

Symposia

Ritter, F. E., & Young, R. M. (Eds.). (1998). Proceedings of the Second European Conference on Cognitive Modelling. Thrumpton (UK): Nottingham University Press. ISBN 1-897676-67-0

Cognitive Models at Work Symposium

Symposium Aims

Cognitive models are used in the design of aircraft and industrial plant; operator tasking; user interface design; and for operations research into the behaviour of complex sociotechnical systems. The purpose of their use is to account for human performance in shaping work environments, developing cognitive aids, evaluating systems and designs, and predicting the outcomes of courses of actions. These models come from a number of intellectual traditions, and the papers included here are from and across disciplines. Rather than focusing on a particular model, this symposium seeks to explore some of the uses to which cognitive models are put, to find which models are being used and to draw some conclusions as where advances has been made and the technical challenges still in front of the cognitive modelling community.

Simon Goss

Programme

1. Introduction
2. *Learning and Testing Cognitive Models*
Simon Goss¹, Sam Waugh¹, Adrian Pearce², and Tim Menzies³
⁽¹⁾DSTO-Air Operation Division,
⁽²⁾Computer Science-Curtin University
⁽³⁾Computer Science- UNSW, Australia
3. *The Cognitive Cockpit: The Application of a Cognitive Model*,
Simon Finne and Robert Taylor
DERA Centre for Human Sciences,
Farnborough, UK
4. *Military Applications of Cognitive Models with COGNET*
Floyd Glenn
CHI Systems, Inc., Spring House, PA, USA
5. *Exploiting Knowledge Engineering for the Construction of Cognitive Models*
Nigel Shadbolt
Psychology, Nottingham University, UK
6. Wrap-up

Learning and Testing Cognitive Models

Simon Goss (1), Sam Waugh(1), Adrian Pearce(2), and Tim Menzies(3)

(1) DSTO-Air Operation Division,
GPO Box 4331, Melbourne, 3001 Australia
{first.last}@dsto.defence.gov.au

(2) Computer Science-Curtin University
GPO Box 1987U , Perth, 6001, Australia,
adrianp@cs.curtin.edu.au

(3) Department of Artificial Intelligence,
School of Computer Science & Engineering, University of. NSW, Sydney, Australia, 2052,
tim@menzies.com

Introduction

In the military environment the physical systems components are there to facilitate operator mission objectives. While analysts have traditionally paid considerable attention to fidelity when modelling physical entities, the physical characteristics of system components are not sufficient to describe operational systems in sociotechnical environments; the human operators contribute significantly to systems outcomes [1]. In supporting operational usage after a capacity acquisition it is in the domain of mission parameters and operator procedures that the scope for change to improve performance lies. In operational research there has been a shift in focus from modelling an operator performing a task in an environment to modelling an entity with a social role performing actions in a dynamic social environment. This involves the recognition of the intentions of other entities. It could be said that the focus has shifted from computational theory of mind to computational theory of other minds.

Two aspects of modelling users are addressed in this presentation : the first, "grey box modelling" is applied to documenting a user's model of simulation software; the second concerns a method for learning to recognise the intentional behaviour of players or simulated agents in an agent-oriented virtual environment.

Testing Models

Grey box modelling, the process whereby a user by means of exploration develops a causal model of a partially understood system is the problem of legacy code maintenance and black box model commissioning. It is the process of acquiring expertise with a system to the point of function practicality. In our current work a method developed for the verification of knowledge based systems is applied to the testing and documentation of a developing user model of software [2-4]. The context is operations research where large models are used; often with large components externally sourced and less than well documented.

Considerable investment of staff time is required in learning and using these systems. An explicit documentation of the mental model the user has of the system has significant potential as a guide and aid to the acquisition of expertise, and the retention of this expertise independent of staff movement.

Learning Plans

In this work the experimental aim was to demonstrate a method of constructing procedures from spatio-temporal data which describe action plans of agent/entities in a virtual environment [5-6]. These are required for testing candidate operator intentions against operator action history, and are interpretable as partial instantiations of intentionality. The capacity of situation awareness possessed by human operators in dynamic social

systems requires is the recognition of plans whilst in execution in addition to than casual physical processes in train. A desirable incidental benefit is a summary method for the massive amount of data obtainable in a human-in-the loop simulation.

We explore this experimentally in the context of flight simulation, and offer a method for learning action plans. This requires three components: an appropriate ontology (model of operator task performance), an appropriate virtual environment architecture (accessibility of data and image generation databases) and a learning procedure (which relates the data stream to the domain ontology).

In simple terms, we are looking at the domain of circuit flight. We have a task analysis for circuit flight. The flight simulator has an authentic flight model for a PC9 aircraft, and a cockpit with generic throttle and stick controls. It also has a particular software architecture conferring special data recording properties. A relational learning technique is used to relate the data from the flight simulator to the task analysis. We build relations which describe generalised flight plan segments.

In practise these run in real-time and announce attributed plan segments *while* the pilot is executing them. This is a compelling demonstration of the feasibility of real-time recognition of intention in a user interface to an immersive virtual environment task. We assert that our results have wider significance and may form part of the foundation for the construction of agent-oriented simulations, and more broadly, virtual environments

References

- [1] Ian Lloyd and Simon Goss, *Simulating Human Characteristics for Operational Studies*, Proc of Australian Cognitive Science Conference, 1997,.
- [2] T. Menzies and S. Goss, *Applications of Abduction #3: "Black-box" to "Gray-box" Models* in Proceedings of AI in Defence Workshop, AI'95, DSTO general document GD-0077, 1997.
- [3] Tim Menzies and Simon Goss Vague Models and Their Implications for the KBS Design Cycle, Proceedings PKAW '96: Pacific Knowledge Acquisition Workshop, 1996.
- [4] Waugh S.; Menzies T.J; Goss, G. Evaluating a Qualitative Reasoner. Advanced Topics in Artificial Intelligence: 10th Australian Joint Conference on AI; Abdul Sattar (Ed.); ISBN 3-540-63797-4. Springer-Verlag, pp 505-514, 1997..
- [5] Adrian Pearce, and Simon Goss, *Learning Action Plans in a Virtual Environment.*, In L. J. Hettinger and M. W. Haas, editors, Psychological Issues in the Design and Use of Adaptive Virtual Interfaces, Erlbaum Publishers Inc., Hillsdale, N.J., USA, (to appear 1998.)
- [6] Adrian Pearce, Terry Caelli, and Simon Goss, On *Learning Spatio-Temporal Relational Structures in Two Different Domains*, Third Asian Conference on Computer Vision - ACCV98 Hong Kong, January 1998.

The Cognitive Cockpit: The Application of an Adaptive Cognitive Model

SE Finnie and RM Taylor
Centre for Human Sciences
F131, DERA Farnborough
Hampshire, GU14 6TD, UK
+44 1252 394282
{sfinnie, rtaylor}@dra.hmg.gb

Abstract

In an increasingly complex and automated aircraft environment, aircrew tasks are now more cognitive than physical in nature. This has led to interest in the requirements for cognitive quality in aircrew systems, and the need for engineering principles to guide the design of cognitive tasks. In symbiotic systems where both human and system cognitive quality is necessary for effectiveness, research is needed into the requirements for cognitive control (strategic, opportunistic) and compatibility (usability, intuitiveness). Such joint cognitive systems require reliable, and adaptive, cognitive models.

DERA CHS is currently developing such a cognitive model which will provide guidance on pilot-system dialogue structures, and cognitive task specification. The model attempts to encapsulate the relationship between human and machine at different levels of control, communication, awareness, and behaviour, and draws upon contemporary psychological theories such as: Rasmussen (1983); Hollnagel (1996); Taylor (1988). The model will provide guidance on the nature of the relationship between human and system. For example, the model will indicate that at no time should the system remove the pilot's control. Instead, a process of critiquing is preferable where the system is able to critique the pilot's errors, and similarly, the pilot is able to critique, and improve, the Cognitive Cockpit's advice. This paper outlines the adaptive cognitive model and the factors that ensure that it is a practical, applicable, framework for implementing automation in the DERA CHS Cognitive Cockpit.

Rasmussen J (1983), *Skills, Rules, and Knowledge: Signals, Signs, and Symbols, and other distinctions in human performance models*, IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-13, No 13, 3.

Hollnagel E (1996), *Modelling the Orderliness of Human Action*, In "Cognitive Engineering in the Aviation Domain", Amalberti R, and Sarter N (Eds.)

Taylor MM (1988), *Layered Protocols for Computer-Human Dialogue I: Principles*, In "International Journal of Man-Machine Studies", 28, 175-218.

Military Applications of Cognitive Models with COGNET

Floyd Glenn

CHI Systems, Inc.

716 N. Bethlehem Pike

Spring House, PA, USA 19002

215-542-1400

floyd_glenn@chiinc.com

ABSTRACT

This paper presents an overview of three cognitive models developed with the COGNET (for Cognition as a Network of Tasks) methodology and toolset. The examples illustrate the broad range of applications for which such models are suitable. They include a model for an air defense gunner which was developed for the purpose of crewstation design evaluation. The second example is a set of models for the watchstanders in an advanced ship's combat information center which are being developed as part of an embedded intelligent training system. The last example is a model of an airborne anti-submarine warfare sensor operator which is being developed to support an intelligent interface for the sensor operator.

Keywords

cognitive model, design evaluation, training, intelligent interface, COGNET, CIC watchstander, air defense gunner, sensor operator.

INTRODUCTION

Development of a cognitive model for a person operating a complex system is always a daunting effort. At a minimum, the cognitive modeler must define the task procedures for system operation, the complete knowledge base that is relevant to performance of these tasks, including both general and task-specific knowledge, and the various component performance models which characterize each aspect of human task performance. Construction of these cognitive models typically entails use of specialized AI programming languages such as LISP and accordingly requires the support of highly trained computer scientists. The COGNET methodology and toolset for cognitive modeling (Zachary, Ryder & Hicinbothom, in press; Zachary et al., 1992) has been developed in order to facilitate the development of cognitive models with a minimum need for support from such computer specialists. COGNET offers an integrated model development environment with a graphical interface for goal and task representation. This paper presents an overview of the COGNET toolset and descriptions of three distinctly different types of application of COGNET for military systems. The three COGNET applications include the primary alternatives that have been conceived for applications of executable cognitive models — (1) detailed performance prediction for design evaluation, (2) an embedded cognitive model for

an intelligent training system, and (3) an embedded cognitive model for an intelligent operational interface.

