-day unit we will discuss, called Guppies, students created mathematical models of biological population growth. As part of this unit, students were to learn both about how to construct mathematical models of population growth and about the exponential functions that underlie them. Our analyses focus on a group of students (Manuel, Lisa, Kera, and Nick) whose improvement on pre/post-assessments placed them about midway in learning of the half-dozen focus groups videotaped by Rogers Hall and his colleagues (Hall 1999) during this unit in a variety of classrooms. These students are chosen to reflect roughly average performance for the class. For more information about the design of the study, see Hall (2000).

2.1 *Three learning incidents*

The following three incidents were chosen from videotapes because they illustrate both successful and problematic learning episodes. The initial incident from the pre-test phase sees the students make at least part of one of the fundamental conceptual discoveries of this field, namely Malthus' equation—population growth has a recursive characteristic that leads to exponential growth if unchecked.

The second incident, from the body of the course, is of interest because it contains an attempt to diverge creatively from the structure of the assigned worksheet by taking a short cut in the calculation. On the one hand, this divergence reveals the germ of another important insight—that functions can be composed. However, in the circumstances, the insight is not fully worked out and leads to error and confusion.

The third incident is chosen to illustrate that the confusion that is not resolved in the previous incident appears to persist into the much later post-test phase of the course. It consists of another attempt to calculate a birth rate for a new modelling problem.

In all of these incidents, the students struggle to coordinate multiple representations. We examine some of their attempts at coordination in detail, seeking to reveal how some episodes are successful and others are not. The transcriptions are compressed for the reader's convenience by leaving out material which does not relate to our analysis. The line numbering gives some impression of this procedure.

2.1.1 Pre-test insight: 'babies have babies'. When the group discovers the recursive nature of population growth, they are engaged in constructing a model of a mouse population. They have obtained an initial number of 20 adults in the starting population from the worksheet, and estimated a birth rate of four per pair. They are now calculating what the population will be after eight breeding seasons. The group initially adopt a linear model implicit in multiplication of a fixed birth rate. Only when they turn to the graphing activity dictated by the worksheet do they begin to think of the process which the calculation is intended to reflect—the semantics of the arithmetic formalism.

60: M. so there's . . . equals 40 babies each season

65: M. it's three hundred and twenty

66: K. (inaudible) is that including adults?

67: M. no, three hundred and twenty plus twenty

69: M. by the end of the winter

70: M. three hundred and forty mouse . . . mice . . . mices. OK.

73: M. Now we need to make a graph of it

.

.

.

82: M. so let's see . . . the first season is over here (making a mark on the graph)

83: L. ×××××× wait a minute

86: M. and then sixty plus is going to be a hundred

89: L. wait a minute its forty (gestures a triangular shape) OK its forty right?

90: L. and then you have to pair those up (brings hands together) and then they have kids (spreads hands apart, while K and M look at her confused)

92: M. oh yeaaah (embarrassed, laughing at himself) we were doing it . . .

94: L. That's a lot of mice

95: K. gosh that's a lot of nasty mice.

The interchange on lines 65/66 is an example of the frequent need to coordinate numbers with their semantics—adults still have to be included in the population, and 'three hundred and twenty' is the number of babies in eight seasons just calculated. Similarly line 69 is a further reiteration of the semantics of the number 'three hundred and twenty plus twenty'—the number represents a population at a time. Line 73 is an appeal to the authority of the worksheet for what has to be done next. What is interesting about this introduction of a new representation (the graph) is that it appears to be what triggers the new thinking that reveals the error (adopting the linear model) that they have all made. M makes a mark of sixty on the Y axis at the origin representing the starting population. But L has realized that something is wrong (line 183). M continues calculating the next graph point. But L persists. She starts by reiterating the number and asking for acknowledgement of it (line 189). The number is the number of first season babies. She then states that these have to be paired up, and themselves reproduce (line 190). The gesture is intuitively an important part of her communication that she has a new insight, both for herself and for the group. M fairly rapidly sees their mistake too. They all realize that this is going to make the growth of the population much more rapid, although they do not have any number for it. They immediately refer back to the experiential world of 'nasty mice'. Perhaps the reality of reproduction lies behind the affective tone of the incident. It was not just a mathematical mistake, but a failure to apply the 'facts of life'?

