How should we evaluate complex multimedia environments?

Shaaron Ainsworth

School of Psychology and Learning Sciences Research Institute,
University of Nottingham
University Park, Nottingham, NG7 2RD, UK

Abstract

Early research with multimedia environments questioned whether these environments are effective in supporting learning. More recently it has been acknowledged that this question should really be “under what conditions, for which learners, performing which tasks and expecting what outcomes is learning with multimedia environments effective?”. However, while the argument has become more sophisticated, the techniques for evaluating learning with multimedia environments have not always followed suit. The dominant approach at present involves factorial designs with novices as participants, learning something for a short period of time with outcomes tested by an immediate pen and paper post-test. In this chapter, I will review the positive aspects of this approach, but will argue that such an approach strongly limits the questions that can be answered. This will be illustrated by reference to a study with three students of population biology (two experienced and one novice) working with a multi-representational simulation for eight hours. The simulation collects detailed process data, and the learners’ interactions with a teacher and the software were videoed to provide thick descriptions of their experiences. I will argue that this approach reveals much about the processes of learning with multimedia and that a multiplicity of methods are needed to truly understand multimedia comprehension.

Introduction

In common with the introduction of other forms of learning technologies when multimedia learning environments were first becoming available, they seemed to promise a solution for the problem of how to teach in complex domains. Learners would be motivated to learn by novel forms of representations such as animations, videos, dyna-linked pictures and text, *etc*. Moreover, their understanding of the domain would be enhanced by the opportunity to interact with many forms of representations. Learners would identify which representations best revealed the particular aspect of the domain they were currently studying and by making connections across these forms of representation, they would come to understand the domain in a less superficial, more expert way.

Research on learning with multimedia and multi-representational software has shown that this rosy promise can be achieved (*e.g.* for reviews see Ainsworth, 1999; Najjar, 1998) but that it is not an invariant feature of learning with multimedia. For every study published showing that such environments facilitate learning, it seems that an equal number show that learners find such environments overwhelming and that it the worse cases, such environments are not just neutral but can even harm learning (*e.g.* Ainsworth, Bibby, & Wood, 2002; de Jong et al, 1998; Moreno & Mayer, 2000). Consequently, the question changed from the simplistic “is multimedia an effective form of learning environment” to “with which design factors, for which learners, performing which tasks and expecting what outcomes is learning with multimedia environments effective?”.

The purpose of this chapter is to argue that now we have acknowledged the complexity of the question, we need to adjust our research methodologies to this new complexity. To this end, a typical experimental approach to understanding multimedia learning will be described and will be contrasted to new experimental studies that are developing beyond this more limited methodology. However, it will be argued that to truly understanding multimedia learning we need employ a variety of methodologies and the chapter will end by presenting a microgenetic account of three learners developing understanding of a multi-representational simulation.

This approach, common in what Goldman (2003) calls "first-generation" research, has been the most commonly used method to explore the effectiveness of multimedia learning. A typical (imaginary) experimental scenario might be considered to have the following characteristics.

> Participants for an experiment are recruited in return for credit on the psychology or education courses they are studying or are paid a small amount in reward for their time. They have no prior knowledge of the area and what they are about to learn will not benefit them in their future studies. They may be given a short pen and paper multi-choice pre-test to check that they have little prior knowledge of the concepts of the domain and then are randomly assigned to two groups. The first group receives the special multimedia condition that has been previously designed to be "good" according to the predictions of a current fashionable theory. The second group receives the same material but in a text-only control. A short orientation phase is provided to ensure that students know how to use the interface. They then learn with this material for 30 minutes and are then immediately given an pen and paper multi-choice post-test of the domain concepts, which typically will include some harder elements than the pre-test. They are debriefed, thanked for their participation and told not to sign up for further experiments, as they are not naïve to the material. The whole experience takes about an hour. The multimedia group do statistically better on the post-test than the control group and the results are interpreted to support the predictions of the current theory

We should acknowledge a couple of things about this scenario. Firstly, a single experiment rarely has all of these characteristics, but many do include a substantial proportion. Secondly, I have conducted a number of them myself (and still do); so lest anyone think I am pointing the finger of blame, imagine it pointing squarely at myself. I often use this methodology because experiments in this area have a number of desirable characteristics:

- **Use of theory to guide experimentation**

One of the strengths of research in multimedia learning in recent years is the emergence of theoretical frameworks, which integrate what might appear to be unrelated findings into a consistent whole. The two most commonly applied theories (which are strongly related) are the Cognitive Theory of Multimedia Learning (e.g. Mayer, 2001) and Cognitive Load theory (*e.g.* Sweller, van Merrienboer, & Paas, 1998). They focus on the nature of working memory (and its relation to long term memory) with its multiple, modality-specific limited capacity subsystems and identify the benefits that can accrue by presenting information that uses multiple modalities so that learners who actively process such information can profit from this.

- **Robust and replicable results.**

The wide spread acceptance of these theories means that a substantial number of researchers are contributing to the development of the theories and showing the robustness of their results across multiple laboratories. For example, research concerning the "split attention effect" – that separating pictures and text results in worse learning than integrating them into a single representation has been confirmed in many experiments (*e*.g. Chandler & Sweller, 1992; Kalyuga, Chandler, & Sweller, 1999; Mayer & Moreno, 1998*).* This can be enhanced when materials are shared across laboratories, (*e.g.* Bétrancourt this volume).

- **Reasonable statistical rigour**

As opposed to relying on intuition about the benefits of multimedia learning environment, these experiments use statistical methodologies to show when these intuitions are justified and when they are not. Furthermore, effect size analysis (Gain in Experimental Group's Scores – Gain in Control Group's Score)/ St Dev in Control Group's Gain Scores) can be used to allow at least some comparison about the relative effects of different treatments. This was not widely seen in the reports of early experimental research but is becoming more commonplace (*e.g.* Mayer, 2003).

- **Publishing "negative" data**

Generally, negative data or results are those that confirm the null hypothesis – in the imaginary experiment above the null hypothesis is that there is no difference in learning outcomes between those students who learnt with multimedia and those who learnt with text. For example, the learning with animation literature is full of experiments showing no difference between those students who learnt with dynamic representations and those who learnt with static materials (*e.g.* Pane, Corbett, & John, 1996; Rieber, 1990; see Price, 2002 for a review). Similarly, a significant amount of published research on learning with multiple representations has found no benefits for this approach (*e.g.* Tabachneck, Leonardo, & Simon, 1994; Guercin, 2001; Van Someren & Tabbers, 1998; Yerushalmy, 1991). By publishing negative as well as positives results, we are in a stronger position to weight up the costs as well as benefits of the impact of new technologies on learning.

- **Use of within system controls**

Typically, these experiments use within system controls, *i.e.* both groups of learners interact with the same technology, which only differs in the specific aspect of the interface that is under investigation. General explanations about the effects of computers of learning are therefore ruled out (increased motivation, immediate feedback, *etc*) and so we can be certain that the results are due to the specific features under investigation**.**

Second Generation Experiments

But, there are many ways that experiments such as the one described above could be improved and current second generation experiments are attempting to address such limitations. In this section, experimental approaches which address the four interacting characteristics of learning with multimedia – *i.e.* learner, representational, task and outcome characteristics will be reviewed.

Learner Characteristics

One of the key benefits that is commonly used to justify multimedia learning environments is that different types of learners may differentially benefit from alternative forms of representation. Multimedia can present the same information in many different ways so that learners can chose to focus on the representations that they find most useful. This intuition is not always backed up by research as it ignores the fact that learners may find representations difficult to integrate or that learning to select appropriate representations is a significant task in itself and learners may not always make sensible decisions. However, it does acknowledge one key aspect of multimedia learning – that people differ in what and how they learn with multimedia.

One problem with much existing experimentation is that it ignores this fact. Firstly, in most experiments even if there are significant differences between conditions, there is also significant overlap between subjects' learning outcomes in all conditions. Yet, this is simply ignored as the error term in an ANOVA rather that considered a subject for exploration. Why are some learners in the "bad" condition able to surpass the performance on those in the "good" condition? Does it simply relate to their prior knowledge or ability to learn new material or are they able to compensate for poor representational design in some way.

