Beaman, C. P. & Morton, J. (1998) *Proceedings of the Second European Conference on Cognitive Modelling, Nottingham, UK.*

Modelling Memory-Updating Characteristics of 3- and 4-Year Olds

C. Philip Beaman
John Morton
MRC Cognitive Development Unit
4 Taviton Street
London WC1H 0BT, UK
+44 171 387 4692
P.Beaman@cdu.ucl.ac.uk
J.Morton@cdu.ucl.ac.uk

ABSTRACT

In this paper a memory perspective on young children's performance at a particular false belief task, the Smarties task, is described. The theoretical analysis focuses on the computational conditions that are required to resolve the Smarties task, on the possible limitation in the developing memory system that may lead to a computational break-down resulting in a failure to resolve, and on ways of bypassing such limitations to ensure correct resolution. A symbolic model of this analysis implemented using the COGENT modelling environment is described, and its fit to the data considered.

Keywords

Developmental modelling, false belief, memory updating, COGENT

INTRODUCTION

One of the many constraints identified by Newell (1990) on any form of cognitive architecture which attempts to model human cognition is that it should be capable of arising from earlier forms by a process of developmental maturation. Developmental constraints, and discrete developmental stages, have received surprisingly little attention from symbolic modellers, although questions of how a mature system might develop from a relatively simple template are now being considered within the connectionist research program (e.g., Elman et al., 1996). The present study considers a developmental stage believed to be crucial to the maturation of memory processes, and aims to demonstrate how the failure of 3- and 4-year olds at a task which adults find trivially easy (the Smarties task; Perner, Leekam & Wimmer, 1987) can be modelled using a destructive-updating process. A subtle alteration of the memory encoding characteristics of this task enables 3- and 4-year olds to perform the task correctly. The patterns of children's performances are modelled as discrete developmental stages using the COGENT (Cognitive Objects in a Graphical EnvironmeNT) modelling environment of Cooper and Fox (in press).

The Smarties Task

The basic procedure for the Smarties task is as follows. The subjects are shown a tube of Smarties (a popular brand of sweet) and asked what the tube contains. Children of around the age of four are usually both able and willing to provide an answer to this question. The top is then taken off the tube, and its contents are shown to the child. The contents of the tube are pencils rather than the anticipated Smarties. The top is then replaced on the tube, and the child is asked two questions, the reality question (what is in the tube?) and the belief question (when you first saw the tube, what did you think was in the tube?). Typically, 70% of 3-year-old children who are able to answer the first question correctly (pencils) now also give the same answer to the second question.

A Memory-Updating Explanation

The original form of the Smarties task implies some peculiar memory characteristics. Children who fail this task are incorrectly reporting a belief which they had held, and told to the experimenter, only seconds previously. Although a conceptual deficit, an inability to comprehend false belief, can be put forward to explain these results, it seems strange to suppose that this deficit manifests itself in the child's inability to correctly recall the contents of this belief, even though they were able to report to the experimenter what the contents of this belief were immediately before it was shown to be false. Instead, it is argued (Barreau, 1997; Morton, 1997) that the child's inability is centred around a memory updating system, such that the false belief (that the tube contains Smarties) is never encoded as a stable, long-term representation, and so is immediately supplanted by the incoming information that the tube contains pencils. Thus, when such children are asked the belief question, the only source of information available to them is the representation of the current state of reality: ***in(tube, pencils)***.

The Bag Experiment.

A variation on this experimental procedure designed to maximise the possibility that the contents of the tube are translated into a long-term format is described by Barreau (1997). Immediately after showing the tube to the child, and asking the child what they believed the tube to contain, the contents of the tube were emptied into a bag. Although the child witnessed this operation, at no time were they able to see the contents of the tube either at first or during the transfer. The tube was then shown to the child to demonstrate that it was empty, and then ostentatiously hidden from view. The child is then asked what they believe to be in the bag. All children replied "Smarties". The contents of the bag were then shown to the child. In this case, the bag contained marbles, rather than Smarties. The child was then asked five questions concerning the contents of the bag and the tube:

1. Before I opened the bag, what did you think was in the bag? (BAG:BELIEF: PAST)
2. What is really in the bag? (BAG:REALITY: PRESENT)
3. When I first showed you the tube, what did you think was in the tube? (TUBE: BELIEF: PAST)
4. What is inside the tube now? (TUBE: REALITY: PRESENT)
5. What was really inside the tube? (TUBE: REALITY: PAST)

In Barreau's (1997) experiment, twenty-four children were questioned in this manner, the results of this experiment are shown in the table below:

TABLE 1: Table of answers to the tube and bag questions.