The original adoption of the linear model arises within the 'mathematical world'. It is, in some sense, the obvious calculation to do—forty babies a season for eight seasons is going to give 320 babies. After all, multiplication is something we learn so as to avoid having to do multiple additions. It is not until the graphing acitivity makes them break this calculation down into a series of calculations, and makes them think of the process of seasonal breeding, maturation, etc. that L sees the error. She thinks about what happens in the world of mice—about the semantics. Her insight is adopted rather rapidly with little overt acknowledgement.

2.1.2 An attempt at creative construction: 'discovering function composition'. When the group brushes up against function composition, they are constructing one of their early models of a population of guppies. They have a worksheet entitled *Building the Birthrate* which gives them a procedure for calculating, or recording from reference sources, the various parameters of the situation (brood size for different ages, birth rate, survival rate). Parts of this worksheet and the computer interface are condensed in figure 2.

Building the Birthrate

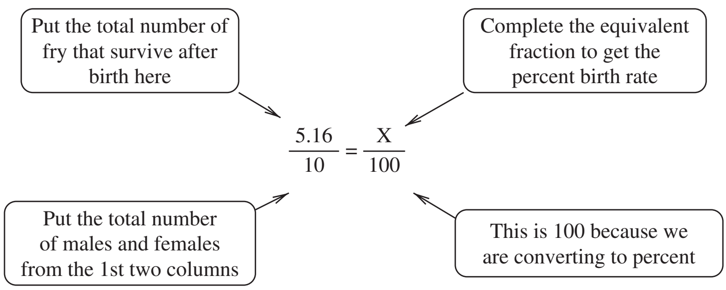
Step 1

Age	# Males	# Females	# Fry	Total
Young	2	1	4	4
Mature	4	2	50	100
Old	0	1	0	0
Total	6	4		104

Step 2 What percent of fry born survive? What happens to the ones who don't make it?
5% of fry survive. They are eaten.

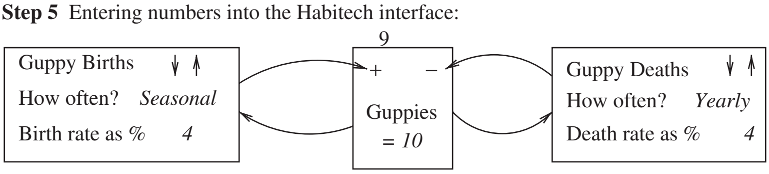
Step 3 Use this survival percent and the total number born to calculate the number that survive.
5.16

Step 4 So what's the birthrate? Now that you have calculated an assumed number of fry that survive past birth, you need to convert this into something that Habitech can use as a birth rate. As you know, Habitech works with percents or constant numbers. You will be using a percent birth rate.



BASED ON YOUR ASSUMPTIONS YOUR BIRTHRATE IS 4%
Congratulations! Now take this birth rate and the death rate you will use and head to Habitech to make your model. Remember this birth rate is based on certain assumptions. If you change an assumption, it will affect your model.

Step 5 Entering numbers into the Habitech interface:



Recording of Models

Initial #	Birth rate %	Death rate %	Years	Descr.
10	4%	4%	2 year	<13

Figure 2. Parts of the worksheet and the computer interface.

The worksheet has its own sequence of activities, although it should be noted that this is not the sequence in which this group performs them. The worksheet's first table (see figure 2) implements the calculation of the total population births in a season from data from reference sources. At step 2 the percentage survival rate is entered from a reference source, and at step 3 it is applied to the total from the table to give a number surviving. The lines represent page breaks in the work booklet. Step 4 then converts the total surviving fry into a percentage birth rate for the computer. The relevant part of the interface appears next. The bottom table of the figure (over the page on the worksheet) keeps track of the model, and will hold several trial models later on.