One explanation that follows from the assumptions of the limited capacity working memory accounts of multimedia learning is that as people differ in their WM capacity (Daily, Lovett, & Reder, 2001) then what may overload some learners will not overload others. What is often overlooked in the applications of such theories by multimedia designers (although not in the original conceptualisation of the theory) is the interaction between the limited capacity and constructivist nature of human cognition. WM capacity refers to chunks not items and so it is as much about long-term memory as short term. Thus, the string 441159515314 will be difficult for you to remember but as I know that is the UK international dialling code followed by the Nottingham dialling code, followed by my office phone number it is easy for me (3 chunks of information rather than 13). Thus, everyone's different long-term memory (schemata) will influence how they can interact with different representations. Broad assumptions about certain representations overloading working memory may not always be justified. Furthermore, complete novices who are the most common participants in these experiments are likely to be the ones without relevant schemata and who are most likely to suffer most from representations that are working memory intensive.

There may be other differences between learners that influence how people learn with multimedia. For example, research on learning with pictures has long acknowledged that there may be aptitude by treatment interactions. For example, Winn (1987) proposed that factors such as IQ, spatial reasoning, locus of control, field dependence, verbal ability, vocabulary, gender and age will mean that learners with different characteristics will differentially benefit from different forms of representation. Early research with non-computational media suggested that that lower achieving learners are more likely than their higher achieving peers to benefit from graphical representations of a task (see Snow & Yalow, 1982). It is often proposed that learners with high spatial or visual preferences will benefit from the graphical representations that are common in multimedia learning (*e.g.* Mayer & Sims, 1994). However, there is not necessarily a simple or face-valid relation between supposed cognitive style, representational preference and task performance. For example, Roberts, Gilmore, & Wood (1997) showed that high visual problem-solvers understood when to abandon visual strategies better than low visual problem-solvers.

Recent experimental approaches to multimedia learning are revisiting aptitude treatment interaction research to ask what sort of learner most benefit from dynamic or multi-representational learning environments. Increasingly common practice is to include pre-tests that examine learners' prior knowledge to ask whether learners with different expertise will benefit from the approach. For example, Mayer & Gallini, (1990) showed that learners with less domain-specific knowledge benefit more from multimedia material (text and pictures) than from text alone. Seufert (2003) placed learners into one of three categories based on their prior domain-specific knowledge of chemistry before they learnt complex chemical concepts with multimedia software. They also received one of three types of help for supporting their understanding of the relation between representations – directive, non-directive or no help. Seufert found that learners with medium levels of prior knowledge increased their comprehension of the material most when given help – learners with too much or too little knowledge did not benefit to the same degree and in some cases, help was even harmful. An important addition to understanding the relation between learners' prior knowledge and learning with multimedia is to disentangle learners' familiarity with domain concepts from their familiarity with the representations employed. This is exemplified by Stern, Aprea, & Ebner (2003) who examined both prior knowledge of the domain under investigation (economics) as well as understanding of linear graphs.

The finding that learners with different amount of prior knowledge benefit from different sort of multimedia experience causes some concern though about the standard experimental paradigm where it is common for people with no background knowledge to participate in return for payment or course credit. One could argue that many multimedia environments are designed for people who are in the early stages of learning a topic and hence experimental participants provide a good analogue of this population. However, it is commonly the case multimedia environments are used by students after initial exposure to the topic in other forms, (*e.g.* lectures, readings) and often as part of a more extensive curriculum. Furthermore, people learning with multimedia outside experimental settings are presumably doing so because they want to know something about the topic, perhaps because they are fulfilling some personal learning objective or because they are required to learn to pass examinations. It may not be wise to generalise about the suitability of different approaches to multimedia from this experimental population to the actual intended users of the learning environments.

An approach that relies on classifying learners into categories is still limited in the questions it can answer, as it does not tell us much about **why** certain learners benefit whilst others do not. Interaction measures can be used to examine why some learners are successful and some not irrespective of design factors or learner characteristics. One of the key advantages of experimental approaches to multimedia learning is the opportunity to collect easily and automatically a wealth of data about how learners interact with the representations. Interaction measures (behavioural protocols) that can be used include time on task data, progression through curriculum, use of various systems features (*e.g.* learner control of dynamic representation, selection of different representations, *etc*), amount of help sought or provided, performance on questions, *etc*. This data is often (though not necessarily) more difficult to collect in more naturalistic situations – for example, time on task data collected in real contexts is normally too noisy to provide reliable information. Such interaction data can be used to explain why some learners are more successful than others. For example, Zahn, Barquero and Schwan (in press) compared subjects learning with different hypervideo designs and with the text and video materials presented without hyperlinks. Irrespective of the specific condition it was those students who activated more links and spent longer reading the content who learnt the most. On balance however, surprisingly few experiments report this sort of interaction data and relate learners' use of representations to either prior knowledge or learning outcomes. This is to be regretted as this information could tell us much about how learners actually use such systems, and collection and analysis is normally relatively straightforward.

A second type of process measure that can be used to understand individual differences in learning outcomes involves changing the nature of the learning experience. Examples include eye-movement data (see Tabbers, this volume), poor men's eye trackers (*e.g.* Romero, du Boulay, Lutz, & Cox, 2003) which only uncover parts of a screen when a cursor is moved over them, videoing learners to capture gesture and other forms of non-verbal behaviour, and various forms of protocol data. For example, Lewalter (2003) collected verbal protocols from students as they learned with text, which in two cases was supplemented with either graphical dynamic or static representations. She found no difference between the two illustrated conditions, but both were better than text only. However, she found that learners provided with static representations in addition to text produced more rehearsal strategies. A similar analysis was conducted by Ainsworth & Loizou (2003) who asked students to self-explain when learning about the cardio-vascular system with either text or pictures. Students presented with pictures produced significantly more self-explanations, were more likely to produce explanations that include goals or principles and that self-explaining was more strongly related to learning outcomes in the pictures rather than the text conditions.

Unfortunately though it can be difficult to get at detailed measures about learners strategies using classical experimental methodology. For example, Ainsworth, Bibby & Wood (2002) examined how learners used combinations of either pictorial, mathematical or mixed pictorial and mathematical representations to answer estimation problems. Interaction data was gathered which was able to relate learners' use of representations to post-test measures. Only learners given pictures and mathematical representation improved their estimation performance. The interaction measures showed that learners in all conditions improved in their use of representations. Learners in the pictures and mathematics conditions also increasingly converged their use of representations over time, but learners in the mixed condition never did so. They suggest it was this failure to successfully translate that was related to the lack of learning. However, what this experiment does not tell us if whether learners were trying and failing to translate between the mixed representations or whether they did not even try to translate between the representations. In the second half of this chapter, a microgenetic study will be presented that was partly aimed at addressing this question.

Representational Factors

In order to answer the more complex question of when and for whom does learning with multimedia prove beneficial we need to increase the depth of our analysis of the representational aspects of multimedia learning. One current problem when trying to generalize across multiple studies is that an insufficient level of detail is presented about the design of the multimedia environment and the representations that are used. In much of multimedia research, representations are described simply in terms of modality - pictorial/graphical or textual. Pictorial representations are depictive in that they explicitly preserve geometric and topological information whereas textual representations are descriptive as they have an arbitrary relationship to the object that they represent. The other common classifying dimension in multimedia research is sensory channel (*i.e.* auditory, visual, less often haptic). The current theoretical focus on dual coding and cognitive load theories has tended to lead to a situation where these dimensions of modality and sensory channel are seen as the only ones important for representational analysis. Yet, there are many of ways that representations can differ from one another. For example, other dimensions that have been used to classify representations include precision (the level of accuracy of information *e.g.* qualitative to quantitative), specificity (the extent to which a representation permits expression of abstraction), perspective (what is represented such as functional or structural relationships) and complexity (the amount of information) (see de Jong et al., 1998).