DESIGN EVALUATION - GUNNER MODEL

The application of COGNET for design evaluation concerns the development of a simulation model for the operator of the U.S. Army's mobile air defense weapons system known as Avenger. The Avenger is an operational mobile Forward Area Air Defense (FAAD) element consisting of a High Mobility Multipurpose Wheeled Vehicle (HMMWV) having a rotatable turret and eight ground-to-air Stinger missiles. Avenger is manned by a driver and operated by a gunner. The gunner sits in the turret where he searches for air targets through a transparent windscreen and also with a forward-looking infrared (FLIR) display. Upon detecting a target, the gunner aims the turret optical site at the target by rotating the turret using a control yoke, and then, upon verification of a hostile identification, a missile can be fired using control buttons on the yoke. The simulation of the Avenger gunner was developed to operate in the software environment of a simulation-based trainer for the Avenger system, called the Avenger Institutional Conduct of Fire Trainer (Avenger ICOFT). This simulation effort was originally conducted in order to enable simulation-based evaluation of contemplated modifications to Avenger, but interest has also developed in the potential use of this simulation for DIS applications. The gunner model includes distinct component performance models for visual search, target detection and identification, target tracking, and associated equipment operation and decision making. The model was developed initially through a task analysis of gunner performance in the ICOFT and later through collection of detailed performance time data in the ICOFT to use for model parameter calibration.

TRAINING APPLICATION - CIC MODELS

The COGNET application for intelligent training involves a series of models to simulate both the behavioral and cognitive activities of the watchstanders comprising the Anti-Air Warfare (AAW) team in the Combat Information Center (CIC) on an Aegis-based Cruiser (see Zachary et al., 1997 for a more detailed summary). This was done to construct simulation-based Advanced Embedded Training Systems for shipboard team training. The various members of the AAW team must identify and appropriately respond to air tracks so as to maintain the self-defense of own-ship and any

protected assets. This is a particularly difficult task in complex geo-political environments characterized by low-intensity conflict such as the Persian Gulf. Models for four different watchstanders have been developed — the AAW Coordinator (AAWC), the Tactical Action Officer (TAO), the Electronic Warfare Supervisor (EWS), and the Identification Supervisor (IDS). Each simulation model is able to:

- * operate the actual watchstation (or a high-fidelity simulation) in the same manner and with the same level of performance as a human expert;
- * generate and respond to voice interactions with other members of the AAW team and other CIC personnel;
- * reason about and solve tactical problems as they arise; and
- * take appropriate tactical actions.

The simulations were built to support shipboard team training based on the embedded training simulation capability of the Aegis Weapon Control System. While an Aegis embedded simulation is running, the behavioral models work the simulation scenarios in parallel to human trainees, generating expert level behaviors that are used as a dynamic benchmark for diagnosing both the behavioral and cognitive performance of the trainees. This diagnosis is then used to provide (real-time or deferred) feedback. Other direct applications of these models include supporting mission rehearsal and tactics development.

INTERFACE APPLICATION - SENSOR OPERATOR MODEL

COGNET has been used in a planned design for an intelligent interface for the U.S. Navy's new SH-60R multi-mission helicopter, designated as the Task-Oriented Interface Layer (TOIL). TOIL is envisioned as separate from the basic crewstation interface planned for the SH-60R and is intended as an alternative means for the sensor operator SO to accomplish essentially all functions provided by the crewstation. TOIL is offered as an option to the SO who may use it as much or as little as seems appropriate given the knowledge and skills that that operator has with the crewstation and tactical domain. Thus, TOIL represents an alternative interface layer for operator interaction with the system. TOIL is implemented in the form of various menu and data windows that appear in the data strip region of the SO's mission display. TOIL is structured to provide interface options and guidance that are specifically tailored to the momentary tactical and task context, and is hence task-oriented. Additionally, TOIL will incorporate intelligent agent software which will enable automation of some interface or tactical functions as part of TOIL.

CONCLUSIONS

The three example applications of COGNET described above provide an indication of the diverse ways that cognitive models are beginning to contribute to complex military systems. With the emergence of such "main stream" applications, it is becoming increasingly important to provide tools and methods to facilitate such model developments. COGNET represents a candidate model development environment that is designed to support modeling of the kinds of real-time, multi-tasking jobs involved in military crewstations such as are described here, but it is also designed to be relatively easy to use by people with minimal special computer training. Clearly, there are still many further enhancements and aids that are feasible and warranted for COGNET, as well as other similar cognitive modeling environments, to improve their usability. Two such areas of particular current interest are a graphical visualization tool for the COGNET declarative knowledge base (i.e., blackboard) and an aid for COGNET-oriented knowledge engineering (e.g., for guiding interviews and supporting information management).

ACKNOWLEDGMENTS

The applications of the COGNET model described in this paper were supported by contracts N61339-97-C-0023 and N61339-96-C-0055 with the Naval Air Warfare Center Training Systems Division, by contract N00421-97-C-1133 with the Naval Air Systems Command, and contract DAAH01-95-C-R124 with the Army Missile Command. All views expressed in this paper are strictly those of the author.

REFERENCES

- Zachary, W., Ryder, J., & Hicinbothom, J. (in press). Cognitive Task Analysis and Modeling of Decision Making in Complex Environments. In J. Cannon-Bowers & E. Salas (Eds.), *Decision Making Under Stress: Implications for Training and Simulation*, Washington, DC: APA.
- Zachary, W., Ryder, J., Hicinbothom, J., Bracken, K. (1997). The Use of Executable Cognitive Models in Simulation-based Intelligent Embedded Training. *Proceedings of Human Factors Society 41st Annual Meeting*. Santa Monica, CA: Human Factors Society, pp. 1118-1122.
- Zachary, W., Ryder, J., Ross, L., and Weiland, M. (1992) Intelligent Human-Computer Interaction in Real Time, Multi-tasking Process Control and Monitoring Systems. in M. Helander and M. Nagamachi (Eds.). *Human Factors in Design for Manufacturability*. New York: Taylor and Francis, pp. 377-402.

Modelling Conceptual Changes in Mechanics: An Interdisciplinary Perspective

Stella Vosniadou¹, Christos Ioannides¹, Aggeliki Dimitracopoulou¹, Daniel Kayser²,
Marc Champesme², Floriana Esposito³, Giovanni Semeraro³, Donato Malerba³,
Stefano Ferilli³

ABSTRACT

The work reported here has been led in a collaboration which took place in the framework of "taskforce 1: Representation Change" of a european project "Learning in Humans and Machines" sponsored by the European Science Foundation during the years 1994-97.

This interdisciplinary and international collaboration has gathered Psychologists and Education Scientists, who collected and analyzed data about the knowledge of students in elementary mechanics, and who hypothesized mental models explaining the data; Computer Scientists specialized in Knowledge Representation who designed a language tailored to express the above models; and Computer Scientists specialized in Machine Learning, who investigated the behaviour of two systems on (part of) the data collected, and evaluated the relevance of the formal study of theory revision to the modelization of the conceptual changes that take place in students.

Keywords

conceptual change, force, knowledge representation, machine learning, mechanics, mental models, validation of cognitive models

INTRODUCTION

We report on a collaborative and interdisciplinary work, which took place in the framework of a project called "Learning in Humans and Machines" sponsored by the European Science Foundation. The objective of the authors in this research was to effectively bring together the know-how from the fields of cognitive psychology and machine learning in view of the fulfilment of two main goals. The first, mostly relevant for cognitive psychology, is to propose a theory of the development of knowledge acquisition in mechanics, with the help of computational models, clearly formalised and precisely testable. The second one, relevant for machine learning, is to obtain powerful guidelines for a more effective design strategy of learning systems, starting from the very basic issue of what knowledge they should handle and how to represent it.

All the research works that are presented have been conducted from the same data that has been collected on Greek students of various ages concerning their concept of force. The format of the present paper will be as follows. It will start with a short presentation of the empirical data

which led to the identification of a small number of mental representations of force in students ranging in age from 5 to 15 years of age. It will continue with a presentation of a computational model which tries to reproduce the conceptual changes that take place when children develop the concept of force with reference to the theoretical framework proposed by the psychology team. We will then proceed with another short presentation of a process model of the solution of three problems in mechanics by elementary school students before and after a six week experimental program of instruction in mechanics. It will be followed by a presentation of a computer model designed to represent accurately the characteristics of the psychological model.

MENTAL MODELS OF FORCE

(Christos Ioannides & Stella Vosniadou)

The purpose of the research reported in this section was to investigate the development of the concept of force in young children and propose a theoretical framework within which we can explain this development.

A total of 105 children ranging in age from 5 to 15 years were interviewed individually using a questionnaire consisting of 27 questions. Each question referred to a drawing depicting objects of different weights and sizes (e.g. big stones and big balloons vs. small stones and small balloons), some stationary and some moving, and were asked about the kinds of forces that were exerted on these objects. Following a methodology developed by Vosniadou and Brewer (1992; 1994), we were able to identify eight mental models of force which were used consistently by 70.6% of the students in order to answer the questions. The mental models of force are presented in Table 1. The younger children in our sample used mental model 1, according to which there is an *internal force* within big and heavy objects regardless of their kinetic state or position. Older children's responses (about 9 - 10 years), could be mostly explained by assuming that they had used mental model 4, according to which there is an *acquired force* only within moving objects. This force was thought to be imparted to the objects by an external agent which set them in motion and serves to explain this motion. Mental models 2 and 3 were based on combinations of the above two interpretations of force (internal and acquired). In contrast to model 4, the









¹ {svosniad@compulink.gr - cioannid@atlas.uoa.gr - adimitr@atlas.uoa.gr} - Department of Philosophy and History of Science - National and Capodistrian University of Athens - 37, John Kennedy street - Athens 161 21 - Greece









² {Daniel.Kayser@lipn.univ-paris13.fr - Marc.Champesme@lipn.univ-paris13.fr} - LIPN - Institut Galilée - Université Paris-Nord - Av. J.-B. Clément - 93430 Villetaneuse - FRANCE

³ {esposito@lacam.uniba.it} - Dipartimento di Informatica - Università degli Studi di Bari - via Orabona, 4 - 70126 Bari - Italy

children who used mental model 5 believed that there is a *force of push or pull* exerted on objects pushed or pulled by an agent (even in the absence of movement). The same interpretation of force is also present in model 7. Most of the children who had received instruction in mechanics developed mental model 6 according to which the *force of gravity* is exerted on all the objects. The force of gravity model was usually added to an already existing *acquired* force model. Various alternative interpretations of the word gravity have been identified. For example, some children believe that the force of gravity increases with the stability of the objects, others that gravity increases as the distance of an object from the ground becomes greater.