Questions	Correct	Reversed	Double
BAG	8	8	8
TUBE	15	3	6

In order to be scored correct, both the bag questions, (belief and reality) had to be correctly answered. To be scored correct in the tube condition, the belief questions and at least one of the reality questions had to be correctly answered. A "double" score refers to a repeat answer, i.e. a reality response to a belief question. This category also includes one child who gave belief answers to reality questions. The reversed response indicates a reversal between the belief and reality answers in the bag questions, and the belief and one of the reality answers in the tube questions.

The assumptions underlying this experiment were that when the tube was removed from view, the tube→ bag transferral episode would be coded as ended, and details of the whole episode would be translated into long-term memory. Thus, when the current representation of the bag's contents is updated, the representation of the tube's contents will be invulnerable.

The data has also been analysed as suggesting that three qualitatively different developmental processes are occurring amongst the children tested (Barreau, 1997).The children were divided into three groups on the basis of the scores they were given for the bag questions. Of the 8 children who were scored as correct for the bag questions, 7 were also correct for the tube question, and 1 gave a "double" response. Of the 8 children who gave reversed responses for the bag question, 6 were scored as correct on the tube question, there was 1 reversed response, and 1 double response, and for the 8 children who scored "double" responses for the bag questions, 2 were correct on the tube questions, 2 gave reversed responses, and 4 gave double responses. This pattern of data was considered to be a little too complex to be easily handled by a traditional verbal theory.

A COGENT IMPLEMENTATION

To properly test the theory against the data, a family of models were produced using the COGENT modelling environment. The basic architecture used in this approach is reproduced below: In this figure, hexagons represent processes, rounded rectangles represent buffers, and diamonds represent data boxes. Square boxes represent compounds, which may contain buffers and processes. Arrows with standard heads indicate message sending. Arrows with black triangular tails indicate buffer reading. Compound arrows (which are denoted by triangular and standard heads) allow both functions.

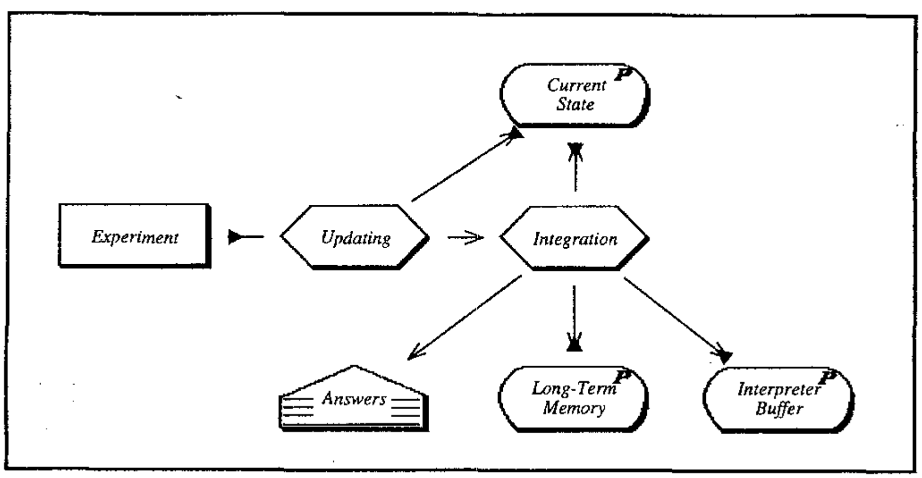

figure 1 - the COGENT object-level representation of the simulation

For the purposes of this paper, the "Experiment" compound is used only as a means of feeding information to the system simulating the child's mental processes, and will not be discussed in any great detail. Note that for the bag experiment, the simulation must include the correct answering of three "belief establishing" questions prior to the five questions of main interest within the experiment. The belief establishing questions were included within the experiment to ensure that the child had formed the correct representations of the state of the world prior to being tested on their memory for the sequence of events. These questions include the initial question of the Smarties task (What do you think is in the tube?), a repeat of the question to ascertain that the child believes the tube is empty (What is in the tube now?) once the transfer operation has taken place, and a question to ensure that the child has tracked the transferral of the supposed Smarties (What do you think is in the bag?). In the experiment, after asking one of the belief establishing questions, the experimenter waited until the child had answered before continuing with the procedure. Accordingly, in the simulation, no further input was fed to the system until the cycle after the system had output the answer to the previous question. This protocol was observed throughout all the simulations.