Similarly, Ainsworth and Van Labeke (in press) argue that the term dynamic representation, in contrast to static representation does not capture the ways that dynamic representations can differ from one another. They maintain that three types of dynamic representation are commonly found in educational software: a) time-persistent, which explicitly shows a range of values using a spatial dimension to indicate time; b) time-implicit, which also show a range of values but not when those values occur; and c) time-singular which only show one value of a variable at a time and are the classical case of animation. These different dynamic forms of representations have different informational and computational properties and that learners are sensitive to these representational features. Furthermore, (Lowe, 2003) shows how subtle perceptual features can strongly influence what novices comprehend and remember from animated weather maps. Participants attended to features that dynamically contrasted with their surroundings, either by changing substantially more or by changing substantially less than their surroundings. However, these features were not necessarily the ones of most conceptual interest. Again this points to the need collect much richer data that is common in much experimental research with multimedia.

These findings suggest that we need to widen our theoretical stance to understand how representations influence learning (see also Reimann, 2003; Klein, 2003) and to do that we need to describe the representations used in multimedia environments in much greater detail. Obviously there is no single 'right' way to describe a representation and different approaches differ in what they emphasise and the level of detail observed (see Blackwell & Engelhardt, 1998). However, a more detailed approach of the sort more common in cognitive science would provide for the theoretical extension for which Reimann argues. For example, representation could be described in terms of how they encode and present information (the 'format') and the 'operators' that are necessary for extracting meaningful information from the representation (*e.g.* in the case of a graph how to find the gradients of lines, maxima and minima, intercepts, *etc*.).

In addition to describing the representations in more detail, we may also need to describe other design factors more precisely. For example:

- What was the interface to the representation? (menu, textual input, point and click, *etc*) as the interface to a representation can effect what is learnt (Svendsen, 1991; O'Hara & Payne, 1998).

- What other representations were present? (Ainsworth et al., 2002) found that understanding of a representation was influenced by others that were presented alongside "representational chemistry". Much research that claims to be contrasting text versus diagrams, for example, is often contrasting text versus text and diagrams.

- Was the representation under learner or system control? (Tabbers, Martens, & Van Merriënboer, 2001) showed that the often-cited advantage of auditory text with animations was only present in situations where the animation was not controllable by the learner.

- How much information is presented in the representations. Ainsworth & Peevers (2003) found that the efficiency of learners' problem solving was influenced by an interaction between the information and computational properties of representations. Participants were given instructions concerning the operation of a complex device in either four simple representations or one complex representation. These were presented either as tables, diagrams or text. There was no difference in the efficiency of problem solving for those using tables or diagrams if given one complex or many simple representations. However, those given complex text spent much longer studying representations than those subjects who saw multiple texts but they were also more likely to find the ideal solution to the task.

Task Characteristics

It is well known that particular forms of representations facilitate performance on some tasks but not on others, (Gilmore & Green, 1984). For example, Bibby & Payne (1993) gave participants instructions on how to operate a simple control panel device using either tables, a procedure or diagrams. Participants given tables and diagrams identified faulty components faster than ones provided with information expressed as specific verbal rules. These participants, however, proved faster given the task of deciding which switches were mispositioned. Taxonomies such as Cheng's (Cheng, 1996) who described specific roles that diagrams can play or Ainsworth's (Ainsworth, 1999) that describes the different functions that multiple representations support can help in examining how to match representations to tasks. One problem with generalizing from the current research to widespread applications of principles is that the majority of research has been conducted in the pure sciences, with much less research addressing social sciences, humanities or arts. These subjects typically don't have a single correct answer and may need different forms of representations to help convey the shades of grey involved in understanding these topics.

Another concern with the nature of the tasks used in most experiments is the timescale of the learning experience. In the vast majority of experiments, participants only interact with the learning environment once and that for a short amount of time (typically under one hour). This leads to a number of questionable practices. Firstly, it tends to confound the time learners must spend in learning to understand new interfaces and representations with the time spent learning the domain though these representations. New forms of representations must be learnt and in many cases this will be a complex task. Learners must come to understand the syntax and semantics of representations, may need to learn how to select and construct representations and in multi-representational cases may need to translate between representations. It is unlikely that the short training sessions that most experiments employ will allow learners to have completed all these tasks. Secondly, the interesting questions of how learners change and adapt as their expertise grows is not addressed in this sort of methodology. Finally, it is questionable whether the results would apply to situations where learners are interacting with environments over more extended periods of time.

Learning Outcome Characteristics

Similarly to task characteristics, the relationship between multimedia and learning outcome characteristics have received some experimental analysis. For example, Mayer (2001) commonly uses retention and transfer tests to examine the interaction between different multimedia designs and types of learning outcomes. Similarly, Ainsworth & Loizou (2003) examined the difference between self-explaining with text and diagrams on explicit, implicit and knowledge inference questions. These types of post-test aim to analyse the differences between the relatively easy to learn, declarative knowledge and more challenging expert knowledge including mental models and transferable knowledge. Typically multimedia design has a larger impact on the latter form of knowledge.

However, there are a number of ways that experiments could improve their approach to learning outcome analysis. For example, there remains a reliance on multi-choice and true-false questions which is understandable as they are quick to administer and mark. However, this has a tendency to bias outcomes towards declarative knowledge. Secondly, the majority of experiments rely on an immediate post-test to assess learning outcomes. Yet, different styles of intervention can differentially impact on the outcomes of learning at different rates. For example, research comparing collaborative to individual learning has shown that benefits of collaboration become more apparent on a delayed rather than immediate post-test (Howe, Tolmie, Anderson, & Mackenzie, 1992). Thirdly, limited emphasis is placed on the modality of the test items. Typically, irrespective of the modality of the learning material, post-test items have a tendency to be textual. We may be underestimating the benefits of different forms of representation this way. If dynamic representations lead to dynamic mental models, should we give learners tools to create dynamic representations at post-test?

Furthermore, the representational aspect of what is learnt is often not assessed separately from the conceptual. One exception is that of Van Labeke & Ainsworth (2002) who designed (multi-choice) post-test that examined learners understanding of the concepts of population dynamics, the representations (*e.g.* time-series graphs, histograms, tables and phaseplots) and the relationship between representations. They found learners understanding of the concepts and representations at post-test far exceeded their understanding of the relationship between the forms of representations. Given that much multimedia research involves novel representations formalisms it would be nice to routinely differentiate between learners understanding of representations and the way they encode domain. For example, Cheng (2002) showed that learners could solve complex electrical circuit problems more easily when solved with a novel form of diagram – a law-encoding diagram. To truly understand the contribution that this diagram made to facilitate problem-solving, it would be beneficial to see if learners could apply this form of representation to domains with similar deep structures and whether they could solve problems in this electrical circuit domain more easily with other forms of representation.

Non-Experimental Approach to Multimedia Learning

Although experiments form the majority of the research on multimedia learning, they are not the only methodology that has been applied. Techniques such as modelling, naturalistic accounts of representation use in "real world" contexts and microgenetic accounts of learning have all been applied to this problem. Although these methodologies differ in the granularity of their accounts, what they all have in common is a focus on the process of learning. Researchers such as Tabachneck-Schijf, Leonardo, & Simon (1997) and Lane, Cheng, & Gobet (2000) have developed computational models of multiple representational understanding. A number of more naturalistic studies have examined the context in which representations are used (*e.g.* Roth & Bowen, 2001; Kozma, Chin, Russell, & Marx, 2000). This contrasts to experiments where typically multimedia systems are considered in isolation without examining the context in which they are embedded (*e.g.* what would teachers do to support multimedia, most children in schools (at least in the UK) use computers in pairs, *etc*). Kozma (2003) explored the difference between expert and novice use of representations that he had first examined experimentally by observing the use of multiple representations by practicing scientists in a chemistry laboratory. He showed that experts reasoned very differently with representations to novices. They were able to coordinate different forms of representation, selecting representations that fit their purposes and furthermore could act as mediator for social interaction so they could come to a shared understanding of the task. By contrast, students did not connect representations spontaneously and demonstrated little awareness of the relation between the phenomena and the representations.

Finally, experiments in this area typically do not examine the timescale of learning. Microgenetic accounts of learning and strategy development are an excellent way of exploring this. Schoenfeld, Smith, & Arcavi (1993) examined one student's understanding of function using a multi-representational graphing environment. Using micro-genetic analysis, they describe in detail the mappings between the algebraic and graphical representation in this domain. Working with one student over a number of sessions, they showed how a student could appear to have mastered fundamental components of a domain both in terms of algebra and in terms of graphs. However, because some of the connections between these modes of representation were missing, her behaviour with the representations was often misguided.