Table 1: Frequencies and percent of mental models of force as a function of grade.

Mental Models of Force	Kind/ garten	4 th grade	6 th grade	9 th grade
1. INTERNAL FORCE: There is an Internal Force within stationary and moving heavy objects  	40%	6.7%	0%	0%
2. INTERNAL and ACQUIRED FORCE: There is an Internal Force within stationary heavy objects - There is an Internal and an Acquired Force within heavy objects that are moving  	20%	26.4%	20%	0%
3. INTERNAL FORCE IN STATIONARY OBJECTS: There is an Internal Force within stationary heavy objects only  	13.3%	6.7%	0%	0%
4. ACQUIRED FORCE: There is an Acquired Force within moving objects only  	0%	6.7%	30%	10%

Mental Models of Force	Kind/ garten	4 th grade	6 th grade	9 th grade
5. ACQUIRED FORCE and FORCE OF PUSH/PULL: There is an Acquired Force within moving objects - Force from an external agent on all objects that are pushed or pulled by the agent  	0%	0%	13.3%	16.5%
6. GRAVITATIONAL FORCE and ACQUIRED FORCE: There is the force of gravity on all stationary and falling objects - There is the force of gravity and an acquired force within moving objects  	0%	3.3%	0%	39.6%
7. FORCE OF PUSH/PULL: There is a force only on objects that are pushed or pulled by an agent  	0%	0%	0%	3%
8. SUSPENDED and ACQUIRED FORCE: Objects at high and unstable positions have "more force" than objects at a lower or more stable position (Suspended Force)- There is an Acquired Force within moving objects  	6.7%	16.5%	13.3%	3.3%
9. Mixed	20%	33.3%	23.3%	26.4%

Mental models are assumed to be formed as children attempt to explain their everyday observations and information (verbal or other) they receive from the culture under the constraints of certain ontological and epistemological presuppositions, such as that force is a property of physical objects, and that the motion of an inanimate object requires an explanation in terms of a causal agent (see Figure 1). The process of conceptual change seems to be a slow affair that proceeds through the gradual suspension or revision of well entrenched beliefs and presuppositions. For example, the older children seem to have differentiated weight from force. Nevertheless, the presupposition that force is a property of physical objects

and that the motion of physical objects requires an explanation, do not seem to have been changed in the conceptual system of the 9th grader, despite the fact that these students have been exposed to systematic instruction in Newtonian mechanics.

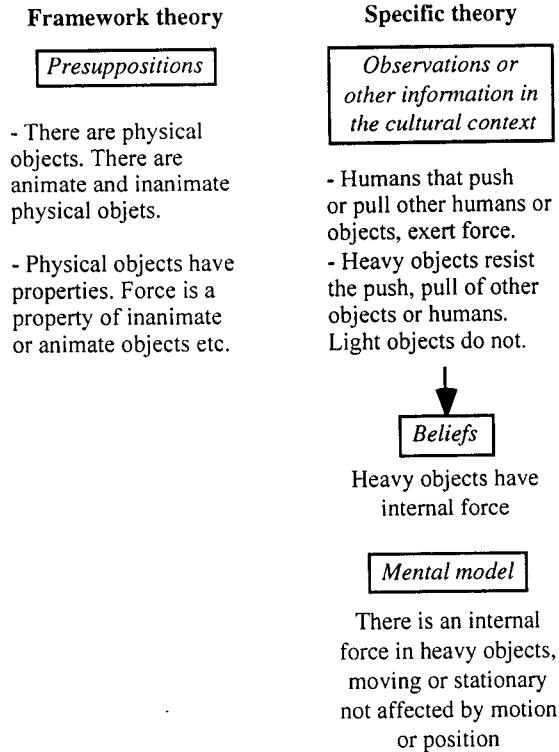


Figure 1: Hypothesised conceptual structure underlying the internal force mode.

ELABORATION OF A MACHINE LEARNING MODEL

(Floriana Esposito, Giovanni Semeraro, Donato Malerba and Stefano Ferilli)

From the Machine Learning perspective, the research focuses on the elaboration of a computational model which tries to reproduce the conceptual changes that take place when children develop the concept of force with reference to the theoretical framework proposed by the psychologist team.

A fundamental characteristic in the use of mental models concerns the possibility of verifying the general validity of a reasoning process based on examples by generating a sequence of significant examples and by applying revision procedures on the models (Johnson-Laird, 1983). Revision processes triggered by inductive mechanisms are an important aspect of learning. The research focused on the elaboration of a computational model of learning based on theory revision. The main objective of the work was to prove the validity of two particular Machine Learning systems: whether they are able to simulate the very complex phenomena related to the process of acquiring concepts of naive physics by creating these conceptualizations and refining them on the ground of new evidences. This could be useful in order to supply an artificial experimental laboratory where to test some of the mental models proposed by the psychologists, by observing the variations in the behaviour of the

computational models, monitoring the process of concept acquisition and refinement.

The proposed computational model considers learning as a process of formation and revision of a logical theory, where a logical theory is viewed as a set of conditions that are necessary and sufficient to explain a number of observations in a given environment. To be useful a theory must be able to explain past events and also predict future situations in the same environment.

A set of concept definitions constitutes a theory: theories are represented as sets of facts and rules, both strict and defeasible, sufficient for proving or explaining how an instance of a concept meets the concept definition. The instances from which a theory is inferred are called the training examples: these may be positive or negative. If we assume that the only source of knowledge available is represented by a set of examples and no prior knowledge can be exploited, the process of formulating a new theory is bound to be progressive. Starting from contingent observations, it is not possible to define concepts that are regarded as true. New observations can point out the inadequacies in the current formulation of the concepts. In such a case, a process of theory revision should be activated. Revisions of a logical theory are caused by a shift in the language and a change in the number and meaning of the involved concepts. In the proposed computational model the theory is refined incrementally: this is necessary when either incomplete information is available at the time of the generation of the initial theory or the nature of the concepts evolves dynamically.

Artificial learning systems can be roughly subdivided into batch and incremental, depending on whether all the examples used to train the system are completely available at learning time (batch) or not (incremental). Incremental learning systems maintain the inferred set of concept definitions consistent with all data (examples or observations) and have to store all previous data as soon as any backtracking is involved. In order to simulate the cognitive models of conceptual change in children learning elementary dynamics, two Machine Learning systems were used: ATRE and INTHELEX. The former is a batch system, while the latter is a fully incremental learning system.

The aim of the study was to see whether learning systems which learn from positive and negative examples by inductive inferential mechanisms could reproduce the changes in the concept of force observed in children. It has been suggested that children develop their concept of force on the basis of their interpretations of observations and information from their cultural background. Given some empirically derived hypotheses about the development of the concept of force, it was possible to extract the kinds of observations and/or information that are needed for the development to take place. These observations were used to validate the inferential power of the above mentioned learning systems.

Two experiments were run. In the first experiment, since ATRE can realize a shift in the representation language, the aim was that of discovering whether the system was able to relate the concept of force corresponding to "internal force in stationary and moving objects" to that corresponding to "acquired force in moving objects only". For humans the problem of learning dependent concepts

is related to the possibility of having an ontology. For machine learning systems the two ways of solving this problem are to supply the system a graph representing the concept dependency or to leave the system discover the dependency while it learns the concepts. ATRE uses both the approaches and some interesting results have been obtained related to the autonomously defined order by which it generalizes the concepts.

The second experiment concentrated on the process of theory revision; INTHELEX was used to emulate the transitions occurring in the human learning process when, starting from an empty theory and providing just an observation a time, it is possible to model and to monitor the refinement process of a theory. Some initial interesting results have been obtained although a direct comparison with the children acquisition mechanisms is not possible.

The batch system allowed us to point out how the formulation of the naive concepts of force is based in part on everyday experience, observations and verbal information and to prove that the dependence between some basic concepts of force can be modeled by a shift in the representation language. The incremental learning system, compared to the batch learning system, seems to be more accurate in performing the conceptualization process, for two basic reasons:

- changes of the initial theory caused by a new observation go through a process of refinement and it is not necessary to re-start the whole learning process from the beginning when a new instance is presented;
- it can take into account temporal relations albeit in a very simplistic way.

Both learning systems were able to produce from examples concepts related to the two models of "internal" and "acquired" force which were found in the developmental studies, although it is clear that students create their concepts only on the basis of observations or only being told about "force". The two systems tried to develop the two models of force through generalization and specification mechanisms. This may be compared with the phenomenon of "tuning" in conceptual change: indeed, both systems try to maintain coherence in the logical theory through their operators. This is an initial very simple form of conceptual change, although only a process of restructuration of knowledge should be considered a real conceptual change.