The group's sequence of work is actually to start by fulfilling steps 1, 2, and 3, followed by entering the result into the computer and recording the model. Step 4 is circumvented initially and is only filled in retrospectively the next day. The incident opens with M proposing to take a short cut in the calculation. This is at first taken by L to be a mistake. She requests an explanation and receives one that she finds satisfying. However, she appears to appreciate that there is consequent book-keeping which needs to be taken care of, but fails to deflect the group from continuing to the entry of data into the computer model.

444: M. hey wait wait wait . . . no but listen. If 4% of the frys survive why don't we just forget about the fry survival and just put that amount for the, for how much are born . . .

445: L. because the number born are not how much survived

446: M. yes. yes, the ones who survive are the ones we count, not the ones who are dead because we don't make room for the ones that are dead

453: M. OK you know how 4% the whoooole fry who were born survive so why don't we just put 4% on the guppies birth because that's how many are going to survive

454: L. I get what you're saying because why put however many more guppies in when they're just going to die anyway?

455: K. so why not just put 4% because that's how many are surviving/that's how many we're going to count?

497: L. but what's that 4%?

498: K. the ones that survive

499: M. The ones that actually survive fryhood

501: L. Yeah, I know, but how many of the guppies are 4%?

502: M. we don't know, we'll let that mechanical thing work and tell us.

At 444, Mopens with a proposal to collapse two stages of calculation into one. In fact, this proposal is perhaps something akin to what is embodied implicitly in the worksheet, and is potentially a creative proposal embodying a concept rather close to one of the core aims of this curriculum—the understanding of mathematical functions and their composition. M is proposing to compose two functions into a single function taking the argument of the first and the value of the last. L objects to this proposal and justifies her objection by pointing out that 'the number born are not how much survived'. In fact, we will see that in the terminology of the worksheet, the number of fry surviving expressed as a percentage of the whole population is the birth rate, which plays its part in this confusion. M appears to understand the objection and explains his proposal's departure from the worksheet with some success. L accepts the sense of the innovation even though she expresses reservations about its coordination with the worksheet. The activity is turned over to the superior calculating powers of 'that mechanical thing'—the computer program Habitech.

Unfortunately, the ''mechanical thing'' does not understand the creative proposal; L's reservations are well motivated but, because she lacks a clear understanding herself, her intervention does not deflect the group (see Greeno *et al.* (2000) for further analysis). There are numerous problems of coordination between the representations in figure 2. The survival rate of 5% at step 2 is copied into the model table as 4% (possibly a memory error, or a correction later). However, the serious error is short-cutting the calculation at step 4 and entering the 4% rate directly into the birth rate box at the end. The algebraic ratio part at step 4 is returned to only later next day when trying to comply with having the whole thing filled in.

What has gone wrong as the group struggles with the welter of representations and numbers? It is hard to give a crisp interpretation of a murky confusion, but we can suggest some of the contributing factors. An important source may be a divergence of the ordering of biological events and the calculation events that refer to them; another is the terminology. In the fish world, fry are born, the vast majority are eaten, and at the end of the season they are counted. In the calculation world, first the number of births are calculated, then a survival rate is applied, and a census number of surviving fry results. So far so good. However, turning the page after step 3, and after recording model parameters on the next page, the students arrive at a further calculation of the 'birth rate', where 'birth rate' now means 'birth-and-survival-to-year's-end rate'.

Therefore, at step 1, the *birth rate* is a set of numbers representing the brood size of the average guppy at different ages (i.e. the numerals 4, 50, 0); at step 2, the *birth rate* is the number of fry born to the whole population (i.e. the numeral 104). In steps 3 and 4, the *birth rate* is the birth-and-survival-to-end-of-season rate expressed as a percentage of the whole population (i.e. the numeral 4). The same idea, a very tangible idea, namely the rate of births, is represented each time by a number, but each time the number counts something different, and complex calculations constitute the inference rules which 'move the number from box to box'.

