Perception, 1983, volume 12, pages 21-34

# A model of the development of the early infant object concept

George F Luger
Department of Computer Science, University of New Mexico, Albuquerque, NM 87131, USA
T G R Bower, Jennifer G Wishart
Department of Psychology, University of Edinburgh, Edinburgh, Scotland

Received 7 October 1981, in revised form 19 July 1982

**Abstract.** A computational model is proposed for the early stages of development of the object concept in infancy. Stages in development are represented as a sequence of grammars or rewrite rules that parse a set of perceptual phenomena. The infant's changes between developmental stages can be described by differences between the grammar rules that model each stage. The program replicates five studies by Bower et al on the development of the object concept and reaffirms the primacy of rest and motion parameters as explanatory invariants in early object-concept development.

## 1 Introduction

The development of the object concept in infancy has attracted a great deal of psychological interest, particularly in recent years (for reviews see Elkind and Sameroff 1970; Gratch 1975; Harris 1975, 1983; Schuberth 1982). One reason for this is that the rate of development of the object concept seems to be the only change in infancy that predicts later rates of change in cognitive functioning (Wachs 1975). Our paper is a beginning attempt at a computer simulation of that development. Some of the psychological data relevant to this attempt are set out in subsequent paragraphs. At this point we would like to clarify the presuppositions of our approach to this and other problems in cognitive development; these presuppositions have shaped not only this paper but also prior and emergent papers (eg see Bower 1979, 1983; Wishart and Bower 1983).

The basic problem which concerns us is development, ie change in functioning with age. For us this requires the description of the age-related stages of any cognitive change before any description of the nature of the transition between these stages. Descriptions of transition processes, it seems to us, are impossible without precise descriptions of the states at the beginning and the end of these transitions. Although the above may seem a truism, it leads into conceptual problems which psychologists at least rarely face: that is, the status of descriptions of behaviour. As will be seen below, our model is not based on records of specific S-R behaviour (eg eyes move left-stop-return 40°) but rather on higher-level descriptions of these. The term 'level' is used here in the sense used by Russell (1910) and introduced to psychology by Bateson (1972). In our own research on object-concept development we have looked at behaviour and tried to formulate first- or second-level rules to account for that behaviour. We have then attempted to test the validity of these rules by presenting babies with new situations in which, if the baby operated by these rules, behaviour would go one way, if by different rules, in another way (eg see Bower and Paterson 1972; Bower 1974)

It should be clear that the rules begin in the mind of the adult experimenter. The adult experimenter is, however, trying to infer the rules that operate in the mind of the baby. If the two match up, we would have a complete theory of at least one segment of development. The computer model is important in that a computer, unlike one's fellow scientists, will not be confused by an ambiguous sentence. To try then to make a computer act like a baby, one must be very clear about the rules one is trying to attribute to the baby. The term 'rules' is very important: in development most psychologists would admit that the same rules or concepts can be indexed by different behaviours, eye movements, hand movements, or even mouth movements in the case of the object concept (Gouin-Décarie 1969). A computer has neither eyes, hands, nor mouth but this is irrelevant. The model is intended to simulate rules and rule changes, without reference to specific physiological or behavioural details.

This is not unimportant in evaluating the model. The computer model we use has, as zero-level characteristics (Bateson 1972), a 'snapshot' model of perception (Neisser 1976) and serial processing of data. Neither of these are necessarily true of the real live human infant. They may or may not be 'true' (Gibson 1969; Rosch 1977). What we are concerned with is what is done with the output of these zero-level processing systems. For our present purposes the nature of zero-level processing is as irrelevant as the precise details of the associated biochemical processes. Until there are data, from the zero level, that indicate that our first-level descriptions are wrong—and as yet such data do not exist—we feel we can proceed with our analysis. The emphasis, we must again insist, is not on the *particulars* of behaviour but on the rules or concepts that generate such behaviour.