A PSYCHOLOGICAL PROCESS MODEL OF THE SOLUTION OF MECHANICS PROBLEMS BY ELEMENTARY SCHOOL STUDENTS

(Stella Vosniadou, Christos Ioannides and Aggeliki Dimitracopoulou)

The present project is based on collaborative and interdisciplinary work with a team of computer scientists from the LIPN (Daniel Kayser and Marc Champesme). The psychology team worked on a model that explained the solution of mechanics problems by elementary school children while and the computer science team validated this model by constructing a computer program. In previous work (Ioannides and Vosniadou, 1993; submitted) we used the theoretical model and methodology described in a series of studies on knowledge acquisition in astronomy (Vosniadou and Brewer, 1992;

1994; Vosniadou, 1994), to investigate the development of the concept of force. The results showed that it is possible to classify approximately 70% of the students in our sample as making use of one relatively well-defined mental model of force which they used in a logically consistent way to answer a number of questions about force. More specifically, six mental models of force were identified. These models were used in different frequencies by students ranging in age from 5 years (kindergarten) to 15 years (9th grade). These models are described in the previous section "Mental models of force".

In the present work we used these models to see if they could explain 5th grade students' responses to problems in mechanics, such as the one presented in Figure 2.



These two stones are just standing there. Is there a force exerted on them ?

Figure 2: Question 1.

The results showed that children's responses could be explained by assuming that the students used one of four models of force. When they gave responses such as "Yes, a force is exerted on the stones because they are big/heavy, etc." we assumed that they used the *internal force* model. On the contrary, if they said "No force is exerted on the stones because they do not move", we assumed that they used the *acquired force* model. Some students said that there is no force exerted on the stones "because the man does not push them". We assumed that these responses indicated use of the *push/pull model*, according to which a force is exerted when an (usually animate) object pushes or pulls another (usually inanimate) object. Finally, some students said that the force of gravity is exerted on the stones (*force of gravity model*). Students' responses to question 1 and the assumed mental models assigned to their responses are presented in Table 2.

Table 2: Students' responses to question 1.

	Response types	Assumed mental model	%
1.	Yes, a force is exerted on both stones because they are big/heavy they have weight/force	Internal force	18.4%
2.	No force is exerted on the stones because they are not moving	Acquired force	26.5%
3.	Yes, a force is exerted on both stones because the earth pulls/attracts them	Force of gravity	20.3%
4.	No force is exerted on the stones, because the man does not push them/exerts effort.	Push/pull	30.6%
6.	Other		4.2%

The results of this study also showed that the above mentioned models of force were not mutually exclusive and that the probability of activating them was influenced by the verbal and pictorial content of the specific questions asked. There were noticeable changes in the frequency of use of the different models in the different questions by the same subject population. The co-existence of different models of force in the same subject raised the issue of internal consistency. In previous work (Vosniadou & Brewer, 1992; 1994; Ioannides & Vosniadou, submitted) we have argued that students are consistent in their use of not more than one mental model of the earth, of the day/night cycle, or even of force. Are the present findings contradictory to our previous claims?

We believe that it is possible to explain the present findings if we assume that in the conceptual system of the 5th grader force is categorised differently for animate versus inanimate objects, as shown in Figure 3. If the child considers the question to apply to animate objects, then the *push/pull* model is used. If, on the other hand, the question is interpreted to belong to inanimate objects, the *internal* or *acquired force* models are used. Such an interpretation would make it possible for the same child to use either the "animate" or the "inanimate" models of force in different contexts, but not in the same context. Our results confirmed this prediction.

The possibility of internal inconsistency still is possible, however, in the case of use of the two inanimate models of force even in different contexts. However, an examination of the data showed that only one child made use of both the *internal* and the *acquired* models of force (the internal force model in questions 1 and 2, and the acquired force model in question 3). All the other children were consistent in their use of only one "inanimate" model of force.

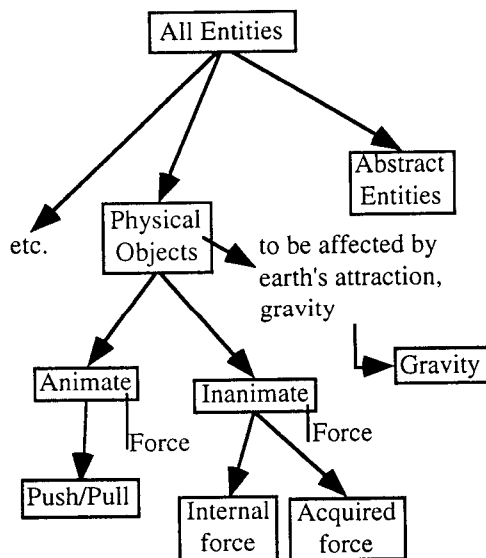


Figure 3: Assumed Categorization of the Concept of Force (for elementary school children).

The above leave unsolved the problem of how the mental model of the force of gravity is used. It appears to us that the gravity model comes through instruction to be added to the existing conceptual system of the 11 year old child and to be interpreted to apply to physical objects in

general. Thus, the gravity mental model can theoretically take the place of any animate or inanimate model of force. When contextual cues lead to the activation of the gravity model, the search stops there and the other mental models of force are not utilized. We understand that this is a very preliminary treatment of the notion of gravity. We know from previous work that there are various misconceptions of gravity. These issues are further investigated in ongoing work.

To conclude, force can be categorised in different places in the conceptual system of a 5th grader. It can appear under inanimate objects either as an inherent internal property (internal force) or as an acquired property (acquired force). It can appear under animate objects as the force exerted by a person's push or pull. Finally it may appear as a property of physical objects to be affected by the earth's attraction (gravity). These alternative representations of force become available as information comes through observation and from the culture in the form of systematic or unsystematic instruction. In previous work we described some of the beliefs and presuppositions about force that underlie these representations. In the present work we note that the different representations of force are associated with different contexts of use. Depending on the nature of the question and on the context, the student activates selected pieces of his or her prior knowledge to eventually create a specific mental representation of force on the basis of which he or she provides a response.

We believe that this work succeeds in capturing important aspects of the concept of force in young children, both in terms of how it is related to assumed underlying beliefs and presuppositions and in terms of its relationship to other concepts and categories.

A COGNITIVE MODEL OF ELEMENTARY SCHOOL STUDENTS' SOLUTION OF THREE PROBLEMS IN MECHANICS

(Daniel Kayser and Marc Champesme)

Modelling the knowledge state of students is an important objective for theoretical and practical (e.g. pedagogical) reasons. The model needs be validated and the best **validation** consists in implementing it and run the computer program on various questions in order to check whether the answers are identical - or at least, analogous - to those provided by the students.

In this section, we report on the experiment described in the previous section. The data have been analyzed by Cognitive Psychologists and the resulting models have been implemented jointly with Computer Scientists specialized in Knowledge Representation.

The Language

Recent work in Artificial Intelligence shows that the most difficult problem is to find appropriate trade-offs between the **efficiency** (of the inference procedures) and the **expressiveness** (of the representation language). Therefore, in this research, we attempt to **tailor the expressiveness to the exact needs** of the model, care being constantly taken that the algorithms using the knowledge represented remain tractable.

Early works in **Knowledge Representation**, such as KRL (Bobrow & Winograd, 1977; Lehnert and Wilks, 1979) also originated from a collaboration between AI and Cognitive Psychology. But their purpose was more

ambitious, because they aimed at a general framework for knowledge, therefore requiring maximal expressiveness, while we aim here at the minimal expressiveness compatible with the data.

The main line of further research (e.g. the KL-ONE family of languages (Brachman & Schmolze, 1985) followed by terminological and description logics) has emphasized on formal limitations in expressiveness in order to remain compatible with polynomial-time inferential mechanisms (a synthesis can be found in (Gottlob, 1996)).

More recently, research concerned with biological plausibility, as e.g. (Shastri, 1993) — notice being taken that biologists contest the relevance of this work, see discussion following (Shastri & Ajjanagadde, 1993) — have opened other insights in the efficiency vs. expressiveness trade-off. Papers by Fahlman (1979) and Shastri might have inspired very indirectly the present study.

Terminological Component

Concepts are structured in an ordinary “is-a hierarchy”, with multiple inheritance. Relations between concepts are represented by roles, which may have cardinalities. Less common, but still very important (e.g. in order to define at least a weak notion of consistency), is the ability to express the disjointness of concepts. Obviously, every statement of the language is translatable into first-order predicate logic, the reciprocal being false.

Assertional Component

The student is given a text from which (s)he is supposed to build a representation. For example, a sentence such as: “a man pushes a stone” is assumed to create an instance M of *man*, an instance S of *stone*, and an instance P of *push* having as arguments respectively M and S.

The assertion of an entity may be direct (entity supposed to be created while reading the text) or indirect (existence asserted as the consequent of an inference rule, see below).

Inference Rules ; Inference Engine

The students also entertain beliefs of the form, e.g. “when an agent xxx, then yyy”; this corresponds to the classical notion of inference rules.

Representing Several Models

Every piece of knowledge belongs to one or several of the mental models identified (see second section). A large part of the “is-a hierarchy” is model-independent (a stone is a physical object in every model), but some critical areas do depend on the model, e.g. the ontological nature of *force*.

We therefore associate to the internal representation of every concept, role, and rule, a **vector of boolean values**. The dimension of the vector is the number of models identified (currently, a dozen of them). Technically, we first provide the computer with the list of the names of all mental models. Each name is assigned an index in the vector. Then comes the description, in the language of the information (terminological component and inference rules) supposed to be valid in every model. It is compiled and stored with a vector filled with “true”. We then repeat a sequence composed of a list of the names of the model(s), followed by statements considered valid only in the models named in the list; the boolean

vector accordingly sets to “true” only the corresponding indices. Once all descriptions have been processed, we begin a “*session*” intended to simulate the behaviour of a student.

Implementing the Models in the Language

The implementation of the psychological models has been a long process with several feedbacks between Computer Scientists (CS) and Cognitive Psychologists (CP).

The first reason is that, from the CS point of view, a large part of the CP theories remains implicit, either because it constitutes the common knowledge assumed in the cultural background of CP, or because CP do not consider making it explicit as a valuable effort.

Another reason is that CS implementation resulted in making some aspects of the psychological theory more explicit, raising new important questions which needed be answered without ambiguity, and in some cases this led to some changes in the modelization (cf. subsection “Adding Psychologically Essential Features” in this section).