Unfortunately, M's insight that two functions can be composed requires attendant housekeeping to keep the ontology straight. Perhaps a contributing factor is that because the pre-survival birth rate in step 1 is never put into the form of a percentage (1040%), M does not appreciate that, after step 3, it has already been implicitly composed with the survival rate, and the calculation at step 4 is intended only to return to a percentage form. Unfortunately, the terminology exacerbates this problem of 'backward causality'—first calculating a survival rate (using births) and then calculating a birth rate from that figure.

2.1.3 Post-test—the persistence of a confusion. We now present an incident from the post-test in which the group displays evidence that the episode of confusion just described has not been fully resolved. Although in the intervening couple of weeks the group has made good progress in understanding population models (see Hall 2000), it is of some concern that the particular confusions surrounding the derivation of birth rate from raw data appear to persist.

The group is working on the post-test problem of constructing a model for a mouse population preyed on by cats. This episode is from fairly early on when they are settling on a birth rate for mice and have not yet considered predation.

76: M. four, five or six? per adult?

77: K. If we're going to go four, five or six, let's go four.

78: L. actually, let's use five. Its four through six. Let's use five.

82: M. OK how do we find out the birth rate? (grabs a piece of paper) We do the . . . five is what we decided on. How many did we start out with (looks at the computer)

83: L. Twenty

86: M. I'm not sure that this is right (as he writes)

87: M. What's 500 divided by twenty?

88: L. What are you doing?

89: M. Finding out the birth rate

90: L. Oh yeah.

91: M. What's 500 divided by 20? (K hands him the calculator and M starts punching in numbers)

92: M. 25% I could have figured that out myself (K laughs; M goes back to the computer) 25% right? (enters it into the birth rate) and how many die?

Segment 82 illustrates the pervasive struggle with the semantics of numbers. M accepts that they will use 5 (babies per litter per season), which one might think is a birth rate, but in this context, 'birth rate' is a specific number that can be entered into certain boxes on the worksheet and the computer screen. They correctly appreciate that they do not have the birth rate, in this sense, and this is precisely where they had problems before. The number they seek is a percentage. At 87, M has implicitly multiplied the 5 by 100 and is now explicitly going about dividing by 20 (the number in the initial population). Not surprisingly, L does not understand where the 500 came from and asks for clarification, but receives only the description at the completely unhelpful level 'finding out the birth rate'. The problem is then accepted as a calculation problem, and the semantics is left unaccessed. Why should the number of babies in one litter divided by the total number adults in the population multiplied by 100 yield a percentage birth rate? The answer would appear to be based on some dim memory of the ratio formula from the episode analysed above (figure 2, step 4).

The group is content to continue to the next stage of the problem and does not question the reasonableness of the figure of 25%. This is testimony to the insulation of the numbers from what they mean. If each couple has five babies, the actual number is 250%. But the group does not discuss finding this number or acknowledge that adults have to be paired up. The group does not even apply the qualitative reasoning that since the parents are outnumbered by their babies, the birth rate must be more than 100%. Such qualitative inferences are only available if the numbers are treated as standing for something other than themselves, i.e. mice.

The group now tries to use this birth rate in a simulation using the interface shown in figure 3. This interface is testimony to the fact that mere substitution of words for other words is no problem for these students. The interface is fixed in always using the terms 'guppies' and 'caribou', but these examples merely stand for 'population modelled' and 'predator', respectively. In the case of this particular problem, these are mice and cats. Even when the model actually turns out to extinguish the mice in short order, the problem is not traced to the miscalculated low birth rate. It is all too easy for a problem to hide in a complex model. The whole point of models is that many parameters contribute to their outcome. However, this means that there are many possible culprits when the outcome is unacceptable, and debugging becomes a major problem.