In the study described in the remainder of this chapter, a similar method to Schoenfeld et al. (1993) is used to examine how learners with very different background knowledge use a multi-representational simulation environment. A number of questions motivated this study. What sources of difficulty did learners find in learning with a complex multi-representational simulation and what strategies did they develop to overcome this difficulty? How did their behaviour change over time as they became more expert with the simulation and the domain? What use did they make of systems features to help them learn? Finally, much of the analysis focussed on how learners came to understand the relationship between the various representations.

Learning the Concepts and Representations of Population Dynamics

DEMIST (Van Labeke & Ainsworth, 2001) allows authors to create multi-representational instructional simulations. Its design is centred on Learning Units that describe a single mathematical model and include a set of representations and instructional activities. So that learners can easily compare the effects of different parameters, learning units can consist of multiple experimental sets, which vary such factors as birth rate, initial population density, *etc*. Authors decide on which representations to include up to a maximum of 16 per unit and which variables and parameters to display in each type of representation. A key feature of DEMIST is that the degree of translation between representations can be varied. Thus authors can chose that representations be dyna-linked (if a learner modifies the information in one representation, this action is reflected onto the other relevant representations), mapped (a learner selects a value in one representation to find the corresponding relationships in other representations) or independent. Finally, authors can describe a small number of additional activities available for learners. Learners can be encouraged to make *predictions* about the values of the model in the future or perform *actions*, which allows the learner to act on a value at the current stage of the simulation and change it.

In this study, DEMIST was authored to describe four simplified models of population dynamics. Population dynamics is the branch of biology that attempts to discover the rules that govern how populations of living organisms grow, decline or oscillate. It describes such features as a species birth/death rate, the maximum number of individuals that can be supported in a population, the rate that predators capture prey and how availability of prey influences the number of predators that can be supported. The relationship between these features is given by fairly complex mathematical expressions. The four models described a single species growing in an unlimited manner (SSUL), single species growing in a limited manner (SSLG), a simple predator-prey relation (PP), and two species competition (TSC) and were introduced in this order to instantiate model progression (White, 1993). A typical mathematical expression for predators and prey is $dN/dt = r\,N(1-N/K) - \alpha\,N\,P$, (where r = prey fertility, N = population density of prey, P = population density of predators, K = carrying capacity of the environment and alpha = the number of prey killed by a predator per unit time). Each model contains four learning units and each unit allowed the learner to select from amongst 16 ERs. Representations available include time-series graphs (*e.g.* the population density of number of prey and predators on the Y axis with time of the X axis), Ln(X) graphs (*e.g.* the Y- Axis represents the natural log of population density with time on the X axis), phase-plots (*e.g.* the number of prey on the X-Axis and number of predators on the Y-Axis), tables, animations, histograms, pie-charts, text that provides definitions of terms and pictures of the animals described. These representations varying also in their dynamic features as some are Time-Persistent (*e.g.* dynamic time-series graphs or tables), some Time-Implicit ERs (phaseplots) and some Time-singular ERS (*e.g.* animations, histograms).

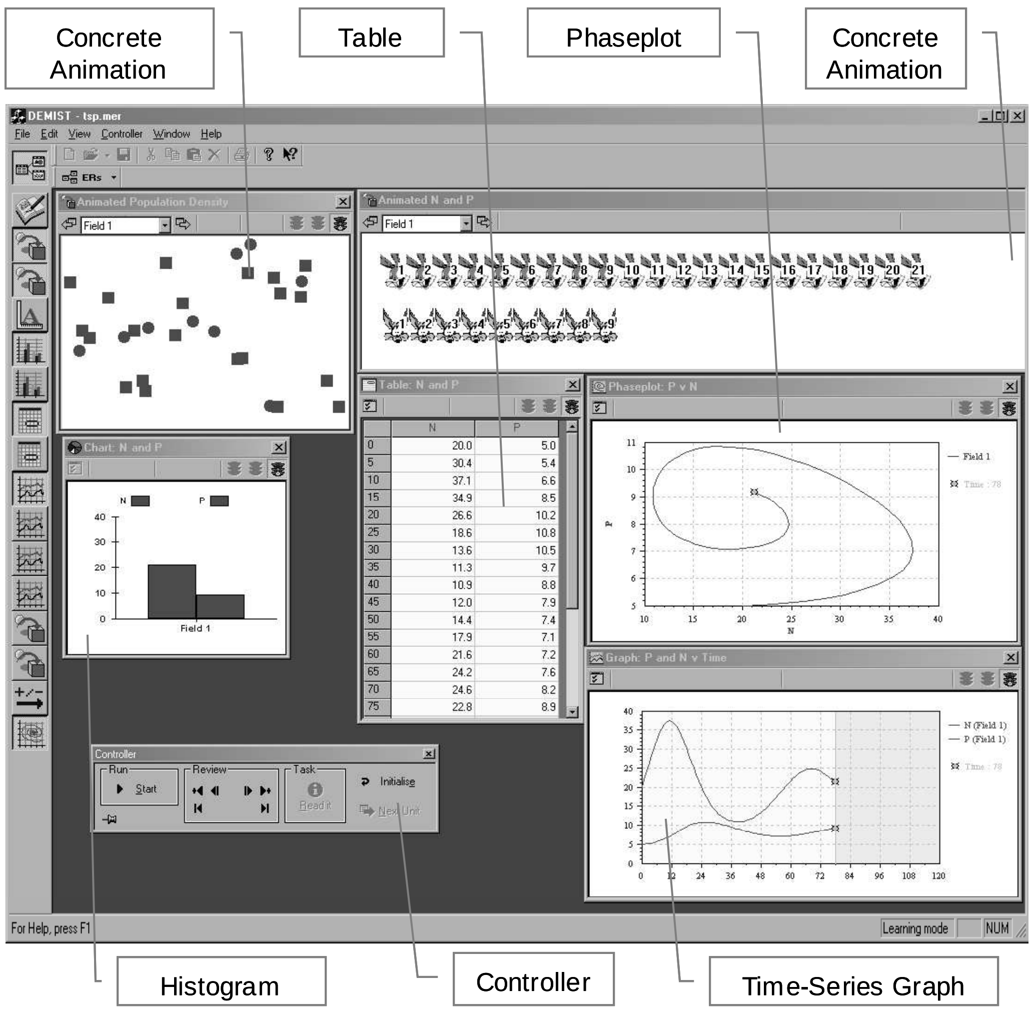

Figure 1 A screenshot from DEMIST with a simple predator/prey model. All representations in this case show two dimensions of information (N & P) with varying precision, dynamic form, & modality.

Each unit include a wide variety of representations of different form and with different dynamic characteristics. Each model begins with a unit that allows learners to explore the characteristics of the models by changing parameters and exploring the effects on the representations. Subsequent units suggest different activities where learners were asked to predict future values (*e.g.* changing prey birth rate or the initial population density to examine the impact on future numbers). Depending on the representation, this prediction could be performed in different ways. For example, in the table, the learner scrolls down to an appropriate point and directly edits a value that can subsequently be compared to the real value. In the time-series graph, the learner clicks on the graph and drags a "hypothesis" marker around which can then be compared to the graph. Types of representation afford different strategies with graphical representations emphasizing visual estimation and tables suggesting precise and accurate calculation. Learners can also act to change values after the simulation started running, for example, one unit asks learners to set up a "safe" area and a "poached" area and see what happens to prey numbers after predators numbers are altered. Again, the ways these actions are supported depends on the nature of the representation and are similar to the way that predictions are made. These activities were described in an accompanying worksheet. Furthermore, an experimenter was present throughout to support learners. Although she rarely taught the concepts directly, she suggested experiments, supported learners when they misinterpreted representations (*e.g.* by drawing their attention to inconsistencies in their inferences) and answered questions. Her role is considered crucial to what learners came to understand.

