# Cognitive processes in solving variants of computer-based problems used in logic teaching

Tessa Eysink, Sanne Dijkstra, Jan Kuper

Running head: COGNITIVE PROCESSES IN SOLVING LOGIC PROBLEMS

# Cognitive Processes in Solving Variants of Computer-Based Problems Used in Logic Teaching

Tessa, H. S. Eysink, Sanne Dijkstra, and Jan Kuper

University of Twente, The Netherlands

Faculty of Educational Science and Technology

Department of Instructional Technology

University of Twente

P.O. Box 217

7500 AE, Enschede

The Netherlands

Tel.: +31 53 4893572 / 4893606

Fax: +31 53 4892895

E-mail: Eysink@edte.utwente.nl

*Eysink, T.H.S., Dijkstra, S., & Kuper, J. (2001). Cognitive processes in solving variants of computer-based problems used in logic teaching. Computers in Human Behavior, 17(1), 1-19. This is the author's version of the journal article, published in Computers in Human Behavior © Elsevier Ltd. 2001. http://dx.doi.org/10.1016/S0747-5632(00)00038-8*

## Abstract

The effect of two instructional variables, visualisation and manipulation of objects, in learning to use the logical connective, conditional, was investigated. Instructions for 66 first-year social science students were varied in the computer-based learning environment Tarski's World, designed for teaching first-order logic (Barwise & Etchemendy, 1992). For all instructional conditions, the scores on the transfer tests showed a significant increase in understanding the conditional. Visualisation, operationalised as presenting only formal expressions or a geometrical reality in addition to these, showed no differences on the transfer test. If only presented formal expressions, about half of the participants needed to make drawings of the objects, especially when the problems increased in complexity. The manipulation condition, in which the participants could either construct a geometrical world or were presented a fixed world, significantly influenced the participants' cognitive processes in solving the logic problems. The students worked affirmatively and were tempted to stay in familiar situations. The results support the authors' view that visualisation facilitates cognitive processing. Moreover, the results are congruent with Piaget's theory of the development of knowledge of formal science concepts from the action with objects.


Keywords: Logical Reasoning, Problem Solving, Cognitive Processes, Computer-Based Instruction, Visualisation, Manipulation.

The central concern of logic is the correctness of human reasoning. Reasoning occurs in all sciences and in all possible contexts. The rules of logic are valid in all these situations. To use the same rules in every possible situation, they must be formulated in a general way, that is, in such a way that they are not restricted to a given context. This makes logic abstract and general. Furthermore, the language of logic is formal. Agreements are made about symbols to be used and about the way these symbols can be connected to each other to form formulas in a formal and precise way.

Various studies have shown that a substantial part of all students have difficulties with learning and using these abstract and formal characteristics of logic. Students experienced logic education as being difficult, too abstract and boring (e.g., Goldson, Reeves & Bornat, 1993; Fung, O'Shea, Goldson, Reeves & Bornat, 1994). Besides this, Barwise and Etchemendy (1998) stated that students often saw logic as the manipulation of logical expressions by applying formal, meaningless rules. They do not get sufficient practice in finding the relation between abstract representations and real-life meanings and therefore have difficulties in applying abstract principles to everyday phenomena (White, 1993). This resulted in students not grasping any real understanding of the concepts and rules of logic.

In addition to this, studies have shown that abstract reasoning is difficult to improve. Only near-transfer effects (Cheng & Hollyoak, 1986) or effects of years of formal training (Lehman & Nisbett, 1990) have been found. Freudenthal (1991) supposed that abstract reasoning is difficult to improve, because common-sense ideas often obstruct scientific ideas. In everyday life, people develop naïve notions about logical reasoning. Sometimes, these often ill-defined concepts and rules do not meet the rules of logic. If the learners develop certain ideas, it is difficult to change their minds and to convince them they should replace the old (incorrect) knowledge with the new

(correct) knowledge. Teachers in logic are often confronted with the problem of how to teach students to solve logic problems and to translate the real world statements into formal statements. Moreover, they have to make clear to the students how to use the logical connectives, conjunction ($\wedge$), disjunction ($\vee$), conditional ($\rightarrow$) and negation ($\neg$). To clarify the meaning of logical expressions, Barwise and Etchemendy (1992) constructed a reality of computer-generated geometrical objects, which students could use to construct logical statements and to check whether these statements were true. This world of geometrical objects, which was labelled Tarski's World, was presented on a screen and could be constructed and manipulated by using a mouse. Tarski's World has been reviewed (Goldson, Reeves & Bornat, 1993) as being an easy and fun to use programme that is 'capable of teaching a great deal about a formal language, its interpretation, models, counterexamples and consequence'. Van der Pal & Eysink (1999) designed an instruction for learning formal logic in which Tarski's World was used. In addition to formal expressions, students could manipulate objects in this geometrical world. This instruction was compared to an instruction in which only formal logical expressions were given. Results showed that students who were given the experimental instruction, performed better on transfer tests than students who were given the formal instruction. In the experimental instruction, however, two instructional variables were confounded: (a) the use of a geometrical reality, and (b) the manipulation of the objects. Thus, it was possible that the effect could be caused either by one of the two variables, by both or by an interaction of the two. The purpose of the present study was to investigate which variable was critical. In this respect, two issues received attention. The first concerned the extent of visualisation, that is, the effect on students' performances of presenting a geometrical reality in addition to formal expressions. The second concerned the extent of manipulation, that is, the effect on students' performances of

manipulating the objects presented. Both instructional variables were supposed to facilitate the students' problem solving process and thus the development of the students' knowledge and skills.