The Smarties Simulation.

We assume that the 30% of 3- and 4-year olds who pass the Smarties test do so by accessing a long-term memory (LTM) representation of the likely contents of a Smarties tube, so we do not attempt to deal with this question in any detail here. This is consistent with the developmental literature, which has focused only upon those children who fail. The initial simulation then, must be one that gives a "reality" answer to a "belief" question under the circumstances of the Smarties experiment. The experimental procedure is modelled by adding propositions about the current state of the environment a cycle at a time to an "environment" buffer, within the Experiment compound, which is read by the updating process. The Current State Buffer is a representation of current environmental contingencies. This is kept up-to-date by **destructive updating** which occurs by the operation of the following rules:

RULE 1.
IF: A is in Experiment: Environment
not A is in Current State
THEN: add A to Current State

RULE 2.
IF: in(X,Y) is in Experiment: Environment
in(X,Z) is in Current State
THEN: delete in(X,Z) from Current State

Thus, if *in(tube,smarties)* is in the Current State and *in(tube,pencils)* appears in the Environment, *in(tube,smarties)* is deleted from the Current State by the second of the above rules and is replaced by *in(tube,pencils)*.
The basic workings of the model of the Smarties task are as follows:

In LTM there is a generic representation of past experience of Smarties tubes,

g(in(tube,smarties)),

and a further rule in the integration process that states the contents can be matched to their containers on the basis of such past experience:

RULE 3.
IF: g(in(X,Y)) is in Long-Term Memory
object(X) is in Current State
not in(X,Z) is in Current State
THEN: add in(X,Y) to Current State

This rule is refracted, so that it only fires the first time its conditions are satisfied within a COGENT run. When a tube representation is added to the Current State Buffer, this rule fires and the inference is made that the tube contains Smarties. This information is overwritten, however, when the further information is added from the environment that the tube contains pencils. Thus, when the question regarding the contents of the tube is presented to the system

question(present(in(tube, What))),

the present representation of the current contents of the tube in the Current State Buffer instantiates the unknown variable in the question, and provides the only possible answer: *in(tube, pencils)*.

Questions are dealt with by being passed immediately over from the Current State Buffer to the Interpreter Buffer. Once a question is received in the Interpreter Buffer, it activates the relevant search processes according to the following rules:

RULE 4.
IF: question(present(X)) is in Interpreter Buffer
X is in Current State
THEN: clear Interpreter Buffer
add answer(X) to Interpreter Buffer

RULE 5.
IF: question(past(X)) is in Interpreter Buffer
record(Y) is in Long-Term Memory
X is a member of Y
not X is in Current State
THEN: clear Interpreter Buffer
add record (Y) to Interpreter Buffer
add answer (X) to Interpreter Buffer

Thus, the unknown variables within the question are instantiated either in the Current State Buffer or in LTM, and translated into an answer format. All answers within the Interpreter Buffer are immediately sent to the output processes represented in the diagram by the triangular "Answers" block.

The Bag Simulation.

In the case of the bag experiment, the simulation is a little more complex. In particular, we have to tackle the creation of event records. To do this, a rule must fire when an event is perceived to end. This rule translates all information currently being processed (the contents of the Interpreter Buffer), together with the current representation of the environment (the contents of the Current State Buffer) into an LTM format. In the hypothesis underlying the experimental procedure, the event was signalled to be at an end by a contextual change, the removal of the tube. In the simulation, a record is closed if there are more objects represented in the Current State Buffer than are present in the environment. This is captured formally by the updating rule:

RULE 6.
IF: Objects is the list of all object(X) such that
object(X) is in Experiment: Environment
Representations is the list of all object(X)
such that object(X) is in Current State
A is the length of Objects
B is the length of Representations
B > A
THEN: send close_record to Integration

Upon receiving the close_record trigger, a further rule fires within the integration process which transforms the information within the Current State Buffer and the Interpreter Buffer into a list structure in LTM. The Interpreter Buffer is then cleared.

Simulation Results.
The basic simulation can easily handle the results of the first group of children, those who were scored correct on the bag question (group A). When asked the bag questions, the simulation of this group of children has a record available containing the previous belief concerning the bag's contents,

in(tube,smarties)

which it can use to answer the first question (BAG: BELIEF: PAST), in accordance with rule 5. When asked the second bag question (BAG: REALITY: PRESENT), a Current State representation of the bag's current contents is employed to answer this question in accordance with rule 4.