Three participants took part in the study. The first, PJ, was male with undergraduate and post-graduate degrees in Computer Science. He had no previous experience of population dynamics and had not studied biology passed the age of 16. The second, LN, was female who was currently studying for an undergraduate degree in Biology. She had taken one course recently, which covered some of the basics of population dynamics. The third participant was a male who had just completed an undergraduate degree in Biology. He had studied population dynamics previously and had some experience in computational modelling of biological phenomena.

The initial session consisted of a 30 minute training course on DEMIST's features. This was provided in the context of a simulation of simple harmonic oscillation. This model consisted of three learning units which showed learners how to select ERS, how to act upon them and the change values of the parameters, run and review the simulation and how to dynalink between representations. In subsequent sessions, the learners worked with the simulation for between 60 and 90 minutes to understand the model, performing the activities suggested by the worksheet and the experimenter. All of these sessions were videoed for later transcription and throughout learners were asked to explain their actions. Periodically, the experimenter interviewed the participants about their learning goals, strategies, focus of their educational activities and sources of difficulty. DEMIST also captures rich interaction data describing which representations participants select and for how long, what actions they perform upon the representations and where dynalinking requests are initiated (see Figure 3).

Analysis

Analysis of the participants' interactions with DEMIST revealed how complex the tasks of learning the concepts and representations of population biology to be. It also revealed that learners had distinctive strategies for handling this complexity that were based on their personal learning goals and background knowledge and skills. To begin, some of the difficulties that these learners faced in understanding the concepts and representations of population dynamics will be summarised. Then, the different strategies they developed in light of these difficulties will be discussed.

Source of difficulties in learning with multi-representational simulations

DEMIST presents learners with a number of external representations. In some cases, they have fairly simple format and operators (*e.g.* text, tables) and learners had little difficulties in interpreting them correctly even if they were not always aware of the implications of the representations. In other cases learners found the representations very difficult. The representation that caused the most difficulty was the phaseplot. This representation plots two variables against one another directly and time is the parameterising variable along the plot line (see Figure 2). The form of representation was used in DEMIST to show such relations population density of two different species (*e.g.* predator and prey, or competing species) or the population growth rate against population density.

This representation is very similar to the time-series graphs and a modality analysis of representations would not distinguish between them. Yet, learners had little difficulties in understanding the time-series graphs (indeed they were the single most used form of representation by all learners) in contrast to their difficulties with this representation. Indeed, one of the most common problems for all the learners was their tendency to assume apply the operators of time-series graphs to interpreting phaseplots. On a number of occasions, participants tended to assume that movement across the axis would be constant. For example, that it should take the same amount of time for the prey population to increase from 20 to 30 as from 30 to 40 in the graph in Figure 2. This misconception was quite robust and would reappear when new dimensions were plotted on the phaseplot. The advantage of the dynamic phaseplot eventually allowed all learners to overcome this misconception by running the simulation and see the values change in a non-constant way. Consequently, this misconception is likely to be even more robust in a pen and paper versions of phaseplots. Even when they had become familiar with phaseplots, learners were not always quick to see the implications of the representations. For example, when PJ was asked to estimate the values of prey (N) and predator (P) population density at the stability point for the population he was unable to do so. Yet, he had run the simulation many times and had seen the phaseplot converge to a single point (in the example above this would be at about 20 and 9). Instead he tried to use either the table or time-series graph and did not succeed at the task until he dynalinked from them to the phaseplot.

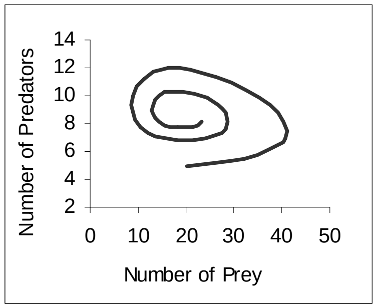

Figure 2 A phaseplot of prey and predator population density

Another problem was simply misreading representations. For example, a participant looked at the N and P time series graph and claimed that both populations peak at the same time rather than a quarter cycle out. This problem was sometimes enhanced by the automatic scaling that DEMIST uses to calculate appropriate minimum and maximum values for graphs. For example, when estimating future values of population growth rate against N, TW relied on the value "always ending up in the top right". He was initially unable to accept that this was not a property of the underlying mathematical model and tried many different values to "prove" that the values always ended up in the top right. Automatically scaling also meant that sometimes graphs that were curved looked straight if participants used low values for some parameters. This caused particular problems in the single species limited growth model. Participants had previously worked with an exponential growth model with meant that values such as population density gives a straight line when plotted on a log scale graph. This fact caused participants to look for other linear graphs and they had a tendency to try to interpret many graphs in this way. A number of other researchers have found that graphs are particularly susceptible to inappropriate overgeneralization (*e.g.* Scanlon, 1998).In addition to problems with interpreting representations, learners also had difficulties in managing other aspects of the learning process. One problem was designing and interpreting experiments. Both TW and LN had a tendency to understand the model by changing values and running the simulation again to compare the results of this experiment with previous ones. One problem that was observed was that it was difficult for learners to design and interpret suitable experiments (hence ideally DEMIST will be designed to support learners like SimQuest does (de Jong & van Joolingen, 1998). Firstly, although the participants' behaviour often looked appropriate (for example, all participants tended to change one value at and time holding the others constants), it could hold limited meaning from them. This was sometimes due to limited understanding of the concepts that were being represented. For example, PJ used the Ln(N) v T graph in the SSUL model to predict the value of Ln(N) after the simulation was run for 20 steps. He did this 11 times with different values of b and d and became very good at the task. However, he did not know the meaning of Ln(N) nor did he know how to relate it to N. The activity was decontextualised and had become purely syntactic – "can I visually estimate the gradient of this line and extend it". Also, participants had a notable reluctance to record values for their experiments (DEMIST does not do this automatically) on the pen and paper provided. Without recording the results of their experiments and the specific values of the parameters, they were able to learn little from their interactions with DEMIST. The impression gathered from their behaviour and comments is that they knew this was what they were supposed to do but did not quite know why!

Secondly, participants would often change parameters by values that were much too small to be noticeable and conclude that changing the parameter had no impact on the aspect of the model they were exploring. No participants spontaneously chose to explore boundary conditions and they had to be strongly encouraged by the experiment to use more extreme values.

Finally, it was quite easy for participants to draw the wrong conclusions from their experiments. For example, in the case of SSUL the equation that governs population growth is birth-rate – death-rate x Population Density (dN/dT = b-d*N). PJ changed the values on b and d to explore their impact on population growth. After running a number of experiments, he concluded that birth rate had a much greater impact on population growth rate than death rate. He drew this erroneous conclusion because of his habit of doubling values in the parameters. For example, he changed birth rate from 0.4 to 0.8 and then death rate from 0.2 to 0.4. The resulting outcomes from the simulation gave him no reason to change his mind about his theory, only direct intervention from the experimenter caused him to reason about b-d instead of b and d separately.

For these reasons (and more), learning with multi-representational simulations is far from straightforward. Learners have to manage a number of complex tasks concerning learning simulations as well as learning from multiple representations. In this next section, the strategies that the learners adopted for handling this complexity are discussed.

Strategies for Learning

Earlier, it was argued that only by exploring learners' strategies in fine detail can a true picture emerge of how people come to understand complex multi-representational situations. To understand each of the participants learning strategies their discussion with the experimenter during and after the intervention was described and matched to the behavioural protocols that DEMIST captures.

Table 1 Descriptions of activities

	Time	Runs of Simulation	Mean No of Co-present ERs	1st Session	Final Session	No of Dyna-Links	Actions	Hypotheses
PJ	7:50	109	5.63	8.48	5.48	581	12	34
LN	7:07	145	4.94	6.05	4.52	68	39	6
TW	7:05	110	6.32	7.05	6.93	154	36	38

Table 1 shows a summary of the learners' interactions with DEMIST. In some way, the learners were similar to one another but in many ways they were subtly different. The first striking result is how many representations learners had open simultaneously – an average of 5.6 across the 16 learning units and 3 participants. This is far larger than the vats number of multimedia environments. However, this does not mean that learners used all of these representations. A very common pattern was for learners to open the concrete animation and then never look at or interact with that representation at all. LN when asked why she did this explained that "*its silly but its comforting to know its there*". Another strategy (particularly common for TW) was to open more representations than screen space comfortably allowed and then deliberately stack representations on top of one another such that all but one representation was obscured. Then he would occasionally select one of these hidden representations for some specific purpose before hiding it again. For example, each unit contained a textual representation that provided definitions of the key concepts. TW would normally selected this to be one of his hidden ERs which he would draw upon if he needed to refresh his memory about the definitions of concepts. In this way, he seemed like he created his own menu of useful representations from the 16 available by pre-selecting ERs rather than opening them from the main menu when he wanted to read one.