## Abstraction and Reality

In order to learn formal logic, cognitive development has to be in a stage in which formal concepts can develop and in which the logical operators can be used in a meaningful way. Piaget (1970) called this stage the formal operational period and he claimed that it is reached at about the age of twelve. During this stage, children start to think abstractly. They can formulate hypotheses without actually manipulating concrete objects, and when more adept in this, they can test hypotheses mentally. They can generalise from a real object to another and from a real object to an abstract notion.

Many adults, however, still have problems learning abstract, formal concepts without any reference to real world objects. They are unable to solve formal problems, in which only symbols are used. They can only do mental operations with real (concrete) objects, events or situations. It was estimated that this is the case for 40 to 70 percent of all adults (Pintrich, 1990). Freudenthal (1991) also recognised the difficulties learners have when studying mathematics. He proposed to connect the formal, abstract mathematics to reality, so that the learners could infer these formal concepts from this reality. This led to the suggestion, that learners need to be offered concrete problem situations, which can be imagined and can be used to develop mathematical knowledge and skills, so that the learners will understand the concepts and be able to work with them.

To study the relationship between abstract concepts and reality, it has to be clear how reality is described in formal sciences, and what will be the best possible way to represent this relationship. The essential characteristic of logic is that it can be applied to all situations and to

all worlds. The world or set of worlds is the reality and the formal language describes this reality. The way in which this reality can be represented can range from a direct, everyday reality via an entirely pre-structured reality (e.g. geometrical figures) to complete abstraction (e.g. abstract mathematical objects, elements, sets). The drawback of learning in an everyday life context is that students are tempted to pay attention to irrelevant aspects of the problem. Students' prior knowledge consists of ideas about often ill-defined concepts and rules. When solving problems in an everyday life context, the students will use these naïve ideas in which pragmatic aspects as preferences, intuition, and hidden assumptions can play a role. Language is also permeated with conversational implicatures (Grice, 1975), that is, sentences often express suggestions without explicitly stating them. When learning logic by solving logic problems in an everyday life context, students will use their everyday life ideas and expressions about what is correct reasoning, whereas they should learn to abstract from the given context and learn to reason according to the rules of logic.

Giving reality as a complete abstraction also shows some drawbacks. Students then receive abstract, conceptual knowledge that is isolated from the situations in which this knowledge is normally used. Students will not always understand what the concepts and rules are about and the knowledge will not be imbedded into prior knowledge. They may only learn to shuffle the abstract symbols without comprehending what they are doing and why. When trying to overcome this abstraction by imagining concrete objects for abstract expressions, they will use their own situations and in doing so mistakes can be made.

Although a common goal of logic education is to be able to apply logic and to reason logically in all situations, the use of a geometrical world to learn to reason logically will probably support learning: it makes all operations possible and at the same time shows what happens when certain operations are applied. The world is completely defined, in such a way that

errors can be precluded, since irrelevant characteristics of the problem situation are left out of the context. Abstract principles are related to concrete meanings, so that meaningfulness and understanding can be reached. Because the context is controlled, unwanted characteristics of the context will have no influence. Stenning, Cox and Oberlander (1995) added to this that a geometrical world shares the property of specificity with the internal representations used by humans in their reasoning. It is the specificity of the world that makes it cognitively manageable and more concrete.

## A Short Description of Logical Notions

Logic is the science of (both human and machine) reasoning, which tries to discover conditions by which conclusions are justified and correct. In order to reach precision, contemporary logic is presented in a formal, mathematical way.

This paper uses the language of first-order logic. The (formalised) language of this logic contains names (a, b, …, x, y, …) to denote individual objects, and predicates to express properties of objects and relations between these objects. For example, Large is a predicate and Large(a) says that object a is large. Likewise, Larger(a, b) says that object a is larger than object b, and Between(x, y, z) expresses that object x is positioned between the objects y and z. Expressions of this form are called (elementary) propositions. They can be combined into more complex propositions by the following connectives: negation ($\neg$, not), conjunction ($\wedge$, and), disjunction ($\vee$, or), and conditional ($\rightarrow$, if … then …, sometimes called implication). For example, the formula Larger(a, b) $\rightarrow$ Smaller(b, a) says that "if a is larger than b, then b is smaller than a".

Propositions express states of affairs about the world and they can either be true or false. The connectives are truth functional, that is, the truth or falsehood of a complex proposition is completely determined by the truth or falsehood of the propositions of which it is composed, as described by the truth tables given in Table 1.

---

Insert Table 1 about here

---

So, for example, if p is true, then $\neg$p is false. And also, if p is true and q is false, then p $\wedge$ q is false, but p $\vee$ q is true. Notice that p $\rightarrow$ q is only false if p is true and q false. In all other cases p $\rightarrow$ q is true.

As mentioned above, the formal language contains names to denote objects. There are two kinds of names: constants (a, b, …) denote specific objects (it is assumed that one knows which objects are denoted), whereas variables (x, y, …) denote arbitrary objects (one does not know which objects are denoted). This difference is exploited by the two quantifiers, the universal quantifier ($\forall$, for all) and the existential quantifier ($\exists$, there is at least one). For example, $\forall$x Large(x) says that all objects x are large, whereas $\exists$x Large(x) expresses that there is at least one object x which is large. A more complicated example is $\forall$x (Cube(x) $\rightarrow$ Large(x)). Literally, this formula says that "every object x is large, if it is a cube". In everyday language this is normally said as "every cube is large", or "all cubes are large". In this formula all occurrences of the variable x are said to be within the scope of the quantifier, and all occurrences of x are bound by the quantifier. If a variable is not bound by a quantifier, it is called free. If a formula contains a free variable, one cannot know whether this formula is true or false, since it is not known which object is denoted by this variable. For example, we do not know whether Large(x) is true or

false, since we do not know which object is denoted by x. On the other hand, when x is bound by a quantifier, as in $\forall x\ \mathrm{Large}(x)$, we can know whether this formula is true or false, since now x ranges over all objects, and for each individual object it can be checked whether it is large or not. In general, one can determine whether a formula is true or false whenever that formula does not contain free variables. Such formulas are called sentences.