3. Discussion

How does this material compare with Hyperproof which we used to exemplify the concept of heterogeneous reasoning? In Hyperproof there were a small number of named rules of inference which were heterogeneous rules (**apply** and **observe** were the

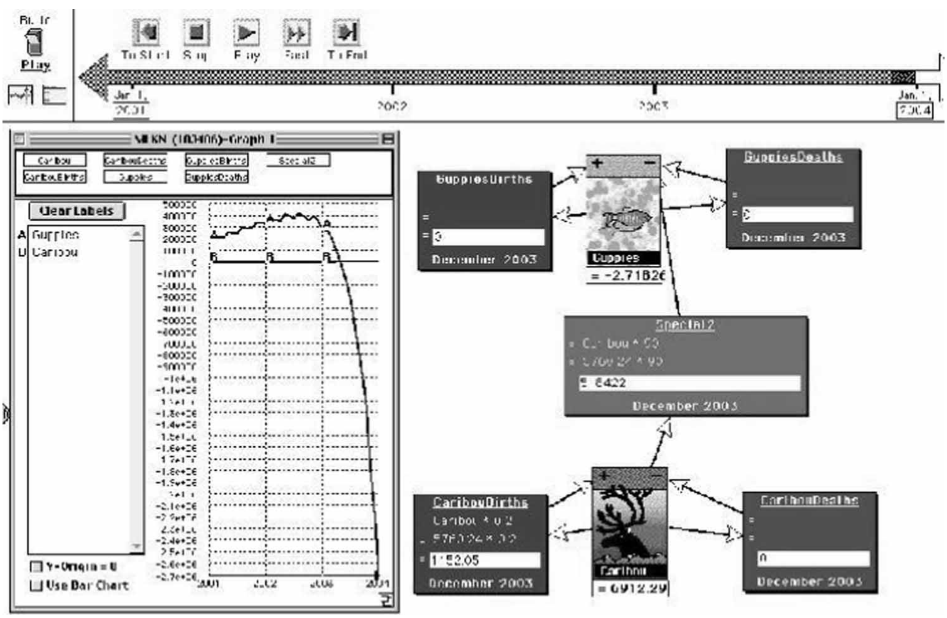


Figure 3. The Habitech interface. On the right are boxes in which various parameters of the model can be changed. On the left, the graph shows the sudden extinction event discussed in the text.

examples we cited) and it was possible to state exact conditions of application of these rules. In the workings of our group, the systems are less completely integrated, but nevertheless, many of the important features of Hyperproof carry over. It is still possible to state formal rules which govern the transformations of our example ‘birth rate’ from context to context, within the worksheet and between worksheet and computer simulation. Where ‘birth rate’ is known in terms of numbers of live-births-to-the-whole-population-in-one-season, it is possible to state exact formal transformations into ‘birth rate’ known in terms of live-births-surviving-to-year-end-per-mother, along with whatever other parameters need to be known to apply the transformation. Or, to take another example, it is possible to state the formal transformation from the parameter in the box labelled ‘caribou’ to the number of predator cats in the model currently loaded into the computer’s software.

It is not that the degree of formality of the rules is any less in this case than in Hyperproof. What may well be different is the account that has to be given of learning in the two situations. In Hyperproof it is at least possible to give an account in terms of learning a very small number of discrete rules (**apply**, **observe**, . . .) and their application. In the case of our group learning about population modelling, viewing what they have to learn as a set of discrete rules seems less natural and probably less insightful. We surely ought to think in terms of learning general concepts about applying alternative ‘units of measurement’ (sets, proportions, percentages, ratios, . . .) and the concept of a ‘horizon-of-survival’ applied to counting offspring (live births, survivals to first season’s end, . . .), along with the many other examples of modifiers on concepts which we could have used from these same episodes. However, even with Hyperproof, some abstraction like this is probably a more insightful way of thinking about what is learned, i.e. successful students learn some partly general concepts about what inference rules must be like to satisfy a reasoning system, and something about strategies of use, rather than simply the particular heterogeneous rules of Hyperproof. The same is true for learning homogeneous reasoning systems. If the student only learns specific rules such as modus ponens and &-introduction, and nothing about what rules of inference are like and what they do in general (and why), then nothing much of use will have been learned.