Seven out of eight of this group of children were also scored as correct for the tube question. In the model, the tube question is handled by the existence of a record available in LTM which can be retrieved to answer the question. The creation of this record was triggered by the removal of the tube. Note that the record does *not* contain a verbatim representation that the tube contained marbles. Instead, the record contains the representation that the contents of the tube were emptied into the bag:

action(empty(tube,bag)),

that the tube is now empty:

in(tube,[]) (where [] denotes the empty set),

and that the bag contained marbles. To correctly answer questions regarding the initial contents of the tube (questions 3 and 5, TUBE: BELIEF: PAST and TUBE: REALITY: PAST) a further rule is necessary to allow the inference that the tube's contents can be ascertained by backwards reasoning from the bag's contents, and the fact that the contents of the tube were entered into the bag. Formally, this rule is:

RULE 7.
IF record(Y) is in Interpreter Buffer
question(past(in(A,B))) is in Interpreter Buffer
action(empty(A,C)) is a member of record(Y)
in(C,D) is a member of record(Y)
THEN: clear Interpreter Buffer
add answer(in(A,D)) to Interpreter Buffer

This rule is triggered if the current representation of the tube's contents is identical to the retrieved LTM representation. Since the child is presumably not expecting to answer a "present" question at this point, the rule allows the search, via inference, for an alternative "past" answer. Note that the simulation demonstrates that Morton's (1997, p. 938) comment that "the conditions are the same" for the tube questions of the bag experiment and for the same questions in the Smarties experiment is not strictly necessary when analysed in terms of the underlying theory. In this simulation, when the inference rule regarding the transferral operation is manually prevented from firing the default answer from the system to the tube questions is that the tube was empty. Since the child was shown the empty tube during the bag episode this forms part of the same record. The full contents of this record are displayed below:

record([[in(bag,smarties), in(tube,[]), object(bag),
action(empty(tube,bag)), object(tube)]
action(remove(tube))]).

With this set of rules, the simulation therefore produces the same answers in the bag experiment as seven out of eight of the children in group A.

The initial results of those children who were scored as giving "reversed" answers (group B) to the bag question need to be explained differently. Recall that these children gave reality answers to belief questions and vice versa. The simulation of this situation uses the same basic structure as the simulation of group A (the "corrects"). However, it is assumed that the group B children attempt to answer all questions initially from their current state representation of the world. Arguably this is less effortful than retrieving information from LTM (see Morton, Hammersley & Bekerian, 1985 for a discussion of the complexities of retrieval from LTM). In effect, we assume that the tagging of questions as referring to past and present is not as well established in this group as in group A. The group B children, then, are not forced to search LTM in response to a PAST question. Rather, they only look in LTM when the Current State search has failed. Since the Current State Buffer representation is one of reality rather than belief, these children's default strategy results in a reversal of belief and reality answers.

Briefly, the simulation of this state of affairs works as follows. The "past" and "present" modifiers in the input are ignored in the integration process by rules 4 and 5, and, instead, all questions are followed by an initial search in CS. This leads to the initial mistake. The reversal of the situation with the next question is simply implemented by making that the look-up rule for information in Current State into a refracted rule so that it cannot be used as a default when the next question is asked. This is the "present" reality question, and the only way the child can answer the question is by searching for a long-term memory representation with information about the contents of the bag. This is found in the record which specifies

in(bag,smarties)

resulting in a reversed pattern of results.

This simulation works well when only the bag question is considered, but runs into problems when the tube questions are also added to the simulation's input, since it produces a further "reversed" pattern of results for these questions. In fact only one child in this group was scored as giving "reversed" responses to the tube question, and six were scored as correct. This failing will be considered in more detail later.

The final group of children to be considered (group C) gave the "reality" answers to "belief" questions. Working on the logic employed in the simulation of group B's results it is assumed that these children also ignore the past/present modifiers and attempt to answer the question in the simplest way possible, by retrieving an answer from the Current State Buffer representation. However, for these children the assumption is that the search rule for the Current State Buffer is not refracted. Accordingly, the simulation produces repeated answers from the Current State Buffer, which are identical to the "double" responses given by this group. Of the eight children who were scored as "doubles" on the bag questions, this simulation matches the repeated "double" scores of four of these children on the tube questions.