## Tarski's World

The computer-based learning environment Tarski's World was designed by Barwise and Etchemendy (1992). By defining a world of visible, geometrical objects with certain characteristics and relations, users learn semantic structures as studied by logic and they learn to determine the truth value of formulae. Within the programme, all situations are completely defined and the programme is able to provide feedback on syntactic aspects and the truth of logical formulae. A typical example is shown in Figure 1.

---

Insert Figure 1 about here

---

The problems that the students have to solve lead to the construction of logical expressions. By examining types of errors and sequences of errors, the students' reasoning process could be mapped out. For example, in Figure 1 sentence 1, the problem "There is a small tetrahedron or a medium dodecahedron" leads to the logical expression

$$\exists x\ ((\mathrm{Tet}(x) \wedge \mathrm{Small}(x)) \vee (\mathrm{Dodec}(x) \wedge \mathrm{Medium}(x))) \qquad (1)$$

If programmed accordingly, the programme is able to register all errors the students made when constructing a formula and when reasoning. As can be seen from the T's and F's at the left of the expressions in the sentence module, two of the displayed expressions are true, whereas one is false. Expression 1 claims that there must be a small tetrahedron or a medium dodecahedron. As object a is a small tetrahedron, the sentence is true in the world given. But as can be seen in the inspector module, the student thought Expression 1 was false. The programme provided feedback by telling the student that the expression was indeed syntactically correct, that it was indeed a sentence, but that the answer that the sentence was false, was incorrect. These three errors, (a) whether the logical expression was syntactically correct, (b) whether the logical expression was a sentence, and (c) whether the truth value given by the student to this logical expression was correct, were being recognised by the feedback mode of Tarski's World.

Another example in Figure 1 is sentence 2, "All cubes are large" which has a hidden conditional. Another way of saying "All cubes are large" is "For all objects it holds that, if it is a cube, then it is large". The correct translation is given in the logical expression

$$\forall \underline{x}\ (\text{Cube}(\underline{x}) \rightarrow \text{Large}(\underline{x})) \qquad (2)$$

However, many subjects do not recognise the hidden conditional and render "All cubes are large" as

$$\forall \underline{x}\ (\text{Cube}(\underline{x}) \wedge \text{Large}(\underline{x})) \qquad (3)$$

which means “All objects are large cubes”. In this case, the student made a faulty translation and Tarski’s World is not able to recognise this. This brings us to two errors not being recognised by the feedback mode of Tarski’s World. These errors are (d) errors in the translation of the Dutch sentences into logical expressions, and (e) errors during the problem solving process of finding an answer. The last two types of errors could only be traced by analysing the log files in which all the students’ actions were logged.

The Visualisation Variable

In order to study whether a simple world of geometrical objects facilitates logic learning, two instructional conditions were designed and constructed. In one condition, the students were presented a geometrical world in addition to Dutch sentences that had to be translated in first-order logic. In the other condition, the students were given a sentential world, that is, a textual description of a world, in addition to Dutch sentences that had to be translated in first-order logic. The authors assumed that the students who were given a geometrical world would outperform the students who were only presented a sentential world. The students in the first group would be able to visually check their reasoning in the available world and easily retain the steps made. When only presented a sentential world, imaginations arise automatically on the basis of the verbal descriptions. Johnson-Laird (1989) called these imaginations the mental models of discourse, making explicit the structure of the situation as it is imagined instead of the exact sentence. This results in a higher load on working memory and thus a greater chance of making errors and less concentration on the conceptions and operations of logic.

Learning results were measured by administering a transfer test. It was supposed that students understood the subject matter, if they were able to apply their newly acquired knowledge and skills to new situations in which the subject matter was not learnt. The transfer

test consisted of items that measured the students' ability to apply the rules of logic to everyday life problems. Psychologists have long been sceptical about the extent in which logical skills generalise to domain independent reasoning skills as people use in everyday life (see Nisbett, Fong, Lehman & Cheng, 1987 for an overview). It is widely held that for most people teaching logic only influences the algebraic symbol shuffling skills. Logic is seen as a syntactic mechanism of reasoning and because humans do not reason syntactically, teaching logic will neither help them to reason, nor to understand what their reasoning means. However, the authors assumed that if instruction is given in which the reasoning can be applied to a geometrical world, understanding is reached, so that this knowledge and these skills can also be used adequately in everyday life problems.

In addition to the transfer results, the authors' interest concerned the acquisition process together with the errors that were made within this process. Therefore, the authors decided to study the errors that the students made during the course of the problem solving process, although they did not have a specific prediction about the errors that would be made. It was assumed that presenting a simple world of geometrical objects would not influence the number or type of errors concerning syntax. However, it was hypothesised that the problem solving process of the students given a geometrical world was different from that of their colleagues in the other condition, because the former could use the geometrical world to check the steps in the reasoning process and to use the objects to retain the steps made, whereas the latter had to imagine the world themselves and had to cognitively operate on the imagined objects. This makes the problem solving process more difficult which will manifest in students needing more time to solve the problems, needing more checks of logical expressions in worlds and making more errors.