Of course, nothing by way of inferences about the causalities or even correlations between the kinds of events observed here can emerge from an analysis of these few isolated examples. Nor is redesign of a curriculum usefully based purely on analyses of single incidents. It is clear from other studies of this curriculum that it is highly effective. Indeed, this very group of students shows a considerable mastery of modelling at the post-test phase. The group repeatedly alters parameters of complex models (including not just birth and survival rates but also predation) in the qualitatively correct direction in response to over- or undershooting of the desired population outcome.

However, we believe that these analyses do make clear just what a sea of semantic complexities the group swims in. They are awash with representations, and the words involved have to travel from one representational system to another to achieve the problem-solving task at hand. As they travel, they change their meanings and their values. Birthrate is rarely the same thing on two occurrences. The whole system cannot be understood as anything other than heterogeneous, and the interpretations as anything other than highly local. The penalty of failing to accommodate these shifts of interpretation is just as severe as concluding from its possession of wings that this building can fly—just as severe, but far more likely to occur, because the equivocations involved are so much more subtle. Each of the birth rates is a quantity of mice (or fish) and mice all look dangerously alike.

If we were to go through the transcript spelling out after each occurrence of a numeral, the type of the entities it enumerates, we would wind up with some splendid and totally incomprehensible sentences. Simply spelling everything out is not to be recommended other than as a way of exposing complexity for the researcher. However, we cannot understand the students' problems until this complexity is exposed.

From a theoretical perspective, this may seem either banal or outrageous. Once we are fluent at the skills of transformation required for coordinating the subsystems of representation, the whole system appears to take on a transparency and homogeneity which is completely illusory. We cease to notice how the very same word means something quite different from occurrence to occurrence. Therefore we can either forget that the system is heterogeneous (and respond with outrage to the claim), or we can, as theoreticians, claim that there is nothing deep in the coordinations that are required (and respond with a yawn). The students do not have the luxury of mastery. For example, one of the banal consequences of the instability of the meanings of the numerals is that there is a huge memory load, as evidenced in the repeated mis-recalls of numbers from sheet to sheet of their workings.

We do not believe that there is any way out of the heterogeneity. Learning mastery of the coordination of representation systems is a requirement of learning mathematics and science, and probably most other things (see Stenning (2002) for a review of these arguments). However, what we can strive to do is to introduce both teachers and students to the quirks of the representational furniture they find themselves surrounded by—the ways that words' meanings wander as words travel about.

Our research experience in classrooms indicates that teachers are rather wary of taking an explicitly metalinguistic stance. They do not often point out the dangers of shifts in meaning of words during an argument. The critical-thinking lecturer warns students about equivocation—the same term being used with different meanings in different ocurrences in an argument—but only at college. Equivocation is treated as a fallacy, usually assumed to be eradicable, and therefore it may be thought that opportunities to commit it should be eliminable from well-kept classrooms. Our analysis in terms of heterogeneity and localness of interpretation strongly suggests that equivocation is not eliminable. We cannot use unique terms for every meaning, and should not if we could. Just imagine if we replaced each distinct meaning of birth rate with a distinct neologism in the transcripts above. The use of the same term (with shifted sense) is often essential to anchor the term to the shared concept as the details shift through its various guises. Perhaps signalling when this is likely to be a problem would help? And perhaps teaching teachers to detect the seams between systems that have become transparent for them is an important aim?