GENERAL DISCUSSION

Successes and Failings

The memory-updating explanation of the Smarties task is outlined by Morton (1997), and the 3-buffer architecture used here to simulate this theory was derived from Barreau (1997), (see Barreau, 1997 for an account of why a 3-buffer system is necessary). The resulting simulation, however, differs in significant ways from either of these accounts. It is intended to be a forerunner of a number of such simulations, building up a set of mutual constraints on later models of on-line processing by this age group (c.f. Barnard, 1985). As such, it has a number of distinct successes and flaws. Not least amongst its successes is that it is - to our knowledge - the only fully specified computational theory of 3- and 4- year olds failings at "false belief" tasks. Other accounts of these phenomena rely upon the assumption that children of this age suffer from a conceptual deficit in representing the beliefs of others, and their own earlier beliefs if inconsistent with current reality (e.g., Hogrefe, Wimmer & Perner, 1986; Perner, Leekam & Wimmer 1987), or else are in other ways not as completely specified as the account given here (Halford, Wilson & Phillips, in press).

Viewed as a modelling project in its own right, a number of flaws become evident with the current account. Firstly, if it is considered to be a straightforward account of the current data independent of theoretical statements put forward elsewhere (Barreau, 1997; Morton, 1997), then it suffers from a rather poor fit to the data in the case of group B, the "reversed" response children. The mechanism which allows for a reversed response to the bag questions should also produce reversed responses for the tube questions. However, the majority of children in this group (six out of eight) were scored as correct in this case.

Elsewhere, the fit to the data is better. The account given by the basic bag simulation is also able to account for the failure of children at the Smarties task with no change to the model, merely altering the input to simulate the change in task. This simulation correctly produces the same results as the "correct" group (A) on all the questions. The modified simulations for groups B and C also give the identical patterns of results to the children they were intended to model for the bag questions, and in the case of group C (the "double" responses) this success is repeated with the simulation giving the same results as the largest subset of these children.

The conclusion to be drawn from this pattern of success and failure is that although there is a large degree of agreement between the performance of the children and that of the underlying model, there is a flaw in the manner in which the model operates. In particular, it should not function in the same way in response to the tube questions as it did to the bag questions. There are two broad ways of accomplishing this. The first is to add other rules which would interpret the material in the record in response to questions concerning the tube. A backwards inference using rule 7 concerning belief could take

action(empty(tube,bag))
in(bag,smarties)

and come up with

in(tube,smarties)

to go along with the *in(tube,[])* already available in the record. The ordering of these two contradictory options in the buffers could give rise to the differences in responding to the tube questions among the children in group B.

The second general approach to the mismatch is to change the way in which the Group A children solve the questions. One approach is to create records of questions and answers. This would make the answer to the initial belief question available, even though the primary representation *in(tube,smarties)* has been deleted. Use of the record

```
record([[in(bag,smarties), in(tube,[]), object(bag),
    action(empty(tube,bag)), object(tube)]
    action(remove(tube))]).
```

would then be restricted to questions about the tube. This resembles the account given by Barreau (1997). To achieve all this, we will have to characterise the differences among the three groups of children somewhat differently. Both these options will be explored in the next phase of simulation.

ACKNOWLEDGEMENTS

The experimental data which provided the impetus for this simulation were provided by Sofka Barreau. The resulting simulation has greatly benefitted in various ways by discussions with Paul Abeles, Sofka Barreau, Rick Cooper, Cathy Dillon, John Fox and Kerry Sims.

REFERENCES

Anderson, J. R. (1996). *Rules of the mind.* Erlbaum, Hillsdale N. J.

Barnard, P.J. (1985) Interacting Cognitive Subsystems: A psycholinguistic approach to short term memory. In A. Ellis (Ed.) Progress in the Psychology of Language, (Vol. 2), Chapter 6, London: Lawrence Erlbaum Associates, 197-258

Barreau, S. (1997) Developmental constraints on theories of memory. PhD thesis, Department of Psychology, University College London.

Cooper, R., & Fox, J. (in press). COGENT: A visual design environment for cognitive modelling. *Behavior Research Methods, Instruments and Computers.*

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* MIT Press, Cambridge, Ma.

Halford, G., Wilson, W. H., & Phillips, S. (in press). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences.*

Hogrefe, G.-J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development*, 57, 567-582.

Morton, J. (1997). Free associations with EPS and memory. *Quarterly Journal of Experimental Psychology,* 50A, 924-941.

Morton, J., Hammersley, R. H., & Bekerian, D. A. (1985). Headed records: A model for memory and its failures. *Cognition*, 20, 1-23.

Newell, A. (1990). *Unified theories of cognition.* Harvard University Press, Cambridge, Ma.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, 5, 125-137.