The Manipulation Variable

To solve logic problems, knowledge of abstract objects as well as the skill to perform logical operations has to be developed. Piaget (1970) stated that learners need to act in the environment if knowledge development is to ensue. Knowledge is constructed through actions on objects in the environment. He added to this that the development of knowledge of formal concepts is realised in a different way from the development of empirical knowledge. Piaget distinguished two kinds of experiences: (1) the physical experience that resembled learning in the experimental sciences and (2) the logic-mathematical experience that resembled learning in the formal sciences. The physical experience consisted of abstracting information from the object itself. For instance, a child picking up balls of different sizes experienced different weights and could infer certain general rules from this. The logic-mathematical experience, however, consisted of abstracting knowledge by operating on the objects and not from the objects themselves. In addition to characteristics already present, new characteristics were attributed to objects. Experience, then, referred to the relation between the characteristics attributed to the objects by manipulating them or operating on them, and not to the characteristics the objects already possessed. In this sense, knowledge was seen to be abstracted from the operations as such and not from the physical features of the object. For instance, a child learned the concept of order by ordering different balls to size. In this case, size was a feature all balls possessed, order was added by operating on the balls. The child understood that operating on the balls did not change the characteristics of the balls themselves.

At a certain moment, the applications of operations on physical objects become superfluous and the logic-mathematical operations are being integrated in symbolic operators, which can be applied in different contexts. Therefore, from a certain moment, pure logic and mathematics are left, for which no (concrete) experience is needed. Formal concepts and operations can be

abstracted from reality and these representations can be operated on mentally. This theory was extended to the acquisition of concepts and rules of logic.

Therefore, it was supposed that for solving logic problems the manipulation of objects would support the development of formal logic concepts and the use of logical operators. By adding and removing objects, by changing size or position, students could see what happened with the truth value of the logical expression they constructed.

To investigate whether operating on objects facilitated the development of formal concepts, two conditions were compared. In one condition, students were given a geometrical world in which they could manipulate concrete objects. In the other condition, students were given a geometrical world in which the objects could not be manipulated. Learning results were again measured by a transfer test. The authors assumed that students who were given the opportunity to manipulate objects would profit more from the environment and better understand the meaning of the logical expression than students who lacked this opportunity. Furthermore, differences in problem solving processes were expected, although the authors did not have specific expectations of the errors made.

Summary

In this study, the effect of two instructional variables was studied. The first variable concerned the extent of visualisation of the subject matter: a geometrically given world versus a sententially given world. The second variable concerned the extent in which the students could manipulate objects. As manipulating objects could only occur in a computer-based geometrical world, the two dimensions, visualisation and manipulation, partly overlapped. As a result, three conditions were administered: (a) the sentential, non-manipulation condition SN; (b) the geometrical, non-manipulation condition GN; and (c) the geometrical, manipulation condition

GM. It was supposed that the students in the third condition would profit most from the instructions.

## Method

### Participants

The participants were 66 first-year social science students (38 male, 28 female; mean age 19.4 years, SD 1.0). They volunteered for the experiment for which they were paid a fee of 50 Dutch guilders (approximately $ 25). None of the students had any experience in computer programming or logic.

### Learning Environment

The computer-based learning environment Tarski's World 4.1 for Windows (Barwise & Etchemendy, 1992) was used. Tarski's World provided an introduction in first-order logic. In the problems to be solved a well-defined, simple world of three kinds of geometrical objects (cubes, tetrahedrons and dodecahedrons) was used. Participants could change the size of the objects (small, middle or large) and the position of the objects (to the left of, to the right of, at the back of, in front of, and between). The learning environment consisted of four main components (see Figure 1): (a) the world module in which students could place the objects of a certain size and shape in the proper position; (b) the sentence module with the same possibilities but in formal notation; (c) the keyboard module for constructing sentences in the sentence module; and (d) the inspector module in which sentences from the sentence module could be checked to verify whether they were well-formed, correct and true/false in relation to the world in the world module.

The programme was adapted to fit the experimental design by making three versions that corresponded with the three conditions. In all three conditions, the students had to translate Dutch sentences into first-order predicate language. These sentences had to be checked in a world. In the sentential, non-manipulation condition SN, this world was given by a textual description. In the geometrical, non-manipulation condition GN, the world consisted of geometrical objects, which could not be manipulated. In the third, geometrical, manipulation condition GM, students had to construct and manipulate the world themselves.

The following changes in Tarski's World were made: (a) the menu bar was made invisible, so that students were not able to give commands themselves; (b) the programme was translated from English into Dutch, so language could not interfere with the results; (c) worlds and sentences were automatically loaded and saved when starting and finishing a task; (d) in the geometrical, non-manipulation condition a certain world was given which could not be manipulated by the students; and (e) in the sentential condition the geometrical world was made invisible.

The instruction accompanying Tarski's World was provided in the browser of Netscape Communicator 4.06. The changes in the browser were as follows: (a) the menu options were disabled, so that students could not navigate completely freely in the browser nor surf on the internet; (b) the browser was linked to Tarski's World, so commands in one programme resulted in actions in the other programme.

## Learning Materials

The learning material comprised the conditional. Two sentences p and q can be combined into a new sentence with the symbol of the conditional. The new sentence will look like $p \rightarrow q$; its English counterpart is “If p, then q”.

Tests and Questionnaires

To measure the students’ knowledge of the meaning of the conditional, a transfer test of eleven items was administered. Figure 2 shows a typical example of an item of the Wason Selection Task (1966), as used in the experiment. In the transfer test, one abstract Wason (card)task, two concrete, non-arbitrary Wason tasks and one near-transfer Wason task in a Tarski’s World setting was used.

---

Insert Figure 2 about here

---

The remaining seven items included one Reduced Array Selection Task (RAST, Johnson-Laird & Wason, 1970), two items to be solved best by using set theory and four items in which a statement was given and the students had to decide whether this statement was true or false or whether you could not tell from the information given. Figure 3 shows a typical example of the latter.

---

Insert Figure 3 about here

---

The students had to complete three comparable versions of the transfer test, namely a pre-, post- and retention test. These tests were designed to measure the knowledge gained after the various instructions. All students received the same tests.

Two questionnaires were administered. The first concerned the former education of the students on mathematics and logic. The second questionnaire was an evaluation of the instruction in combination with Tarski's World.