## 2 The identity theory of object-concept development

Forty years ago Piaget first observed and described the problems infants have in understanding the nature of objects (Piaget 1936, 1937, 1946). Piaget identified six invariant stages of development through which the infant comes to cope with and understand external objects. According to Piaget, the symbolic processes necessary for later cognitive activity and development have their origins in these acquisitions of infancy; the cognitive achievements of infancy form the foundations from which mathematical thought and logical reasoning will develop. Although Piaget's descriptions of the six stages are still widely accepted, his account of the underlying cognitive processes is often disputed.

The theory which underlies the model presented in this paper attempts to bring two components of the development of the object concept under a single umbrella: the two components are the development of object permanence (Piaget 1937 *The Child's Construction of Reality* chapter 1) and the development of comprehension of spatial relations (chapter 2 of *The Child's Construction of Reality*). Of the two, object permanence is the most studied. The development of object permanence is often described as overcoming an 'out-of-sight-out-of-mind' strategy for dealing with the world. The young infant is described as believing that objects that are no longer seen no longer exist. With development that belief is overcome. The out-of-sight-out-of-mind hypothesis has been under attack since its inception. Piaget himself in chapter 2 of *The Child's Construction of Reality* pointed out that, for one stage at least, exactly the same behaviour could be elicited by a visible object placed upon another object. This observation has been replicated and refined by several subsequent investigators, who have shown that all of the age-typical behaviours normally elicited by hiding an object can be elicited by placing the object on another object (Bresson et al 1977; Wishart 1979; Wishart and Bower 1983). In addition it has been shown that all of the same age-typical behaviours can be produced by placing one object inside or behind another transparent object (Bower 1967; Harris 1974; Butterworth 1977; Neilson 1982). In neither of these cases is there any question of the target object being out of sight, yet typical out-of-mind behaviour is elicited.

Since featural information is incorporated in the rule for identifying an object, any event sequence that violates these new corollaries will now be treated by the infant as the replacement of the original object by another object rather than as a transformation process, as on rule 1. As a result, search behaviour in such situations (eg where an object is covered by a cup) will be directed first by the rule:

> To find an object that has disappeared mysteriously, remove the object which has replaced it (which will lead to successful search in the stage III–IV transition task).

And then by the more specific rule:

> To find an object that has disappeared mysteriously, remove the object which is now in the place where it was last seen (which will lead to successful search in the stage IV–V transition task).

It can be seen from the above that although a separate search rule is proposed to account for the behaviour seen at each of the first five stages of object-concept development, only two basic conceptual rules are believed to underlie these changes. Development between stages III and V results from elaboration of search behaviour rules rather than from any true advance in object understanding.

This paper describes a computational model of the above set of rules. The computational model will be run in several experimental situations and the results discussed in terms of the notions of *motion* and *place* proposed by Bower and Wishart as explanatory invariants of the early developing object concept.

In the next section the computational model will be described. Section 4 will contain the experimental tests of the model. The final section will discuss the results of these experiments and offer an explanation of the differences in development by comparing the rewrite rules that explain the behaviour at each stage. The paper will close with a description of the continuing research.

## 3 The computational model

### 3.1 *Description*

The rules proposed by Bower and Wishart (see the preceding section) can be formulated as a sequence of grammars or rewrite rules that translate perceptual phenomena into sets of behavioural responses. The perceptual phenomena will remain constant throughout the period of testing, ie across the running of the sequence of grammars that represents the early stages of development.

The a priori commitment of this model of development should be made clear. Two issues are critical: first, there is no interaction between the percept and the cognizing subject that in any way changes the nature of the percept. Rather the changes come in the cognizing subject's transformation of or interaction with the percept. This facet of the model is meant to capture the Bower–Wishart notion of the primacy of motion and rest as explanatory invariants in the early stages of object-concept development.

Second, this belief in the primacy of perception allows description of its origin and presence according to a number of differing theories (Marr 1976, 1978; Ullman 1978; Kolers and von Grunau 1976). This particular program does not parse retinal arrays to detect edges or perform figure-ground separation. (However, it is able to detect boundary violations such as partial occlusion.) The perceptual variables given to the sets of grammars include position, size, colour, and shape. Further, the perception of motion and changes of motion by the calculation of differences in positions over time is an irrelevant implementation detail. That is, like feature extraction, how this is accomplished in the human is an empirical question to be addressed by researchers who are considering these aspects of human response (references as above).