Refining the Ontology

Implementation first requires, once the representation formalism is designed, to translate the psychological theory into that formalism. Now the theory was initially expressed in very heterogeneous forms, ranging from rather general theoretical statements to concrete responses of students in natural language extracted from the experimental data.

The first proposed implementation was strongly guided by the most explicit parts of the theory. Therefore, it corresponded to a rather literal interpretation of the psychological data: many concepts were created in an attempt to capture all subtleties of the psychological models. In view of this preliminary modelisation, CP's feedback led to a pruning of the hierarchy of concepts, resulting in a clarification of which concepts were the most important, and what they meant.

After this clarification, all concepts were classified into three main categories: physical objects, which denote the concrete objects of the real world (e.g. human, stones...), propositions, which express statements about physical objects and abstract entities like measures which are neither concrete objects, nor statements about physical objects.

After these clarifications were completed, only minor changes revealed necessary in the ontology.

Adding Psychologically Essential Features

During the refinement of the first implementation, it turned out that some characteristics of the psychological models of the students were not represented accurately. As these features were considered as essential from the CP point of view, the CS had to modify their proposal.

This fact must be pointed out as an important result of this work, since these features were not explicitly stated in the initial model provided by CP, and would perhaps remain unnoticed otherwise. Another important point is that, although the representation issues constitute in general difficult problems for AI research, a careful analysis of the psychological model showed that several

features could eventually be represented, at least in this case, in a rather simple way.

Validation

Internal Validation

A compiler transforms the language into an internal form, performs several syntactic verifications (e.g. checks the well-formedness of the chain of roles), and goes beyond: using **partition** and **exclusive** statements, it checks that an entity does not inherit from two entities declared as incompatible with each other. Such checks proved useful to point at problems that were overlooked when writing the models.

A student can shelter simultaneously inconsistent beliefs, but in a given situation, (s)he should not use beliefs generating inconsistencies. Therefore, during a session, care is taken that every newly created entity is compatible with the is-a hierarchy, and obeys the cardinality restrictions declared in the **relation** statements.

External Validation

The above controls are more debugging aids than genuine validation. It is by far more important to compare the output of the program with the behaviour of a student supposed to work under the model(s) selected for a given session.

Obviously, some differences are irrelevant, as are also some similarities between student and computer answers. What matters is whether, for every situation in the domain of investigation, the computer outputs a result considered as plausible from a student supposed to use the corresponding model(s). Of course, this judgment is theoretically biased, since the models identified exist only in the theory. A better test, which has been used in (Chaignaud, 1996), consists in coding the student reactions in a way formally indiscernible from computer outputs, and to evaluate statistically whether experts are able to discriminate between human and computer protocols. Even this method is not completely immune to criticism. Anyhow, validating cognitive models raises deep epistemological issues, which we are not willing to develop further here.

Model Selection

After having declared which model(s) S the student is supposed to have access to, we describe the situation as a list of entities.

Introducing entities triggers inference rules. The information relevant to this process (both the existence of "is-a" links propagating the search for the rules, and the rules themselves) is indexed by the set M of models in which it is assumed to hold. Three cases are to be considered:

- M and S are disjoint: nothing happens;
- M contains S: the information is used;
- $\emptyset \subset M \cap S \subset S$; here, we need to know more about the influence of the context over the behaviour of the student. The only empirical data at hand being probabilistic, we select at random, obeying to the probabilities measured by CP, the (unique) model in which the student is assumed to reason in this case, and the decision of using or not the information is taken accordingly.

CONCLUSION

This research had two kinds of benefits:

- globally, models of the knowledge of students have been hypothesized, specified in a precise way, tested, and machine learning systems have been run in order to reproduce the acquisition of some concepts;
- locally, each team has taken advantage of the presence of the others in the following way:
 - the Psychologists have been forced to refine their models, and to resolve some inconsistencies which were not perceptible before the Computer Scientists had to implement them;
 - the analysis of the Psychologists has in turn influenced the design of a knowledge representation language having, per se, interesting features in terms of expressiveness and efficiency;
 - finally, the researchers in Machine Learning have been able to test their ideas on theory revision on real data.

Several other works concerning the change in understanding have been conducted in "taskforce 1". They are described in (Kayser & Vosniadou, in preparation).

Now that we have stronger tools to represent the knowledge state of a student, promising perspectives are opened to ask new questions about the evolution of this knowledge state under the influence of teaching, and the answer to these questions has obviously a great importance for Education.

ACKNOWLEDGMENTS

This work has been done in the framework of Project "Learning in Humans and Machines" funded by the European Science Foundation. Project leaders: Hans Spada, Univ. of Freiburg, Germany and Lorenza Saitta, Univ. of Turin, Italy.

REFERENCES

- Bobrow, D. G., & Winograd, T. (1977). An Overview of KRL, a Knowledge Representation Language *Cognitive Science*, 1, 3-46.
- Brachman, R. J., & Schmolze, J. G. (1985). An Overview of the KL-ONE Representation System *Cognitive Science*, 9, 171-216.
- Chaignaud, N. (1996). *Étude cognitive et informatique de la résolution d'un problème : analyse, modélisation et implantation*. Thèse d'Informatique, Université Paris-Nord.
- Fahlman, S. E. (1979). *NETL - A System for Representing and Using Real-World Knowledge*. M.I.T. Press.
- Gottlob, G. (1996). Complexity and Expressive Power of KR Formalisms. In *Proceedings of KR'96* (pp. 647-649). Cambridge Mass..
- Ioannides, C., & Vosniadou, S. (1993). *Mental models of force*. Paper presented at the Fifth European Conference for Research on Learning and Instruction. Aix-En-Provence, France.

- Ritter, F. E., & Young, R. M. (Eds.). (1998). *Proceedings of the Second European Conference on Cognitive Modelling*. Thrumpton (UK): Nottingham University Press. ISBN 1-897676-67-0
- Ioannides, C., & Vosniadou, S. (submitted). Aspects of the development of the concept of force. *Child Development*.
- Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge University Press, Cambridge.
- Kayser, D., & Vosniadou, S. (Eds.) (in preparation). *Change in Understanding in Humans and Machines: Case Studies in Physical Reasoning*.
- Lehnert, W. G., & Wilks, Y. (1979). A Critical Perspective on KRL, *Cognitive Science*, 3, 1-28.
- Semeraro, G., Esposito, F., Fanizzi, N., and Malerba, D. (1995). Revision of Logical Theories. in *Topics in Artificial Intelligence*, Lecture Notes in Artificial Intelligence 992, M. Gori and G. Soda (Eds.), Springer, 365-376.
- Shastri, L. (1993). A Computational Model of Tractable Reasoning — taking inspiration from cognition. In *Proceedings of the 13th IJCAI* (pp.202-207).
- Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning. *Behavioral and Brain Sciences*, 16, 417-494.
- Vosniadou, S. (1994). Capturing and modelling the process of conceptual change. *Learning and Instruction*, 4, 45-69.
- Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24, 535-585.
- Vosniadou, S., & Brewer, W. F. (1994). Mental models of the day/night cycle. *Cognitive Science*, 18, 123-183.

Implicit and explicit learning in ACT-R

Christian Lebiere

Department of Psychology
Carnegie Mellon University
Pittsburgh, USA
cl+@andrew.cmu.edu

Dieter Wallach

Department of Psychology
University of the Saarland
Saarbrücken, Germany
dwallach@cops.uni-sb.de

Niels Taatgen

Department of cognitive
science and engineering
University of Groningen
Groningen, Netherlands
n.a.taatgen@bcn.rug.nl

ABSTRACT

A useful way to explain the notions of implicit and explicit learning in ACT-R is to define implicit learning as learning by ACT-R's learning mechanisms, and explicit learning as the results of learning goals. This idea complies with the usual notion of implicit learning as unconscious and always active and explicit learning as intentional and conscious. Two models will be discussed to illustrate this point. First a model of a classical implicit memory task, the SUGARFACTORY scenario by Berry & Broadbent (1984) will be discussed, to show how ACT-R can model implicit learning. The second model is of the so-called Fincham task (Anderson & Fincham, 1994), and exhibits both implicit and explicit learning.

Keywords

ACT-R, implicit learning, explicit learning, skill acquisition, instance theory.

INTRODUCTION TO ACT-R

Knowledge Representation

ACT-R (Anderson, 1993; Anderson & Lebiere, in press) is a hybrid production system architecture for cognitive modeling. It is a hybrid architecture because it works at two interdependent levels: a symbolic level and a subsymbolic level. Each level is divided into a procedural and declarative component.

Symbolic Level

Declarative knowledge consists of chunks. Chunk structures are composed of a number of labeled slots, each of which can hold a value which can also be another chunk. Each chunk is an instance of a particular chunk type, which defines the name and number of slots. Procedural knowledge consists of productions. A production is a condition-action pair, which specifies the action to be taken if a particular condition is satisfied.

ACT-R is a goal-directed architecture. At any time, a goal is selected as the current focus of attention. Goals are organized on the goal stack, on which a goal can be stored (pushed) and later restored (popped). ACT-R operates in discrete cycles. At the start of each cycle, each production is matched against the state of the current goal. The productions that match enter the conflict set. A production is selected from the conflict set. The rest of the production condition can specify a number of chunk retrievals from declarative memory. If the retrievals are not successful, then the next production in the conflict set is selected. If the retrievals are successful, then the production action is executed. The action can modify the current goal, push it on the stack or pop it and restore a previous goal.

Subsymbolic Level

At the symbolic level, ACT-R operates in discrete, deterministic steps, but the subsymbolic level provides a measure of continuity and randomness. The previous section left two points unspecified: how are productions ordered in the conflict set, and if several chunks match a particular declarative retrieval, which is selected?

The productions are selected in order of decreasing expected utility. The current goal is assigned a value, or gain, equal to the worth of successfully achieving it. To each production is associated the probability and cost of achieving the goal to which it applies. The expected utility of a production applied to a goal is equal to the gain of the goal times the probability of success of the production, minus its cost. Noise is also added to the expected utility of a production, making production selection stochastic.