All tests and questionnaires were administered on the computer.

## Log Files

Two log files were generated during the experiment. The first log file logged all the actions of the students while working in Tarski's World. It logged the status of the sentences and the matching world every time the students checked this combination on Well Formed Formula (WFF), sentence and/or truth. The second log file logged all the actions of the students while working in the browser. This, among others, concerned answers of students on two questionnaires, answers of students on the transfer tests, and time registration.

From these log files, different results could be taken. First, the mean time students were working on the problems during general training and experimental problems could be calculated. Second, the mean number of checks per student could be computed. As the subject matter concerned the conditional with the general format $p \rightarrow q$, these checks could be divided into the four possible truth-falsity combinations ($1 \rightarrow 1$, $1 \rightarrow 0$, $0 \rightarrow 1$, $0 \rightarrow 0$, in which 1 = true, and 0 = false). Third, the mean number of errors in the final answer per student could be determined. These 'final' errors could be divided into two kinds: (a) 'indicated' errors, that is, errors indicated by the feedback module of Tarski's World, but deliberately ignored by the students;

and (b) ‘non-indicated’ errors, that is errors not indicated by the feedback module of Tarski’s World, so that the students did not detect them. When the latter type of errors occurred, it was mostly because Tarski’s World did not check whether the logical expressions were correctly translated from the Dutch sentences.

## Design and Procedure

Experimental Conditions. The subjects were randomly assigned to one of three conditions. In all conditions, the students were given Dutch sentences, which they had to translate in first order predicate language. In the sentential, non-manipulation condition SN, the students had to check the sentences to be true or false in a given, sentential world. In the geometrical, non-manipulation condition GN, the students had to check the sentences to be true or false in a given, static geometrical world. In the geometrical, manipulation condition GM, the students had to make geometrical worlds in which they had to check the sentences to be true or false. In all conditions, students were allowed to use scrap paper if they wished.

Procedure. The experiment was held in three consecutive sessions; the first (pre test, instruction and general training) and the second (exercises and post test) on two successive days and the third (retention test) three weeks later.

The first session started with an introduction after which the students had to complete a questionnaire about their previous education in mathematics and logic. This questionnaire was followed by a pre test, which consisted of eleven puzzles measuring the knowledge of the subjects of several aspects of the conditional. Successively, the subjects received a verbal instruction in which they got an introductory course into first-order logic, as used in Tarski’s World. This instruction gave the subjects an idea of what logic can be used for, what Tarski’s World can do, what logic operators and quantifiers are available, how these operators and

quantifiers can be used and what truth and falsity meant. This, together with some examples was the knowledge the students were equipped with. After the instruction, the students received a general training of about two hours depending on the condition to which they were assigned. During the training the students learnt to work with Tarski's World and with the logic operators. For this, model progression was used, an idea introduced by White and Frederiksen (1990). One of the general principles of model progression is to structure the rich information source and to keep the environment manageable by not introducing too many ideas at one time. Model progression entails starting with a simplified version of a model and gradually offering more complex versions of the model. In this case, the model was the field of predicate logic. The concepts were introduced in the following order: (a) predicates and constants, (b) connectives and parentheses, (c) quantifiers and variables, and (d) conditional. If the students had any questions, assistance was given by one of the experimenters present.

The second session started with six problems that had to be solved by the students in the three conditions. The first two problems were presented to refresh the knowledge acquired the day before. Consequently, four exercises addressed the conditional. In all the exercises students were asked to translate Dutch sentences into first-order logic and to check the truth of the sentences in the geometrical or sentential world. After the students had completed the exercises, a post test was administered. In this post test, the students were again tested on their knowledge of the conditional. The post test consisted of the same type of items as used in the pre test. Also the second questionnaire in which the instruction and Tarski's World was evaluated, was administered. Three weeks after the experiment, the students had to return for the retention test. This test consisted of comparable items as were used in the pre- and post test.

## Results

Reliability

The reliability of pre-, post- and retention test, as measured with Cronbach's $\alpha$, was $\alpha = .49$; $\alpha = .68$; and $\alpha = .75$ respectively. Deleting items from the test did not lead to significant higher reliabilities.

Pre-, Post- and Retention Tests

Table 2 shows the means and standard deviations on the tests for the three conditions GM, GN, SN. The maximum score was 11. Scores on the pre-, post- and retention tests increased significantly for condition GM ($F(2, 42) = 3.85$, $p < .05$), condition GN ($F(2, 42) = 7.91$, $p < .01$), and condition SN ($F(2, 42) = 11.39$, $p < .001$).

---

Insert Table 2 about here

---

To study the effect of the visualisation variable, the SN- and the GN-condition were compared. To study the effect of the manipulation variable, the GN- and GM-condition were compared. In both cases there was no difference between the conditions on the pre test ($F(1, 42) = .27$, $p > .05$ and $F(1, 42) = .03$, $p > .05$ respectively). The use of visualisation did not yield significant differences between the conditions SN and GN on the post test ($F(1, 42) = .02$, $p > .05$) and on the retention test ($F(1, 42) = .00$, $p > .05$). Neither did the manipulation of the objects in the world show a significant difference between the conditions GN and GM on the post test ($F(1, 42) = .40$, $p > .05$) and on the retention test ($F(1, 42) = 1.29$, $p > .05$).