This paper hypothesizes that the symbolic output of the featural and motion detection mechanisms is available to the cognizing subject. As stated above, this model offers no explanation of the physiological origins of such phenomena, but rather emphasizes the descriptive adequacy of the internal symbol structures and the interpretive adequacy of the cognizing subject's manipulation of such symbol structures. Further, the changes in the computational rules that express the interpretive adequacy of infants at various stages in development will offer explanation of that development.

Each experiment of this study will be composed of a sequence of 'snapshots' representing the physical situation according to a time-parameter. Snapshots represent objects by themselves, partially or totally obscured by occluders, and replaced by other objects. An example of such a set of snapshots may be seen in figure 1 (see below). This figure represents a subset of the snapshots taken from experiment 3 of section 4, where $t$ indicates the time parameter.

A set of symbols represents each snapshot and a set of rules characterizes the grammar that interprets each sequence of snapshots. Each rule of a grammar represents a different interpretive capacity of the subject such as to locate an object symbol structure within a fixed radius $r$ of a spatial position $(x, y, z)$. This rule takes position $(x, y, z)$ and radius $r$ and returns an actual object structure at location $(ox, oy, oz)$ such that

$$r^2 \leqslant (x-ox)^2+(y-oy)^2+(z-oz)^2 .$$

A further rule checks for parallax in an object symbol structure by checking whether the structure 'has' mass or volume. This is accomplished by testing it from two slightly different views. Each rule of a grammar is a procedure (in PROLOG, cf next section) for interpreting the sequence of snapshots.

In particular the grammars of this model are designed to implement the rules outlined by Wishart (1979) and presented in section 2. An implicit assumption is that the infant is motivated to maintain contact throughout the event sequence with the object initially identified as interesting (see below).

### 3.2 *Implementation*

This computational description of the object-permanence phenomenon is written in PROLOG (Warren and Periera 1977). The action of PROLOG is that of a unification algorithm which operates on a set of record structures. These structures are of two general forms: a set of facts and a set of inference rules. The PROLOG facts are used to make up the object structure for the description of each snapshot. For example the facts loc(obj*n*, *x*, *y*, *z*, *t*), colour(obj*n*, *cl*, *t*), and size(obj*n*, *sz*, *t*) indicates that obj*n* has colour *cl*, size *sz*, and location $(x, y, z)$ at time $t$. The combination of these and other descriptors make up each snapshot of the experiment. The grammar rules interpret these object structures.

PROLOG rules are of the form "A ⊢ B, C, D", which may be described procedurally as "to accomplish A attempt to accomplish B and C and D". B, C, and D may be facts (checked to be true) or may themselves be rules that lead to the proof or performance of B, C, and D. For example, grammar 1 says to test for a permanent object at a location

(i) look for an object structure within a fixed radius $r$ of the location (described in section 3.1).

(ii) Check whether previous snapshots would indicate that a permanent object should be at this location.

(iii) Test the object structure for interest (parallax, as described above).

(iv) Check whether the object structure is intact (that is whether it or any of its boundaries are occluded). And finally,

### 4.2 *The results*

4.2.1 *Experiment 1.* Grammar 1 identifies an interesting object, say OBJ1, and associates this object with a path across the field of view. When OBJ1 stops moving grammar 1 continues to follow the former path, not finding an object it looks back to the new stationary object, calls it OBJ2, and looks back searching again for OBJ1 on its original path. Grammar 2 in this situation identifies OBJ1 as interesting and follows it. When it stops, grammar 2 continues on the path and then looks back to the stopped object, notices its similarity in size and shape to the original OBJ1 and continues to look at it.

4.2.2 *Experiment 2.* Grammar 1 identifies an interesting object, OBJ1, stationary before it. When the object starts to move, grammar 1 follows it, looks back to its former location, and then looks to the new moving object and calls it OBJ2. Grammar 2 does the same as grammar 1 except that it recognizes the size and shape of the moving object to be the same as that of the original stationary object, and does not create a new object.