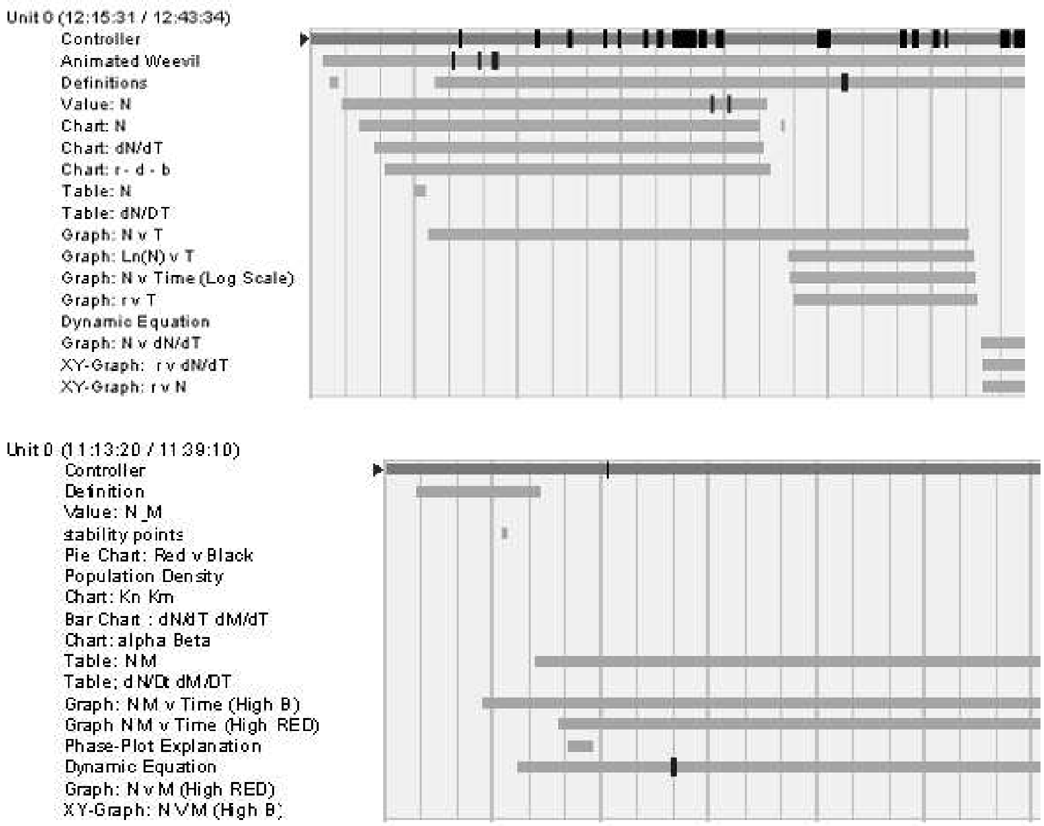

Figure 3 DEMIST use traces from the first unit of a) SSUL and b)TSC for PJ.

Two participants (LN and PJ) significantly decreased the number of co-present ERs they used as their experience with DEMIST and the simulation grew. Figure 3 shows how PJ's behaviour changed over the course of the session. The first unit of every model introduces learners to the key concepts and representations associated with that model. In the SSUL unit, PJ selects many representations and has eight representations open simultaneously for the majority of his interaction with the simulation. This interaction is dynamic and interactive (this can be seen by the bars in the controller line of the representation which show when he interacted with the controller to run or rewind the simulation). He also for the most part selects ERs simply by the order they are presented in the menu. In the TSC unit by contrast, he selects far fewer ERS, is much more strategic about the ERs he does chose (identify them by description and selecting many graphs and a table) and runs the simulation far less often. At the beginning of the sessions, PJ (the participant without background in the area) spent a lot of time learning what each representation showed him. He had to invest a lot of time and effort in learning the format and operators of the representation and relating them to (unfamiliar) domain concepts. By the end of the sessions, he had learnt to select ERs that he felt were most beneficial (for him) when learning about the domain. This was after nearly six and half hours of intensive work with the simulation and again suggests that the sort of behaviour we observe from short-term experimental participants may not be representative of a the wider population.

PJ had no knowledge of population biology when he started to work with DEMIST, but he was confident in the use of computers and familiar with most of the forms of representation (with the exception of the phaseplots). When asked what he considered to be his learning goal, understandably he replied that he was trying to learn "population biology". He also stated that he found the hardest task to be "looking at a representation and understanding what it means". Without any knowledge of domain concepts, PJ only had his knowledge of representations to guide him and much of description of his understanding remains at this level. For example, when describing the Killed (N) and Potential (N) graph (aggregates from the predator-prey model defined in text and by equations for the participants) (LHS Figure 4) he explained the graph as *"First there is more Potential N on this side and then kill N goes down…. And then the green one goes above it and* *then the black ones goes".* He rarely looked at the equations or definitions to help him understand the terms focusing on the visual pattern of the representations

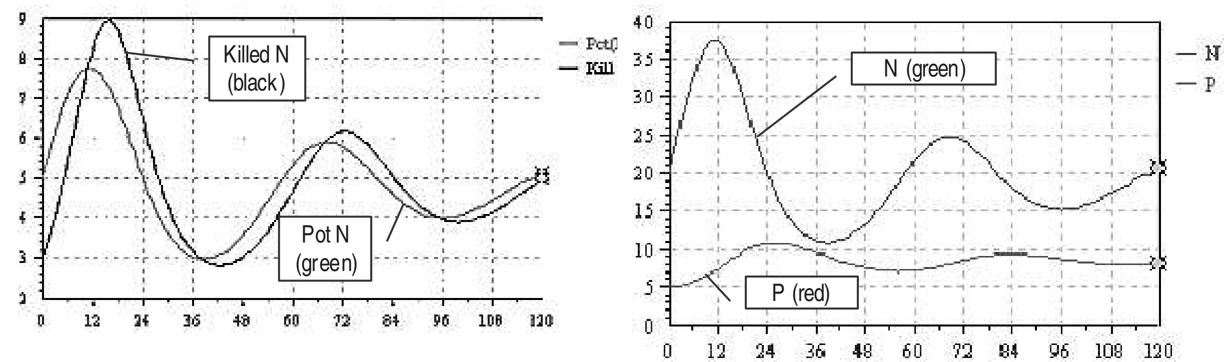

Figure 4 Two time-series graphs a) Killed (N) and Potential (N) and b) N and P

Similarly, his strategy for translating between representations was based on a visual and syntactic strategy. He described it as follows, "Look at axes and scales, see if the numbers match and whether the pattern on the graph matches". In the example for Figure 4 "*They are both the same shape, the maximum points correspond on the X scale*". This syntactic translation strategy was strongly supported by the use of dyna-linking which PJ used far more than any other participant. Dyna-linking in DEMIST shows the relationship between two representations by indicating the common points and as such suits PJ strategy perfectly.

By contrast, LN was the least computer experienced of the three participants and had received recent although not extensive teaching in population biology. She described her learning goal as "*to learn maths and representations to begin with and the biology towards the end*". She focused on the meaning of the representations working hard to understand how unfamiliar representations expressed concepts of the domain. For example, when first introduced to a time-series graph showing red and black ant density (two competing species) "*When the red ants are the stronger ones they reach stability and black ants can grow at the beginning until the number of red ants increase*" She rarely discussed the visual appearance of the representations focusing on what they meant rather than how they looked.