Process Data

The students' actions, which were stored in the log files, are summarised in Table 3. As can be seen, condition GM distinguished from condition GN and SN on several aspects: (a) students in condition GM needed more time to complete the experimental problems than students in condition GN ($F(1, 42) = 14.31$, $p < .001$) and students in condition SN ($F(1, 42) = 5.42$, $p < .05$); (b) students in condition GM used more checks on all sentences than students in condition GN ($F(1, 42) = 10.72$, $p < .05$) and students in condition SN ($F(1, 42) = 13.24$, $p < .001$), especially on sentences in which both antecedent and consequent were true ($1 \rightarrow 1$); and (c) students in condition GM deliberately ignored less 'indicated' errors, although this was not significant compared to students in the condition GN ($F(1, 42) = 3.08$, $p > .05$) and compared to students in the condition SN ($F(1, 42) = 3.14$, $p > .05$), and they made far more 'not-indicated' errors, that is errors of which Tarski's World did not indicate they were made, compared to students in the GN-condition ($F(1, 42) = 7.78$, $p < .01$) and compared to students in the GM-condition ($F(1, 42) = 7.68$, $p < .01$).

---

Insert Table 3 about here

---

All students were allowed to use scrap paper. It turned out that 59% of the students in the SN-condition made use of this possibility. In all cases, the paper was used to draw the given, sentential world. Students in the other two conditions did not use the scrap paper.

Questionnaire

At the end of the experiment, 82 % of the students stated that they enjoyed working with the programme. Furthermore, 72% of the students in the GM- and GN-condition stated that the concrete, visual representation made it easy to work with the logical formulae.

## Discussion

### Reliability

The fact that the reliability of the pre test was lower than the reliabilities of the post- and retention test is explained by the small number of correct answers on the pre test ($M = 4.24$, $SD = 1.73$). Apparently, the pre test was difficult, so that the students may have been guessing when answering the items, which has a negative influence on the reliability of the test.

### Learning Results

The scores on the pre-, post- and retention tests clearly show that all students profited from the instructional conditions. The students were able to solve significantly more logic problems correctly on the post- and retention tests in comparison to the pre test. Because the items of the post- and retention test also comprised the Wason selection task, the knowledge and skills acquired in Tarski's World were transferred to very different problem situations. This is evidence that far transfer is possible. The results show that, although scores on the retention test are still rather low, even non-technical students are able to do better on the difficult items of the transfer tasks after instruction. The authors suppose the findings are fostered by the advantages of Tarski's World. In this learning environment, it is easy to construct logical expressions, as the programme automatically shows parentheses, commas and the number of arguments that go with a predicate. This allows the student to quickly concentrate on the conceptions and operations of

logic. Also, the sentences are easily checked on syntactical correctness by using the inspector module. The worlds allow the student to visualise the objects and their relationships to one another and to test the truth of logical expressions in a given world. Learning occurs by successively testing logical formulae in worlds and by the immediate feedback that is given to the student.

The information on the pre-, post- and retention tests suggests, that in the interval between administering the post- and retention test, the students continued reflecting on the logic problems, which had a positive effect on the learning results. The scores on the retention tests were significantly higher than scores on the post tests. The students needed time for integrating their newly developed knowledge with their existing knowledge.

Though for all conditions the scores on the post- and retention tests increased, no differences were found between the three instructional conditions. In the experiment, the instructional variables, visualisation and manipulation, did not influence the test results. The authors assume that this result is probably due to the time spent on the instructional conditions. The students had to solve only six introductory problems. The number of steps in solving the problems was small and the students may have imagined a world and remembered the cognitive steps they made without the need for help from visualisation and/or manipulation. Also the use of scrap paper may have provided support that interfered with the visualisation variable. In a next experiment, the complexity of the problems will be increased to study the relevance of the variables for instruction.

## Cognitive Processes

The process data of the students in the geometrical condition GN did not differ much from the process data of the students in the sentential condition SN. This might be due to the design of

the SN-condition. By introducing scrap paper, the students in this condition were still able to use visualisations, if they needed it. In this condition, 59 % of the students used this opportunity, especially when the problems became more complex. This shows that most students need a geometrical representation in which their problem solving process can be made concrete and in which steps can be retained instead of keeping these into working memory. The other 41% of the participants did not use the scrap paper to solve the problems. They were apparently able to use mental objects for this situation instead of perceivable objects.

The process data of the students in the manipulation condition GM clearly differed in number of checks and amount of time used from the data of the students in the non-manipulation condition GN. Because of these differences, the working method of the students in condition GM was given a closer look. Three findings will be discussed here.

First, students in this condition had more freedom to explore the different combinations of sentences in worlds. The students were expected to manipulate the objects in the world and to infer the behaviour of the conditional by induction. However, this was not what happened. In the exercises, students first translated the Dutch sentences into first-order logic. Then, they constructed a world that matched the first sentence. However, if it was the case that the world had to be changed to have the second sentence match the world, the students only added objects to the world. They never removed objects or started all over, not even when the problem could then not be solved correctly.

Second, it appeared that students played around in such a way that they were confronted with the subject matter they already understood, but that they did not confront themselves with the subject matter put in a new situation. They were tempted to stay in familiar situations, even when given freedom in exploring. This can be concluded from the high amount of checks in

situations in which both the antecedent and consequent were true ($1 \rightarrow 1$). They kept on the safe side, repeatedly checking things they already knew, instead of trying out something new. They headed straight for the solution without straying from their path, even if this could have resulted in a better solution. Apparently, more guidance is needed to lead them to less familiar situations.

The last finding was that students in the GM-condition made a world on the basis of information from the sentences. If a sentence was about something, the students put these objects in the world. This is a way of working human beings use in everyday life. They work affirmatively, they do not start sentences using negations, and they do normally not reason about things that are not present. For instance, the sentence “all cubes are large” is a complete nonsense sentence in everyday life when no cubes are present. In logic, however, this sentence is true. Differences in language between everyday life and logic are explained by the theory of Grice (for an overview, see Gamut, 1982).