4.2.3 *Experiment 3.* Grammar 1 identifies OBJ1 and follows its path, ignoring the midpath substitution. Grammar 2 follows OBJ1 as interesting, and, on noting the perceptual changes of the substituted object, calls this new object OBJ2, looks back to the previous object, and continues to follow the new object. For a detailed trace of experiment 3 see the appendix.

4.2.4 *Experiment 4.* Grammar 1 sees OBJ1 as interesting and follows it along. It continues to follow the path behind the occluder and follows the path as the new object emerges. Grammar 1 sees only one bounded volume of space moving on a path. Grammar 2 notes perceptual changes and sees two objects: the original and the new emerged object.

4.2.5 *Experiment 5.* With every movement and pause, grammar 1 sees a different object for each movement and each place. Grammar 2 sees only one object, since the perceptual checks at each motion translation allow transition to new motion sequences without confusion.

## 5 Summary and conclusions

Several researchers have recently reported on computational models of infants' and young children's problem-solving skills (Prazdny 1980; Young 1976). The behaviours modelled in these studies are nonverbal, such as simple eye movements, reaching, and gross body movements. Modern research techniques such as video tapes, however, allow researchers to be quite comfortable with the interpretation of such motor behaviour. Such data make up the Bower and Wishart studies, and these are the data our program interprets.

Grammars 1 and 2 follow fairly closely the two rules suggested by Wishart (1979) and presented in section 2. Grammar 2, with its perceptual checking for size, shape, and colour factors, allows proper detection of changes between the two objects when these factors change, and proper positing of no change when only the motion and rest factors differ. Thus *motion* and *rest* are established in the 'identity' model as explanatory invariants of the first stages of infants' acquisition of the object concept.

Both in experiments 1 and 2, grammar 1 identifies two objects, one stationary and the second in motion. Grammar 2, with more advanced perceptual checking, notes the similarity of the object in motion and the object at rest and does not posit a second object.

In experiments 3 and 4, grammar 1 sees only one object that follows a fixed path. Grammar 2, on noting perceptual differences, posits new different objects for each perceptual change in the objects.

In experiment 5, grammar 1 posits a new object for each path of movement and each stationary object position. Grammar 2 sees only one object.

There are several directions for continued research. The first is to develop further the two grammars presented in this paper. With 'disappearing' objects, for example objects hidden by a screen, there is an age × time interaction that is not yet incorporated. It should also be possible to vary an object's path over time to test the robustness of the 'path' hypothesis.

An even more important area for continued work is to develop a third grammar capable of interpreting object structures in more complex situations than those dealt with above: in experimental situations, for example, that involve partial occlusion of an object or close spatial interaction between a number of objects. Bower and Wishart have posited a third conceptual rule (rule 3) to account for these final stages of the development of object understanding (Piaget's stage VI). This rule states that two or more objects cannot be in the same place or on the same path of movement at the same time unless they bear a spatial relationship to each other which involves the sharing of common boundaries. The interpretive adequacy of this rule could be evaluated against experimental situations in which just such boundary sharing occurs, as, for instance, when an object is placed on top of a platform, inside a cup, or behind a screen (Wishart and Bower 1983). Experiments of this form make up a valuable part of the research to explore the object concept and will be a focus of the next stage of this research.

The final aim of the research will, of course, be to move from describing each of the stages to modelling the rules which produce the actual changes upward from one conceptual stage to the next. A series of cost-gain acceleration studies with infants is at present in progress in an attempt to produce data which will give us some insight into these mechanisms for change (Bower 1981). Change seems to occur when the infant's erroneous descriptions of an object give rise to an uneconomic description of what is actually happening.