If several chunks satisfy a declarative retrieval, then the most active one is retrieved. The activation of a chunk is the sum of a base-level activation and an associative activation. The associative activation is spread from the sources of activation, which are the components of the current goal, to all related chunks in memory. Noise is added to each activation, making the retrieval of chunks stochastic. If no chunk activation reaches a retrieval threshold, then the retrieval fails. Furthermore, chunks which only partially match the retrieval pattern can also be retrieved, but their activation level will be penalized by an amount proportional to the degree of mismatch between the retrieval pattern and the actual chunk values.

Finally, the time to retrieve a chunk from memory is an exponentially decreasing function of its activation level. Therefore, although ACT-R operates in discrete cycles, the latency of each cycle, which is equal to the sum of the time to perform all the chunk retrievals plus the action time of the successful production, is a continuous quantity. Whereas the specification of an ACT-R model at the symbolic level has a precise, algorithmic quality, its operation at the subsymbolic level matches the stochasticity and continuity of human performance.

Learning

The previous section describes the performance of ACT-R assuming a certain state of knowledge. However, to provide an adequate model of human cognition, it is also necessary to specify how that knowledge was acquired. In ACT-R, knowledge is learned to adapt the system to the structure of the environment (Anderson, 1990; Anderson & Schooler, 1991).

Symbolic Learning

When a goal is popped, it becomes a chunk in declarative memory. That (and the encoding of environmental

stimuli) is the only source of declarative knowledge in ACT-R. The chunk resulting from a goal represents the statement of the task addressed by the goal and usually its solution. Therefore, the next time that task arises, its solution, depending upon the activation of the chunk, might be directly retrieved from declarative memory instead of being recomputed anew.

Productions are created from a special type of chunk called dependency. When a goal is solved through a complex process, a dependency goal can be created to understand how it was solved (e.g. which fact was retrieved or which subgoal was set). When that dependency goal is itself popped, a production is automatically compiled to embody the solution process. Thus the next time a similar goal arises, the production might be available to solve it in a single step instead of a complex process.

Symbolic knowledge is learned to represent in a single, discrete structure (chunk or production) the results of a complex process. Subsymbolic knowledge is adjusted according to Bayesian formulas to make more available those structures which prove most useful.

Subsymbolic Learning

When a production is used to solve a goal, its probability and cost parameters are updated to reflect that experience. If the goal was successfully solved, then the production probability is increased. Otherwise, it is decreased. Similarly, the production cost is updated to reflect the actual cost of solving that goal. Declarative parameters are adjusted in the same way. When a chunk is retrieved, its base-level activation is increased. The strength of association between the current sources and the chunk is also increased.

Subsymbolic knowledge does not result in new conscious knowledge, but instead makes the existing symbolic knowledge more available. Chunks which are often used become more active, and thus can be retrieved faster and more reliably. Productions which are more likely to lead to a solution and/or at a lower cost will have a higher expected utility, and thus are more likely to be selected during conflict resolution.

IMPLICIT LEARNING IN THE SUGARFACTORY TASK

Introduction

In contrast to rule-based approaches that conceptualize skill acquisition as learning of abstract rules, theories of instance-based learning argue that the formation of skills can be understood in terms of the storage and deployment of specific episodes or instances (Logan, 1988; 1990). According to this view, abstraction is not an active process that results in the acquisition of generalized rules, but that rule-like behaviour emerges from the way specific instances are encoded, retrieved and deployed in problem solving. While ACT-R has traditionally been associated with a view of learning as the acquisition of abstract production rules (Anderson, 1983; 1993), we present a simple ACT-R model that learns to operate a dynamic system based on the retrieval and deployment of specific instances (i.e. chunks) which encode episodes experienced during system control. It is demonstrated that the ACT-R approach can explain available data as well as an alternative model that is shown to be based on critical assumptions.

The Task

Berry & Broadbent (1984) used the computer-simulated scenario SUGARFACTORY to investigate how subjects learn to operate complex systems. SUGARFACTORY is a dynamic system in which participants are supposed to control the sugar production sp by determining the number of workers w employed in a fictitious factory. Unbeknown to the participants, the behavior of SUGARFACTORY is governed by the following equation:

$$sp_t = 2 * w_t - sp_{t-1}$$

The number entered for the workers w can be varied in 12 discrete steps $1 \leq w \leq 12$, while the sugar production changes discretely between $1 \leq sp \leq 12$. To allow for a more realistic interpretation of w as the number of workers and sp as tons of sugar, these values are multiplied in the actual computer simulation by 100 and 1000, respectively. If the result according to the equation is less than 1000, sp is simply set to 1000. Similarly, a result greater than 12000 leads to an output of 12000. Finally, a random component of ± 1000 is added in 2/3 of all trials to the result that follows from the equation stated above. Participants are given the goal to produce a target value of 9000 tons of sugar on each of a number of trials.

The models

Based on Logan's *instance theory* (1988; 1990) Dienes & Fahey (1995) developed a computational model to account for the data they gathered in an experiment using the SUGARFACTORY scenario. According to instance theory, encoding and retrieval are intimately linked through attention: encoding a stimulus is an unavoidable consequence of attention, and retrieving what is known about a stimulus is also an obligatory consequence of attention. Logan's theory postulates that each encounter of a stimulus is encoded, stored and retrieved using a separate memory trace. These separate memory traces accumulate with experience and lead to a „gradual transition from algorithmic processing to memory-based processing“ (Logan, 1988, p. 493). In the following, we contrast the Dienes & Fahey (1995) model (D&S model) with an alternative instance-based ACT-R model and discuss their theoretical and empirical adequacy.

Algorithmic Processing

Both models assume some algorithmic knowledge prior to the availability of instances that could be retrieved to solve a problem. Dienes & Fahey (1995, p. 862) observed that 86% of the first ten input values that subjects enter into SUGARFACTORY can be explained by the following rules:

- (1) If the sugar production is below (above) target, then enter a workforce that is different from the previous input by an amount of 0, +100, +200 (0, -100, -200).
- (2) For the very first trial, enter a work force of 700, 800 or 900.
- (3) If the sugar production is on target, then respond with a workforce that is different from the previous one by an amount of -100, 0, or +100 with equal probability.

While this algorithmic knowledge is encoded in the D&F model by a constant number of prior instances that could be retrieved in any situation, ACT-R uses simple production rules to represent this rule-like knowledge. The number of prior instances encoded is a free parameter in

the D&S model that was fixed to give a good fit to the data reported below. There is no equivalent parameter in the ACT-R model.

Storing Instances

Logan's instance theory predicts that every encounter of a stimulus is stored. The D&F model, however, does *only* store instances for those situations, in which an action successfully leads to the target; all other situations are postulated to be forgotten immediately by the model. Moreover, the D&S model uses a „loose“ definition of the target that was unavailable to subjects: While subjects were supposed to produce 9000 tons of sugar as *the* target state in the experiment, a loose scoring scheme was used to determine the performance of the subjects. Because of the random component involved in the SUGARFACTORY, a trial was counted as being on target if it resulted in a sugar production of 9000 tons with a tolerance of ± 1000 . The D&M model stores only instances that are successful in this loose sense and thus uses information about a range of target states that subjects were not aware of. ACT-R, on the other hand, encodes *every* situation, irrespective of its result. The following chunk is an example for an instance acquired by the ACT-R model as a restored goal.

```
(transition1239
  ISA transition
  STATE 3000
  WORKER 8
  PRODUCTION 12000)
```

The chunk encodes a situation in which an input of 8 workers, given a current production of 3000 tons, led to subsequent sugar production of 12000 tons. While the model developed by Dienes & Fahey (1995) stores multiple copies of instances, ACT-R does not duplicate identical chunks.

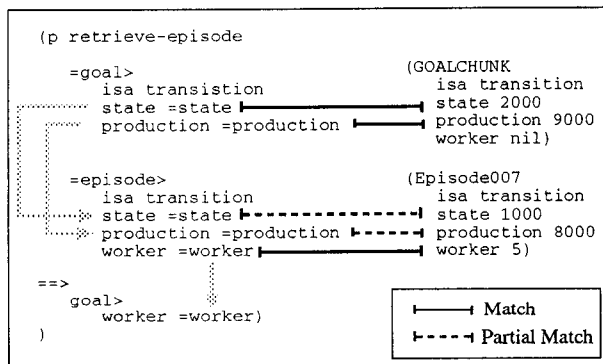


Figure 1. Matching process in the Sugar Factory model

Retrieving Instances

In the D&F model each stored instance „relevant“ to a current situation races against others and against prior instances representing algorithmic knowledge; the first instance after a finishing post determines the action of the model. An instance encoding a situation is regarded to be „relevant“, if it either matches the current situation exactly, or if it is within the loose range discussed above. As with the storage of instances, memory retrieval in the D&F model is again based on specific information not available to subjects. Retrieval in the ACT-R model, on the other hand, is governed by similarity matches between

a situation currently present and encodings of others experienced in the past (see Buchner, Funke & Berry, 1995 for a similar position in explaining the performance of subjects operating SUGARFACTORY). On each trial, a memory search is initiated based on the current situation and the target state '9000 tons' as cues in order to retrieve an appropriate intervention or an intervention that belongs to a similar situation. The production rule retrieve-episode (figure 1) is used to model the memory retrieval of chunks based on their activation level. Instances which only partially match the retrieval pattern, i.e. which do not correspond exactly to the present situation, will be penalized by lowering their activation proportional to the degree of mismatch. As a parameter of the ACT-R model, normally distributed activation noise is introduced to allow for some stochasticity in memory retrieval.

As figure 2 shows, the use of instances over the initial algorithmic knowledge increases over time, resulting in the gradual transition from algorithmic to memory-based processing as postulated by Logan (1988, p. 493).