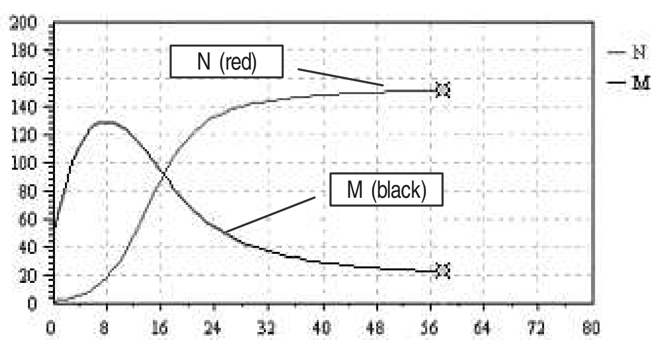

Figure 5 Population Density of two competing ant species (N red ants) and (M black ants)

LN was a "representational resistor" and said she would have been quite happy to have worked exclusively with graphs. She stated that the hardest task was "*relating representations"* and actively avoided it if possible. In contrast to PJ, when describing the relation between the two representations in Figure 4, she focuses on the domain concepts of each representation. *"When number of prey killed by predators peaks then that's when number of prey numbers are at their highest, no just before"* To some extent she does not have a translation strategy at all, as in many cases, her preference is to select the single best representation for the task she is trying to perform. As such her avoidance of dyna-linking can be seen as matching her strategy for learning the domain, as it does not support her way of reasoning.

The final participant, TW had experience with computational modelling of biological phenomena and so began the session with more background knowledge than either of the other participants being familiar with concepts, representations and the process of modelling. He described his goal as "*to learn the relationships between the representations because its another dimensions to what is going on*". He stated that the task he found most difficult was "*relating values in equations to graphs and any sort of visual representation*". This focus on relating representations is likely to be the reason why TW was the only participant to continue with the same high number of co-present representations throughout all four models. When relating the two representations in Figure 4, TW reasoned as follows *"While this is going up [gestures to Killed N (a)] I think there are more hares living than dying. At the point where it crosses potential = dead, and at each of those dN/dt is zero, and so that must be the maxima [points to N (n)]* ".

In this cases, TW is using is understanding of the domain concepts and the operators of graphs to reconcile the two representations. He describes his strategy as follows "*to find time or another dimension that was common to all of the representations but then ignore irrelevant dimensions. Find what is changing in both, like gradient or shape of curve*". TW was a moderate user of dyna-linking. When asked if he found it useful, he replied that he thought it was cheating. As he stated, he wanted to learn how to relate the representations and did not want the software to do it for him! However, TW had the most varied translation strategies of all the three participants. When reasoning about new models or unfamiliar terms, he sounded very like PJ and focused on the visual surface features of the representations. Then as he related these new models and terms to his prior knowledge, he began to relate the visual features to the underlying concepts. This finding is reminiscent of Roth & Bowen (2001) who showed that scientist's graphing practices are highly related to their familiarity with the phenomena.

All three participants tackled the same task, learning the concepts and representations of population biology in very different ways. They interpreted the goal differently, used the features of DEMIST in such ways as to support that goal and each developed their own strategy for relating representations – a task they all agreed was very difficult. PJ with no background in the area had little choice but to use his knowledge of the syntax of representations to understand both unfamiliar concepts and representations simultaneously. LN disliked relating representations and attempted to understand the domain through a known representation and then use her knowledge of the domain to understand the unfamiliar representation. Finally, TW knew both the representations and the domain but tended to use known surface features of a representation to relate to an unfamiliar representation.

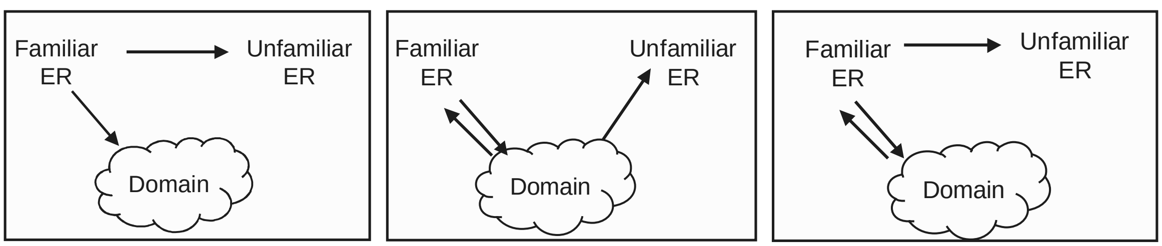

Figure 6. Graphical representation of a) PJ, b) LN and c) TW's translation strategies

It is not the point of this chapter to argue that one of these strategies is more successful than the others as no data was collected to support this argument. Furthermore, the participants started the study with such differences in knowledge a comparison of learning outcomes would have been meaningless. However, it of concern that PJ still reasoned primarily about the surface features of the representations at the end of sessions and was consequently easily thrown by differences in the visual appearances between representations. The reliance on mapping syntactic features and the role that dyna-linking played in supporting this, gave a success that at times seemed too easy. A number of researchers (*e.g.* Kozma, 2003; Scaife & Rogers, 1996) noting the difficulties that learners have in relating representations have called for learning environments to automatically dyna-link representations. However, dyna-linking may be appropriate when people are learning new concepts and when they have become familiar with the representations and are trying to use them to reason about new ideas, but systems may need to fade dyna-linking contingently to encourage learners to develop this understanding for themselves. Otherwise we may be guilty of over-scaffolding the very knowledge we are trying to help learners build.

Conclusion

There are many benefits to evaluating multimedia learning experimentally. However, such an approach is not the only way that multimedia environments can be analyzed. Fine-grained process accounts offered by more detailed accounts of fewer participants working over extended periods when they are engaged in a more authentic learning tasks has much to recommend them. The microgenetic study with DEMIST revealed the complexity of learning the concepts and representations of population biology. A previous experimental study (Van Labeke & Ainsworth, 2002) had given participants only 90 minutes to learn the interface, representations and three models of population biology. This we now recognize was not close to sufficient. It also revealed how important the experimenter's role as a tutor was in supporting learners' understanding. Much of the time students needed far more support in recognizing and overcoming their difficulties than was provided by DEMIST's free discovery approach. However, the main benefits of the approach lies in the insight it has provided into the relationship between learners' goals, their strategies and the way they interacted with the environment. Each learner chose to focus on a different goal - to learn the domain concepts, the representation and mathematics or the relationship between the representations. Their strategies were aimed at achieving those goals and the way the interacted with the system features supported these strategies. Consequently, it reveals that asking a question such as "is dynalinking useful?" is not likely to produce an unequivocal answer as it depends on upon whether dynalinking acts to support each learners' goals and strategies. Finally, this sort of study provides more insight into how learners come to understand the relationship between two representations. Although, the importance of coordinating multiple representations is well recognized in the multimedia comprehension community, much of the research to date has focused on the conditions when coordination is more likely to occur. Process accounts provide a starting point to understanding not just whether representational coordination is likely to occur but also the cognitive processes involved in the complex, yet crucial task.

References

Ainsworth, S. E. (1999). The functions of multiple representations. *Computers & Education, 33*(2-3), 131-152.

Ainsworth, S. E., Bibby, P., & Wood, D. (2002). Examining the effects of different multiple representational systems in learning primary mathematics. *Journal of the Learning Sciences, 11*(1), 25-61.

Ainsworth, S. E., & Loizou, A. T. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science, 27*(4), 669-681.

Ainsworth, S. E., & Peevers, G. J. (2003). The Interaction between informational and computational properties of external representations on problem-solving and learning. In R. Altmann & D. Kirsch (Eds.), *Proceedings of 25th Annual Conference of the Cognitive Science Society.*

Ainsworth, S. E., & Van Labke, N (in press) Multiple forms of dynamic representation. Learning and Instruction

Bibby, P. A., & Payne, S. J. (1993). Internalization and the use specificity of device knowledge. *Human-Computer Interaction, 8*(1), 25-56.

Blackwell, A. F., & Engelhardt, Y. (1998). *A Taxonomy of diagram taxonomies.* Paper presented at the Thinking with diagrams, Aberystwyth.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*, 233-246.

Cheng, P. C. H. (2002). Electrifying diagrams for learning: principles for complex representational systems. *Cognitive Science, 26*(6), 685-736.