In addition to differences in working method, the students in the GM-condition had an extra difficulty in their instruction. As Tarski’s World was not able to recognise Dutch sentences in natural language, the programme only checked whether the sentence in first-order language was correct. As a consequence, the programme did not check whether the logical sentence was the correct translation of the Dutch sentence. Therefore, it could happen that the students thought they correctly translated the sentences, whereas this was not true. If the students then checked the Dutch sentences in the world instead of the logical sentences, it was possible that they deduced the wrong principles. So, students in the GM-condition can have learnt wrong conceptions. The data in Table 3 support this assertion. Students in the GM-condition were more often not aware of making wrong worlds or sentences compared to students in the other two conditions.

Affective Reception

The students in the GM- and GN-condition were positive about the use of a world. Of these students, 72% stated that the concrete, visual representation made it easy to work with the logical formulae. Of all students, 82% stated that they enjoyed working with the programme.

# References

Barwise, J., & Etchemendy, J. (1992). The language of first-order logic: Including the Microsoft Windows program Tarski's World 4.0 for use with IBM-compatible computers. Stanford, CA: CSLI.

Barwise, J., & Etchemendy. J. (1998). Computers, visualization, and the nature of reasoning. In T. W. Bynum & J. H. Moor (Eds.), The digital phoenix: How computers are changing philosophy (pp. 93-116). London: Blackwell.

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. Cognitive Psychology, 18, 293-328.

Freudenthal, H. (1991). Revisiting mathematics education: China lectures. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Fung, P., O'Shea, T., Goldson, D., Reeves, S., & Bornat, R. (1994). Why computer science students find formal reasoning frightening. Journal of Computer Assisted Learning, 10, 240-250.

Gamut, L. T. F. (1982). Logica, taal en betekenis 1: Inleiding in de logica. [Logic, language and meaning 1: Introduction into logic]. Utrecht, The Netherlands: Het Spectrum.

Goldson, D., Reeves, S., & Bornat, R. (1993). A review of several programs for the teaching of logic. The Computer Journal, 36(4), 373-386.

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), Syntax and semantics: Vol. 3. Speech acts (pp. 41-58). New York: Academic Press.

Johnson-Laird, P. N. (1989). Mental Models. In M. I. Posner (Ed.), Foundations of cognitive science (pp. 469-499). Cambridge, MA: The MIT Press.

Johnson-Laird, P. N., & Wason, P. C. (1970). A theoretical analysis of insight into a reasoning task. Cognitive Psychology, 1, 134-148.

Lehman, D. R., & Nisbett, R. E. (1990). A longitudinal study of the effects of undergraduate training on reasoning. Developmental Psychology, 26(6), 952-960.

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P .W. (1987). Teaching reasoning. Science, 238, 625-631.

Piaget, J. (1970). Epistémologie génétique. [Genetic epistemology]. Paris: Presses Universitaires de France.

Pintrich, P. R. (1990). Implications of psychological research on student learning and college teaching for teacher education. In W. R. Houston, M. Haberman & J. Sikula (Eds.), Handbook of research on teacher education (pp. 826-857). New York: MacMillan.

Stenning, K., Cox, R., & Oberlander, J. (1995). Contrasting the cognitive effects of graphical and sentential logic teaching: Reasoning, representation and individual differences. Language and Cognitive Processes, 10(3/4), 333-354.

Van der Pal, J., & Eysink, T. (1999). Balancing situativity and formality: The importance of relating a formal language to interactive graphics in logic instruction. Learning and Instruction, 9, 327-341.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), New horizons in psychology: Vol. I. Harmondsworth: Penguin.

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

White, B. Y., & Frederiksen, J. R. (1990). Causal model progressions as a foundation for intelligent learning environments. Artificial Intelligence, 42, 99-175.

# Figure Captions

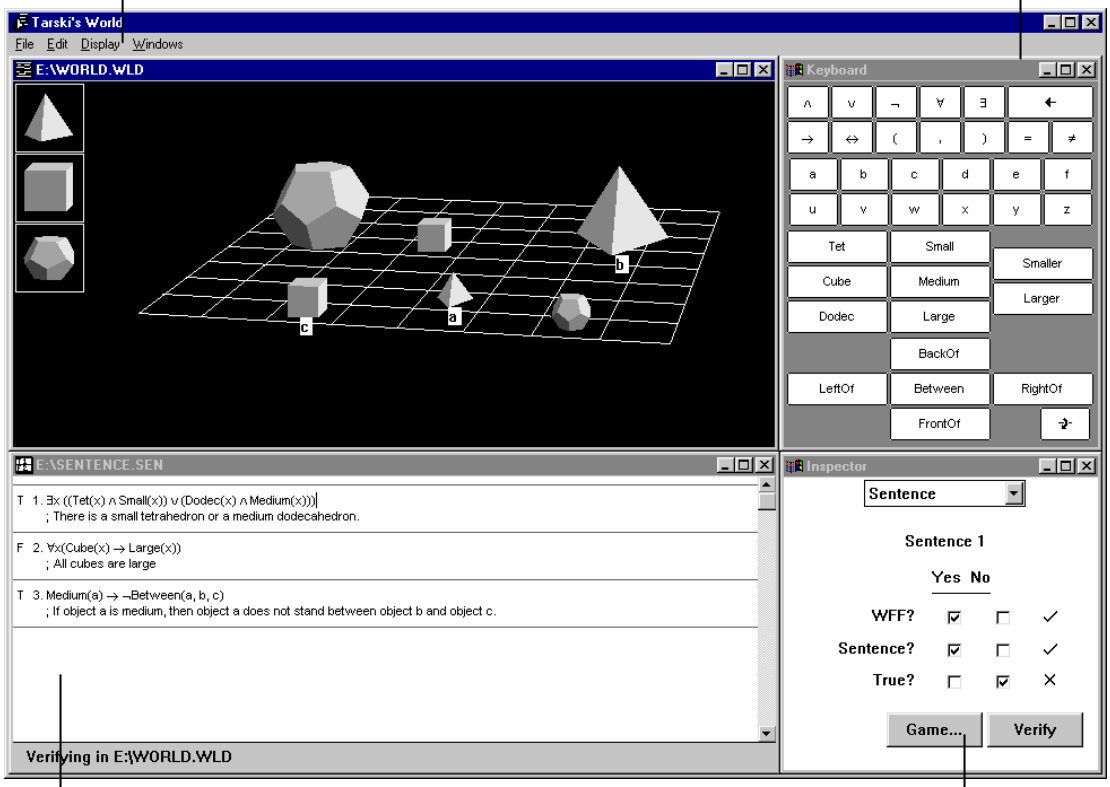
Figure 1. Tarski's World: world module, sentence module, inspector module and keyboard module.