Theoretical Evaluation

While both models of instance-based learning share some

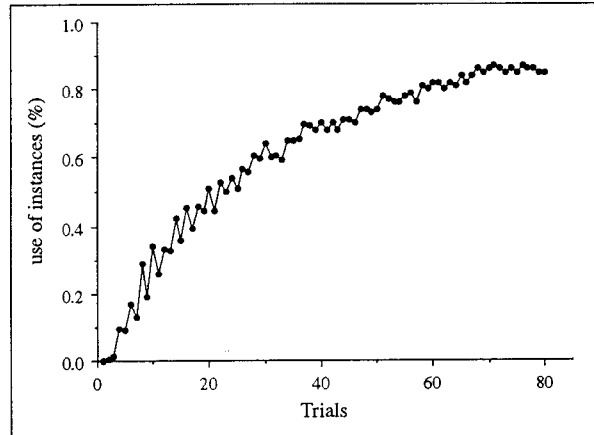


Figure 2. Relative use of instance retrieval per trial

striking similarities, the theoretical comparison has shown that the D&F-model makes stronger assumptions with respect to the storage and the retrieval of instances that can hardly be justified. Dienes & Fahey (1995) found out that these critical assumptions are essential to the performance of the D&F model (p. 856f):

„The importance to the modeling of assuming that only correct situations were stored was tested by determining the performance of the model when it stored all instances. ... This model could not perform the task as well as participants: The irrelevant workforce situations provided too much noise by proscribing responses that were in fact appropriate ... If instances entered the race only if they exactly matched the current situation, then for the same level of learning as participants, concordances were significantly greater than those of participants“.

Since the ACT-R model does not need to postulate those critical assumptions, this model can be regarded as the more parsimonious one, demonstrating how instance-based learning can be captured by the mechanisms provided by a unified theory of cognition.

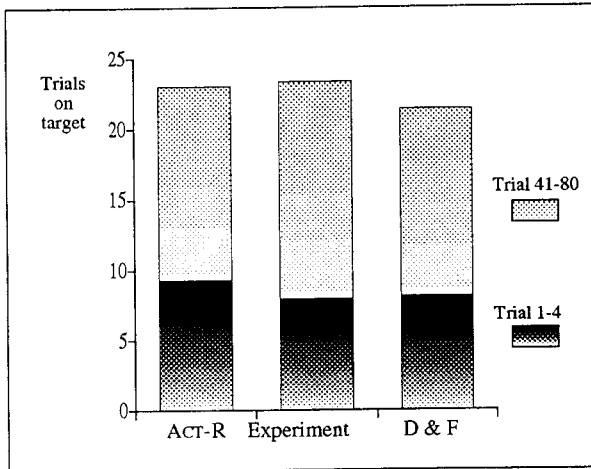


Figure 3. Results of the experiment, ACT-R model and D&F model

Empirical Evaluation

While the theoretical analysis of the assumptions underlying the two models has favoured the ACT-R approach, we will briefly discuss the empirical success of the models with respect to empirical data as reported by Dienes & Fahey (1995). Figure 3 shows the trials on target when controlling SUGARFACTORY over two phases, consisting of 40 trials each. ACT-R slightly overpredicts the performance found in the first phase, while the D&F model slightly underpredicts the performance of the subjects in the second phase. Since both models seem to explain the data equally well, we cannot favour one over the other.

Figure 4 shows the performance of the models in predicting the percentage of times („Concordance“) that the subjects gave the same (correct or wrong) response in a questionnaire as they did when controlling SUGARFACTORY. Again, both models seem to do a similar good job in explaining the data, with no model being clearly superior. Although space limitations do not allow for a detailed discussion, the picture illustrated by these two empirical comparisons remains the same after

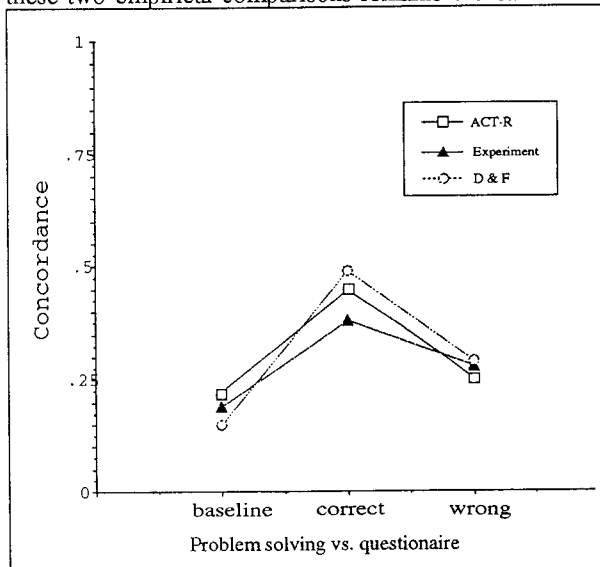


Figure 4. Concordances for the experiment and both models

several additional model comparison tests. We are currently running an experiment, exploring different predictions of the models in more details.

Conclusion

We discussed and compared a simple ACT-R model to an approach based on Logan's instance theory with respect to their ability to modeling the control of a dynamic system. While both models were similar in their empirical predictions, the ACT-R model was found to require fewer assumptions and is thus preferred over the model proposed by Dienes & Fahey (1995). Generally, ACT-R's integration of an activation-based retrieval process with a partial matcher seems to be a very promising starting point for the development of an ACT-R theory of instance-based learning and problem solving.

IMPLICIT AND EXPLICIT LEARNING IN THE FINCHAM TASK

The learning mechanisms in ACT-R are all quite basic, and can be used in several different ways to achieve different results. The idea of a learning mechanism as an integral part of an architecture has properties in common with the psychological notion of implicit learning. Both types of learning are considered to be always at work and not susceptible to change due to development or great variation due to individual differences. One of the defining properties of implicit learning, the fact that it is not a conscious process, is harder to operationalize within the context of an architecture for cognition. The closest you can get in an architecture is the notion that implicit learning is not guided by learning intentions, but is rather a by-product of normal processing. The SUGARFACTORY model discussed in the previous section is an example of implicit learning, since ACT-R uses old goals that are stored unintentionally to improve its behavior.

Explicit learning, on the other hand, is tied to intentions, or goals in ACT-R terms. Since there are no learning mechanisms that operate on goals, explicit learning can best be explained by a set of learned learning strategies. An example of a learning strategy to improve memorization of facts is using rehearsal to improve base-level learning. Base-level learning increases the activation of a chunk each time it is retrieved. If this increase of activation through natural use is not enough for the current goals, rehearsal can be used to speed up the process. By repeating a fact a number of times, its base-level activation can be boosted intentionally.

In this section we will discuss a paradigm for skill learning that involves both an implicit and an explicit strategy. The implicit strategy corresponds to instance-based learning, and the explicit strategy to rule-learning. Figure 5 shows an overview of this paradigm. First we assume that a participant has some initial method or algorithm to solve the problem. Generally this method will be time-consuming or inaccurate. Each time an example of the problem is solved by this method, an instance is learned. In ACT-R terms an instance is just a goal that is popped from the goal stack and is stored in declarative memory. Since this by-product of performance is unintentional, it can be considered as implicit learning.

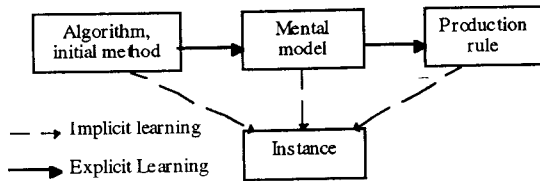


Figure 5. Diagram that illustrates the learning scheme used in the Fincham-task model

Other types of learning require a more active attitude from the participant. If the initial method is too time consuming, the participant may try to derive a re-representation of the information needed for the task to increase efficiency, which we will call, using Johnson-Laird's (1983) terminology, a mental model. If the initial method leads to a large number of errors, the participant may try to deduce or guess new relationships in the task in order to increase performance. The next step, from mental model to production rule, can only be made if the mental model is simple enough to convert to a production rule. Both the application of mental models and firing new production rules will create new instances. So regardless of what is going on due to explicit learning, implicit learning keeps accumulating knowledge.

So, if we have that many ways of learning, what type of learning will we witness in a particular experiment? To be able to answer this question we go back to the principle of rational analysis. According to this principle, we will principally witness that type of learning that will lead to the largest increase in performance. If we have task in which it is very hard to discover relationships or mental models, learning can probably be characterized primarily by implicit instance learning. Tasks in which there are too many instances too learn, but in which relationships are more obvious, will probably be better explainable by rule and abstraction learning. The SUGARFACTORY task is an example in which it is very hard to discover the rules that govern the system due to the influence of the previous sugar production and random factor in the output.

The Fincham Task

An example of a task in which both rule learning and instance learning are viable strategies is described by Anderson & Fincham (1994). In this task, participants first have to memorize a number of facts. These facts are in the form of

"Hockey was played on Saturday at 3 and then on Monday at 1."

We will refer to these facts as "sport-facts" to prevent confusion with facts and rules in the model. A sport-fact contains a unique sport and two events, each of which consists of a day of the week and a time. After having memorized these facts, participants were told the facts are really rules about the time relationships between the two events. So in this case "Hockey" means you have to add two to the day, and subtract two from the time. In the subsequent experiment, participants were asked to predict the second event, given a sport and a first event, or predict the first event, given the sport and the second event. So participants had to answer questions like: "If the first game of hockey was Wednesday at 8, when was the second game?" In this paradigm, it is possible to

investigate evidence for both rule-based learning and instance-based learning. Directional asymmetry, evidence for rule-based learning, can be tested for by first training a sport-fact in one direction (by predicting the second event using the sport and the first event), and then reverse the direction (by predicting the first event using the sport and the second event) and look how performance in the reverse direction relates to performance on the trained direction. If the performance is worse in the reverse direction, this is evidence for the use of rules. Evidence for instance learning can be gained by presenting specific examples more often than other examples. Better performance on these specific examples would indicate instance learning. Anderson & Fincham (1994), and later Anderson, Fincham & Douglass (1997) performed five variations on this basic experiment. The basic findings we will focus on are as follows:

- In general, reactions times improve according to the power law of practice, starting at around 35 seconds for the first few trials and improving to around 7 seconds at the third session.
- There is evidence for rule learning as witnessed by directional asymmetry. However, the effect only starts at the third or fourth session, and is relatively small.
- There is evidence for instance learning, since problems that are repeated more often than others are solved faster.
- Although it can not be inferred directly from the data, participants report they use abstract versions of the rules, for example by memorizing "Hockey day +2" and "Hockey time -2".