However, these observations from the classroom are just as important for theories of the semantics of representations. The conventional response to the kind of observations of language that we have made here is that everyone knows that natural language is 'ambiguous'. The need to avoid equivocation in reasoning cannot be all that is important about heterogeneity of reasoning. Heterogeneity, or so this argument goes, is exciting if a system contains language and diagrams. Here, the heterogeneity is on the surface and we have to ask how disparate elements can be made to correspond. However, the idea that natural language consists of many heterogeneous subsystems is generally resisted, and explained away as 'mere polysemy'.

There are at least two problems with this dismissal. The number of polysemous readings required is essentially infinite; to take our particular example, 'birth rate' could be contextualized to yield different interpretations in an open-ended number of ways, and so this is not polysemy in any conventional sense. There are not separate lexical entries in the dictionary for the different interpretations of 'birth rate' which we tracked. Secondly, the interpretation of one ocurrence of one word is systematically related to the interpretations of ocurrences of others. Words in these discourses do not function atomistically—they are part of subsystems of correlated interpretation. If 'birth rate' is construed one way, then its contrasting terms, such as 'death rate' and 'survival rate', will also be construed in correlated ways. We saw above that 'birth rate' and 'survival-until-the-end-of-the-year-rate' can actually express the same concept in different contexts, even if their occurrence in the same context is likely to be differentiated.

So, where do our original proposals about heterogeneous reasoning stand? We have actually said rather little about the diagrams (both paper and computer screen examples) which are involved in the reasoning episodes analysed here, although some of the words we have focused on actually ocurred within diagrams, tables, or other graphical representations. Instead, we have concentrated on variation in interpretation of the same word across ocurrences. Our argument has been that natural languages in these discourses have to be regarded as composed of mixtures of sub-languages, each with its distinct interpretation—what we have called interpretational heterogeneity. The importance of such heterogeneity for reasoning is that it guarantees that rules of reasoning do not apply uniformly across the language which occurs; wherever heterogeneity of interpretation exists, equivocation is an ever present danger.

What does this do to the argument that heterogeneity may only be skin deep—that human processing immediately translates surface variability into deep homogeneity? Well, in the language examined here we do not have surface variability ('birth rate' has the same surface wherever it occurs), and we do have deep variability of interpretation. Perhaps a sceptic bent on preserving homogeneity could still argue that this deep variability can be 'translated' into a uniform interlingua for homogeneous reasoning to proceed.

The first observation here is that this is a pretty stretched notion of translation—translating English into English in fact. The second is that this 'translation' will require access to all sorts of information about how the sub-language fragments map onto the world—that is, the semantic and pragmatic information which we as analysts (and the pupils too) had to resort to in order to understand what was meant and what understandings and misunderstandings were in play. For the claim that reasoning is really homogeneous to carry any force as a claim about reasoning processes it surely must carry some implication that the use of the uniform reasoning rules that operate on the interlingua is guaranteed by sameness (or translation invariance) of form of the representations. If it is only guaranteed by ensuring some eventual correspondence between the semantics of the systems, then there is only a kind of foundational homogeneity at stake. The claim becomes something along the lines that there is always an interlingua (perhaps different in each case) within which all the apparently heterogeneous information can be recast, which is then susceptible to homogeneous reasoning procedures. This sort of foundational homogeneity may be an issue for the foundations of mathematics, but it hardly seems to be an issue for artificial intelligence or cognitive science, which seek an account of the mental and social processes of reasoning and learning. Besides, the tide in the foundations of mathematics has been running in the opposite direction for the last 70 years.

Acknowledgements

We would like to acknowledge invaluable comments from and discussion with Randi Engel, MuffieWeibe, Jim Greeno, and Rogers Hall. We are grateful for their generous support in allowing access to their data gathered under an NSF grant to Hall. We also acknowledge fellowship support from grant GR #R000271074 from the Economic and Social Science Research Council (UK), and CSLI's support for KS's research at Stanford.