Cheng, P. C.-H. (1996). *Functional roles for the cognitive analysis of ciagrams in problem solving.* Paper presented at the Proceeding of the Eighteenth Annual Conference of the Cognitive Science Society, Hillsdale, NJ.

Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: a source activation account. *Cognitive Science, 25*(3), 315-353.

de Jong, T., Ainsworth, S. E., Dobson, M., van der Hulst, A., Levonen, J., Reimann, P., Sime, J., van Someren, M., Spada, H., & Swaak, J. (1998). Acquiring knowledge in science and math: the use of multiple representations in technology based learning environments. In M. W. Van Someren & P. Reimann & H. P. A. Boshuizen & T. de Jong (Eds.), *Learning with Multiple Representations* (pp. 9-40). Amsterdam: Elsevier Science.

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*(2), 179-201.

Gilmore, D. J., & Green, T. R. G. (1984). Comprehension and recall of miniature programs. *International Journal of Man-Machine Studies, 21*, 31-48.

Goldman, S. R. (2003). Learning in complex domains: when and why do multiple representations help? *Learning and Instruction, 13*(2), 239-244.

Guercin, F. (2001). Can children process complex information from different media. In J.-F. Rouet & J. L. Levonen & A. Biardeau (Eds.), *Multimedia learning* (pp. 59-64). Amsterdam: Pergamon.

Howe, C., Tolmie, A., Anderson, A., & Mackenzie, M. (1992). Conceptual knowledge in physics: The role of group interaction in computer-supported teaching. *Learning and Instruction, 2*, 161-183.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*(4), 351-371.

Klein, P. D. (2003). Rethinking the multiplicity of cognitive resources and curricular representations: alternatives to 'learning styles' and 'multiple intelligences'. *Journal of Curriculum Studies, 35*(1), 45-81.

Kozma, R. (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. *Learning and Instruction, 13*(2), 205-226.

Kozma, R., Chin, E., Russell, J., & Marx, N. (2000). The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. *Journal of the Learning Sciences, 9*(2), 105-143.

Lane, P. C. R., Cheng, P. C.-H., & Gobet, F. (2000). CHREST+: A simulation of how humans learn to solve problems using diagrams. *AISB Quarterly, 103,*, 24-30.

Lewalter, D. (2003). Cognitive strategies for learning from static and dynamic visuals. *Learning and Instruction, 13*(2), 177-189.

Lowe, R. K. (2003). Animation and learning: selective processing of information in dynamic graphics. *Learning and Instruction, 13*(2), 157-176.

Mayer, R. E. (2001). *Multimedia Learning*. Cambridge: Cambridge University Press.

Mayer, R. E. (2003). The promise of multimedia learning: using the same instructional design methods across different media. *Learning and Instruction, 13*(2), 125-139.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82*(4), 715-726.

Mayer, R. E., & Moreno, R. (1998). Split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312-320.

Mayer, R. E., & Sims, V. K. (1994). For whom Is a picture worth 1000 Words - extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*(3), 389-401.

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92*(1), 117-125.

Najjar, L. (1998). Principles of educational multi-media user interface design. *Human Factors, 40*(2), 311-323.

O'Hara, K. P., & Payne, S. J. (1998). The effects of operator implementation cost on planfulness of problem solving and learning. *Cognitive Psychology, 35*(1), 34-70.

Pane, J. F., Corbett, A. T., & John, B. E. (1996). Assessing dynamics in computer-based instruction., *Proceedings of ACM CHI'96 Conference on Human Factors in Computing Systems.* Vancouver.

Price, S. J. (2002). *Diagram Representation: The Cognitive Basis for Understanding Animation in Education* (Technical Report CSRP 553): School of Computing and Cognitive Sciences, University of Sussex.

Reimann, P. (2003). Multimedia learning: beyond modality. *Learning and Instruction, 13*(2), 245-252.

Rieber, L. P. (1990). Animation in Computer-Based Instruction. *Educational Technology Research and Development, 38*(1), 77-86.

Roberts, M. J., Gilmore, D. J., & Wood, D. J. (1997). Individual differences and strategy selection in reasoning. *British Journal of Psychology, 88*, 473-492.

Romero, P., du Boulay, B., Lutz, R., & Cox, R. (2003). The effects of graphical and textual visualisations in multi-representational debugging environments., *Proceedings of 2003 IEEE Symposia on Human Centric Computing Languages and Environments*.

Roth, W.-M., & Bowen, G. M. (2001). Professionals read graphs: A semiotic analysis. *Journal for Research in Mathematics Education, 32*, 159-194.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies, 45(2)*, 185-213.

Scanlon, E. (1998). How beginning students use graphs of motion. In M. W. Van Someren & P. Reimann & H. P. A. Boshuizen & T. de Jong (Eds.), *Learning with multiple representations* (pp. 9-40). Amsterdam: Elsevier Science.

Schoenfeld, A. H., Smith, J. P., & Arcavi, A. (1993). Learning: The microgenetic analysis of one student's evolving understanding of a complex subject matter domain. In R. Glaser (Ed.), *Advances in Instructional Psychology* (Vol. volume 4). Hillsdale, NJ: LEA.

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction, 13*(2), 227-237.

Snow, R. E., & Yalow, E. (1982). Education and intelligence. In R. J. Stenberg (Ed.), *A Handbook of Human Intelligence* (pp. 493-585). Cambridge: Cambridge University Press.

Stern, E., Aprea, C., & Ebner, H. G. (2003). Improving cross-content transfer in text processing by means of active graphical representation. *Learning and Instruction, 13*(2), 191-203.

Svendsen, G. B. (1991). The influence of interface style on problem-solving. *International Journal of Man-Machine Studies, 35*(3), 379-397.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Tabachneck, H. J. M., Leonardo, A. M., & Simon, H. A. (1994). *How does an expert use a graph ? A model of visual & verbal inferencing in economics.* Paper presented at the 16th Annual Conference of the Cognitive Science Society, Georgia Institute of Technology, Atlanta, Georgia.

Tabachneck-Schijf, H. J. M., Leonardo, A. M., & Simon, H. A. (1997). CaMeRa: A computational model of multiple representations. *Cognitive Science, 21*(3), 305-350.

Tabbers, H. K., Martens, R. L., & Van Merriënboer, J. J. G. (2001). The modality effect in multimedia instructions. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd annual conference of the Cognitive Science Society* (pp. 1024-1029). Mahwah, NJ: Lawrence Erlbaum Associates.

Van Labeke, N., & Ainsworth, S. (2001). Applying the DeFT framework to the design of multi-representational instructional simulations. In J. D. Moore & C. L. Redfield & W. L. Johnson (Eds.), *Proceedings of the 10th International Conference on AI in Education* (pp. 314-321). Amsterdam: IOS Press.

Van Labeke, N., & Ainsworth, S. E. (2002). Representational decisions when learning population dynamics with an instructional simulation. In S. A. Cerri & G. Gouardères & F. Paraguaçu (Eds.), *Intelligent Tutoring Systems: Proceedings of the 6th International Conference ITS 2002* (pp. 831-840). Berlin: Springer-Verlag.

Van Someren, M. W., & Tabbers, H. (1998). The role of prior knowledge qualitative knowledge in inductive learning. In M. W. Van Someren & P. Reimann & H. P. A. Boshuizen & T. de Jong (Eds.), *Learning with multiple representations* (pp. 102-119). Amsterdam: Pergamon.

White, B. (1993). ThinkerTools: Causal models, conceptual change, and science education. *Cognition and Instruction, 10*(1), 1-100.

Winn, B. (1987). Charts, graphs and diagrams in educational materials. In D. M. Willows & H. A. Houghton (Eds.), *The psychology of illustration: I. Basic research* (pp. 152-198). New York: Springer.

Yerushalmy, M. (1991). Student perceptions of aspects of algebraic function using multiple representation software. *Journal of Computer Asssisted Learning, 7*, 42-57.

Zahn, C, Banquero, B. & Schwann, S (in press) Learning with hyperlinked videos – Design criteria and efficient strategies for using audiovisual hypermedia. Learning and Instruction