Figure 2. A typical example of the Wason Selection Task (Wason, 1966).

Figure 3. A typical example of a problem in the transfer tests.

**World module:** the student can construct a world of geometric objects by mouse clicking on the objects on the left of the screen. The objects can vary in shape (Tet, Cube, Dodec), size (Small, Medium, Large) and position (LeftOf, RightOf, BackOf, FrontOf, Between).

**Keyboard module:** the student can construct logical formulae by mouse clicking on the keys.



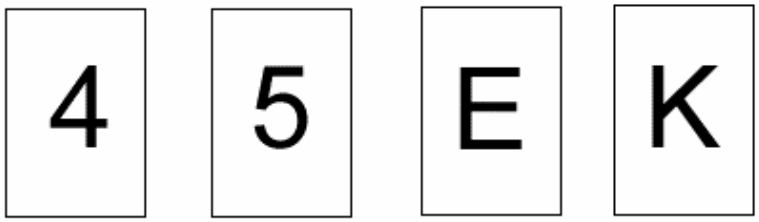

**Sentence module:** the sentences the student is constructing by using the keyboard module appear in this screen. The computer ignores the semicolons followed by sentences in English. By giving T(rue)s and F(alse)s, the computer gives feedback about the truth value of the sentences in the world given in the world module.

**Inspector module:** the student can check whether a logical sentence is syntactically correct (WFF?), whether all variables are bound to a quantifier (Sentence?) and whether the formula is true (True?) in the world given in the world module, by selecting a box. The computer gives immediate feedback (✓, ✖).

Below are four cards. On each card there is always a letter on one side and a number on the other side. A card never contains two numbers or two letters. There is a rule that says:

**If there is an E on one side, then there is a 4 on the other side.**

Which cards do you have to turn over in order to decide whether the rule is true or false?



The following statement is given:

**If I go to the city today, I will eat an ice cream.**

I am going to the beach today and I am eating an ice-cream.

Is the above given statement true or not?

- ◯ yes
- ◯ no
- ◯ that depends

Table 1

Truth tables of the four connectives negation (¬), conjunction (∧), disjunction (∨), and conditional (→)

Elementary Propositions

| p | ¬p |
|---|---|
| 0 | 1 |
| 1 | 0 |

Complex Propositions

| p | q | p ∧ q | p ∨ q | p → q |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 1 |
| 0 | 1 | 0 | 1 | 1 |
| 1 | 0 | 0 | 1 | 0 |
| 1 | 1 | 1 | 1 | 1 |

Note. p and q denote arbitrary propositions. 0 = false; 1 = true.

Table 2

Means and Standard Deviations for Each Condition on Pre-, Post- and Retention Tests

| | Condition | | |
|---|---|---|---|
| Test | SN | GN | GM |
| Pre test | | | |
| M | 4.45 | 4.18 | 4.09 |
| SD | 1.97 | 1.50 | 1.74 |
| Post test | | | |
| M | 4.91 | 5.00 | 4.59 |
| SD | 2.31 | 2.27 | 1.99 |
| Retention test | | | |
| M | 6.05 | 6.00 | 5.14 |
| SD | 2.48 | 2.60 | 2.44 |

Note. Maximum score = 11. SN = sentential, non-manipulation condition; GN = geometrical, non-manipulation condition; GM = geometrical, manipulation condition.

Table 3

Summary of Process Data of Students Working in Tarski's World

| | Condition | | |
|---|---|---|---|
| | SN | GN | GM |
| time[a] | | | |
| General training | 1:41:20 | 1:28:51 | 1:41:46 |
| Exp. Problems | 0:24:39 | 0:21:32 | 0:30:27 |
| Total | 2:05:59 | 1:50:23 | 2:12:13 |
| # checks | | | |
| all sentences[b] | 31.5 | 29.7 | 47.4 |
| $1 \rightarrow 1$[c] | 4.3 | 3.9 | 12.6 |
| $1 \rightarrow 0$[c] | 3.7 | 3.5 | 4.2 |
| $0 \rightarrow 1$[c] | 1.8 | 2.1 | 1.3 |
| $0 \rightarrow 0$[c] | 3.8 | 4.0 | 2.1 |
| # final errors[d] | | | |
| Indicated[e] | 29 | 27 | 15 |
| Non-indicated[f] | 44 | 44 | 80 |

Note. SN = sentential, non-manipulation condition; GN = geometrical, non-manipulation condition; GM = geometrical, manipulation condition.

(table continues)

[a] mean time students were working during resp. the general training, the experimental exercises, and the sum of these two.

[b] mean number of checks per student over all sentences during the experimental exercises.

[c] mean number of checks per student made on the four possible checks of the conditional ($\underline{p} \rightarrow \underline{q}$) during the experimental exercises (1= true; 0 = false).

[d] mean number of final errors students made during the experimental exercises.

[e] errors indicated by the feedback module of Tarski's World.

[f] errors not indicated by the feedback module of Tarski's World.