References

J. Barwise and J. Etchemendy, *Hyperproof*, Stanford, CA: CSLI Publications, 1994.

S. Goldman, J. Moschkovich and The Middle-school Mathematics through Applications Project Team, "Environments for collaborating mathematically: The Middle-school Mathematics through Applications Project", in *Proceedings of CSCL '95: The First International Conference on Computer Support for Collaborative Learning*, J.L. Schnase and E.L. Cunnius, Eds., Mahwah, NJ: Erlbaum, 1995.

J. Greeno, M. Sommerfeld and M. Weibe, "Practices of questioning and explaining in learning to model", in *Proceedings of the 22nd Meeting of the Cognitive Science Society*, L.R. Gleitman and A.K. Joshi, Eds., Mahwah, NJ: Erlbaum, 2000, pp. 669–674.

R. Hall, "Case studies of math at work: exploring design-oriented mathematical practices in school and work settings", Final report to the National Science Foundation, 1999.

R. Hall, "Working the interface between representing and represented worlds in middle school math design projects", in *Proceedings of the 22nd Meeting of the Cognitive Science Society*, L.R. Gleitman and A.K. Joshi, Eds., Mahwah, NJ: Erlbaum, 2000, pp. 675–680.

J.M. Moravscik, *Meaning, Creativity, and the Partial Inscrutability of the Human Mind*, Stanford, CA: CSLI Publications, 1998.

S. Newstead, "Gricean implicatures and syllogistic reasoning", *J. Mem. Lang*., 34, pp. 644–664, 1995.

J. Oberlander, P. Monaghan, R. Cox, K. Stenning and R. Tobin, "Unnatural language discourse: an empirical study of multimodal proof styles", *J. Logic Lang. Inform*., 8, pp. 363–384, 1999.

K. Rinella, S. Bringsjord and Y. Yang, "Efficacious logic instruction: people are not irremediably poor deductive reasoners", in *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, J.D. Moore and K. Stenning, Eds., Mahwah, NJ: Erlbaum, 2001, pp. 851–856.

K. Stenning, *Seeing Reason: Language and Image in Learning to Think*, Oxford: Oxford University Press, 2002.

K. Stenning and R. Cox, "Rethinking deductive tasks: relating interpretation and reasoning through individual differences", *Quarterly J. Exp. Psychol*., 59, 2006.

K. Stenning and M. van Lambalgen, "Semantics and psychology: Wason's selection task as a case study", *J. Logic Lang. Inform*., 10, pp. 273–317, 2001.

K. Stenning and M. van Lambalgen, "A little logic goes a long way: basing experiment on semantic theory in the cognitive science of conditional reasoning", *Cogn Sci*., 28(4), pp. 481–529, 2004. Available online at: http://www.hcrc.ed.ac.uk/~keith/WasonSelectionTask/cognition.pdf

K. Stenning and M. van Lambalgen, "Semantic interpretation as computation in nonmonotonic logic: the real meaning of the suppression task", 29(6), pp. 919–960, 2005. Available online at http://www.hcrc.ed.ac.uk/~keith/InterpretationandReasoning/suppression.pdf

K. Stenning, R. Cox and J. Oberlander, "Contrasting the cognitive effects of graphical and sentential logic teaching: reasoning, representation and individual differences", *Lang. Cogn. Process.*, 10, pp. 333–354, 1995.

K. Stenning, P. Yule and R. Cox, in *Cognitive Science Conference*, La Jolla, CA, 1996, pp. 678–683.

K. Stenning, J. Greeno, R. Hall, M. Sommerfeld and M. Wiebe, "Coordinating mathematical with biological multiplication: conceptual learning as the development of heterogeneous reasoning systems", in The Role of Communication in Learning to Model, M. Baker, P. Brna, K. Stenning and A. Tiberghien, Eds, NJ: Lawrence Erlbaum Associates, 2002, pp. 3–48.

M. van Someren, P. Reimann, E. Boshuizen and T. de Jong, Eds, *Learning with Multiple Representations*. Dordrecht: Kluwer, 1999.

P. Wason, "Reasoning about a rule", *Q. J. Exp. Psychol.*, 20, pp. 273–281, 1968.