On basis of this evidence, Anderson et al. conclude that participants use four strategies: analogy, abstraction, rule and instance. The interesting question is what learning processes play a role in changing strategies. Each of the four strategies can be related to one of the learning stages from figure 5.

The analogy strategy is the initial strategy: first the memorized example that has the same sport as the new trial is recalled, the relationship in this example is determined, and this relationship is mapped on the current trial. Analogy is not very efficient, since it consists of many steps.

The abstraction strategy assumes the participant has created and memorized a mental model of the sport that corresponds to the current trial, like "Hockey day +2". The strategy involves retrieving and applying the abstraction, which is easier and faster than the analogy strategy.

The rule strategy assumes a production rule has been learned that can fill in the answer directly. An example of this rule is (variables are indicated by italics):

```

IF  the goal is to find the day of the second event
    the sport is hockey
    and the day of the first event is day1
    AND day1 plus two days equals day2
THEN put day2 in the second event slot of the goal
  
```

The rule strategy is more efficient than the abstraction strategy, since it requires only a single step in stead of two.

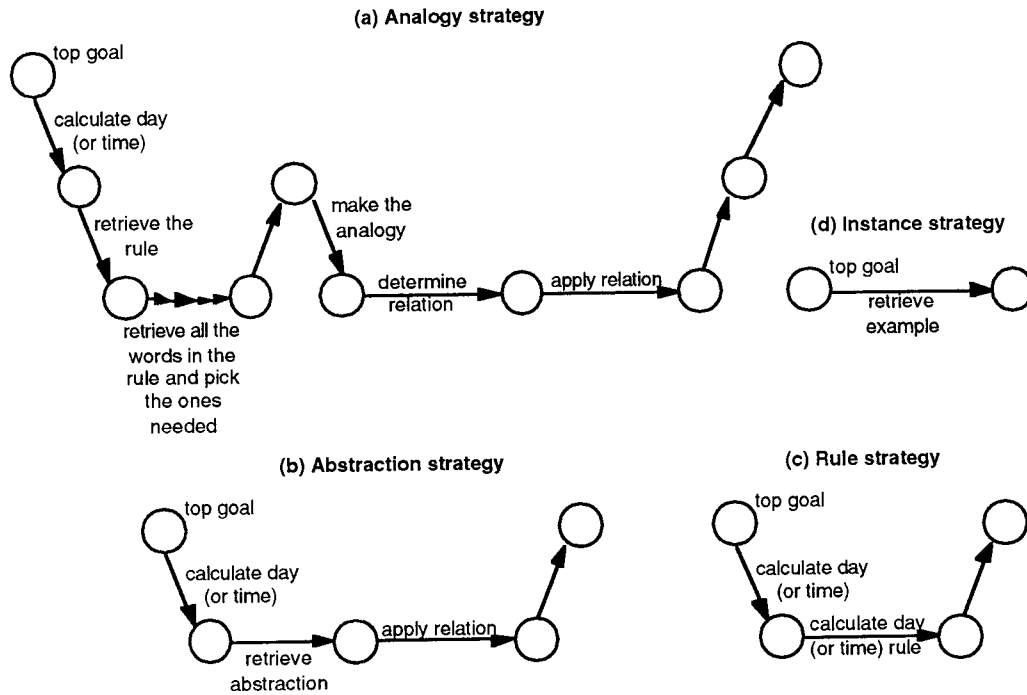


Figure 6. Overview of the four strategies in the Fincham task as modeled in ACT-R

The instance strategy assumes the answer can be given using a previous example. This previous example must be the same as the current trial. So an instance may contain the following information:

```

item1434
  isa instance
  sport hockey
  type day
  left sunday
  right tuesday
  
```

To use the instance strategy, it is sufficient to retrieve the right instance. This will of course only succeed if this instance is present in memory and is retrievable.

An ACT-R Model

We will now briefly discuss the ACT-R model of the task and its results. A more extensive discussion can be found in Taatgen & Wallach (in preparation). Figure 6 shows a schematic diagram of the implementation of the four strategies.

The analogy, abstraction and rule strategies are performed in a subgoal, that focuses on calculating either the day or the time. The instance strategy attempts to retrieve one of these subgoals, and fill in the answer directly in the topgoal. So learning instances is an implicit process in ACT-R, since past goals are always stored in declarative memory, an reoccurrence of the same goal just increases the activation of that goal. Knowledge for the other two strategies has to be acquired in an explicit fashion. An abstract mental model of a sport is no automatic by-product of the analogy strategy, so an explicit decision must be made to memorize an abstraction. To learn a new production rule in ACT-R, a special dependency structure must be created in declarative memory, which is also an explicit decision. In the current model, learning a new

production rule is only successful if there is already an abstraction present in declarative memory, else it is too difficult to collect the necessary information.

Results of the Model

In this paper we will only discuss results of the model on the second experiment of Anderson & Fincham (1994). In this experiment, participants had to learn eight sport-facts. In the first three days of the experiment, four of these sport-facts were tested in a single direction: two from left to right and two from right to left. On each day 40 blocks of trials were presented, in which each of the four sport-facts was tested once. On the fourth day all eight sport-facts were tested in both directions. On this day 10 blocks of trials were presented, in which each of the eight sport-facts was tested twice, once for each direction. Figure 7 shows the latencies in the first three days of the experiment, both the data from the experiment and from the model. The fit between the model and the data is quite good ($R^2=0.94$).

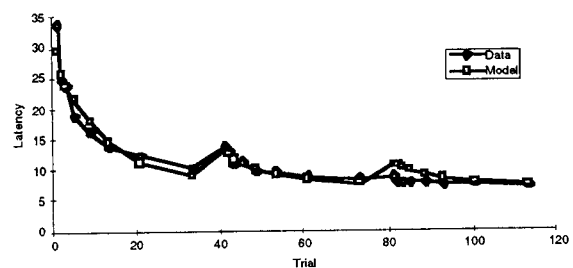


Figure 7. Latencies in experiment 1 for days 1-3

The results on day 4 can be summarized in the following table:

	Data	Model
Same direction, practiced	8.9 sec	8.4 sec
Reverse direction, practiced	10.9 sec	9.3 sec
Not practiced	13 sec	16 sec

Both in the data and in the model there is a clear directional asymmetry, since items in the practiced direction are solved faster than reversed items. The fact that unpracticed items are slower than the reversed items indicates that rule learning can not be a sufficient explanation for all of the learning in the first three days of the experiment.

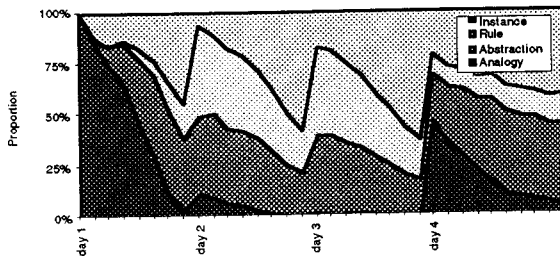


Figure 8. Strategy use in experiment 1 for days 1-4

Figure 8 shows how the model uses the four strategies in the course of the experiment. At the start of the experiment, analogy is used most of the time, but both the abstraction and the instance strategy gain in importance after a few blocks of trials. The rule strategy only appears later, and only plays a minor role during the first day. At the start of the second day, there is a large shift toward using rules at the expense of instances. This can be explained by the fact that the activation of a large portion of the instances has decayed between the two days, so that they can not be retrieved anymore. Since only few rules are needed for successful performance, they receive more training on average and are less susceptible to decay. Note that the abstraction strategy remains relatively stable between the days since it is also less susceptible to decay than the instance strategy. This pattern is repeated at the start of the third day, although the instance strategy loses less ground due to more extended training of the examples. At the start of the fourth day, the frequency of use of the analogy strategy goes up again, since there are no production rules for the new four sport-facts. The abstraction strategy can take care of the reversed items though, so in that case the expensive analogy strategy is not needed. This explains the fact that reversed items are still faster than completely new items.

Except for a model of this experiment, the model has successfully modeled two other experiments as well, using the same parameters. The following additional phenomena could successfully be explained:

- The reaction time for examples that are repeated more often is shortened, since instance learning is more successful and the facts it represents have a higher activation.

- Directional asymmetry increases between day 2 to 4, but decreases again on day 5. The model can explain this by the fact that by day 5 the instance strategy starts dominating the rule strategy.
- The results of the model concur with participant's reports on whether they use a rule or an example to solve a particular trial.

Conclusions

The ACT-R architecture is an ideal platform to study implicit and explicit learning. It not only allows insights in both types of learning separately, but, more importantly, also in the interaction between them.

REFERENCES

- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- Anderson, J.R. & Fincham, J.M. (1994). Acquisition of Procedural Skills From Examples. *Journal of experimental psychology: Learning, Memory, and Cognition*, vol. 20, no. 6, 1322-1340.
- Anderson, J.R., Fincham, J.M. & Douglas, S. (1997). The role of Examples and Rules in the Acquisition of a Cognitive Skill. *Journal of experimental psychology: Learning, Memory, and Cognition*, vol. 23, no. 4, 932-945.
- Anderson, J. R. & Lebiere, C.. (in press). *The atomic components of thought*. Mahwah, NJ: Erlbaum.
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396-408.
- Berry, D. & Broadbent, D.A. (1984). On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Experimental Psychology*, 36A, 209-231.
- Buchner, A., Funke, J. & Berry, D. C. (1995). Negative correlations between control performance and verbalizable knowledge: Indicators for implicit learning in process control tasks? *The Quarterly Journal of Experimental Psychology*, 48A, 166-187.
- Dienes, Z. & Fahey, R. (1995). Role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21 (4), 848-862.
- Johnson-Laird, P.N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard University Press.
- Logan, G.D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492-528.
- Logan, G.D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, 22, 1-35.
- Taatgen, N.A. & Wallach, D. (in preparation). Models of rule and instance-based skill learning.