One last point is worth mentioning because of its implication for future attempts at modelling infant cognitive development. Prazdny has also recently produced a computer model of the rule 1 behaviours described above, a fact which might seem to render the present attempt redundant (Prazdny 1980). Prazdny examined twelve Bower tracking experiments. He then excluded one of them (experiment 5 here) from his model for reasons that are not entirely clear to us. In the other eleven situations he suggested that the experimental results could not fully support the place-movement analysis of Bower. He quoted one experiment in particular which he felt clearly demonstrated that stage II infants do attend to features and do not identify objects solely in terms of places and movements, an experiment in which infants showed upset at an unseen replacement of an object (Prazdny 1980, experiment 7). This would indeed undermine the analysis presented above. Such an experiment, however, although attributed to Bower (1977) has not to our knowledge been done by anyone as yet, least of all ourselves.

Prazdny also rejected the Bower analysis of some of the other tracking results on the basis that an infant would not look for a 'missing' object (eg producing place or movement errors) if he was as dependent on immediate visual input as this analysis would have it. This represents a very basic misunderstanding of our analysis. It is central to the entire analysis that the infant's tracking behaviour is directed by conceptual rules and *not* by straightforward perceptual input. The missing experiment 12 of Prazdny (experiment 5 here) demonstrated this point conclusively—in that experiment the infant tracked forward or back to an empty space, a behaviour that obviously cannot be directed by any perceptual input.

## APPENDIX

**A more detailed explanation and trace of the computer analysis of experiment 3**
To begin the experiment the object structures are created. In the case of experiment 3 this includes a sphere which moves across the field of vision and changes, in full view of the infant, to a cube. The cube continues to move on the same path and with the same speed as the sphere.

To start then, an object structure of type sphere is created at time $t = 1$. The object is to move in three-dimensional space with coordinates $(x, y, z)$ in front of the infant. The $x$ parameter will change from 0 to 120, the $y$, height, will be constant at 4, and $z$, the depth, a constant at 10. The infant is sitting at location (60, 0, 0). The object structure is then created over the time period, here from 1 to 30 with $t = 15$ occurring when the object is in the middle of the path at (60, 4, 10). For each time $t$, the object structure, called OBJ1, is created (asserted into the data base) with location, size, shape, and colour variables attached. For example, at $t = 1$ the object structure is (comments in brackets):

| | |
|---|---|
| OBJ1 - time 1 | the object, time 1 |
| loc(4,4,10) | location $(x, y, z)$ |
| shape(s) | shape is spherical |
| size(3) | radius of 3 |
| colour(g) | green |

The sphere structure continues until $t = 16$ when OBJ1 turns into a cube. The structure for $t = 15$ and $t = 16$ is:

| OBJ1 - time 15 | OBJ1 - time 16 |
|---|---|
| loc(60,4,10), | loc(64,4,10) |
| shape(s) | shape(c) |
| size(3) | size(3) |
| colour(g) | colour(g) |

The cube continues to move until $t = 30$, when it is at location (120,4,10), at which time it disappears.

This concludes the 'object creation' phase of the program. Now any particular set of rewrite rules (to represent analysis at that 'stage' of development) will consider the object structures created. The code for grammar 1, presented in the text, is called to consider a point (3,7,10) near the object structure. The actual object structure at (4,4,10) is located, found interesting (parallax), and under the motion–rest invariant is expected at the appropriate new location. OBJ1 is not occluded by any other object throughout the sequence of time changes. In the trace it is referred to both by its number (the number 1,2,3, ... of 'interesting' objects found in the sequence) and by its shape, first a sphere and then changing to a cube. A sample of the trace follows.

Object 1 has interest at time 1
and is expected at place (4,4,10).
The sphere called Object 1 is not occluded at time 1.

Object 1 has interest at time 2
and is expected at place (8,4,10).
The sphere called object 1 is not occluded at time 2.

.
.
.

.
.
.

Object 1 has interest at time 15
and is expected at place (60,4,10).
The sphere called Object 1 is not occluded at time 15.

Object 1 has interest at time 16
and is expected at place (64,4,10).
The cube called Object 1 is not occluded at time 16.

.
.
.

Object 1 has interest at time 30
and is expected at place (120,4,10).
The cube called Object 1 is not occluded at time 30.

There are no further objects in view at time 31.

To run grammar 2, the sequence of object structures is recreated (exactly as above) and the code for grammar 2 is called. The trace is in most ways the same as that for grammar 1 except that a perceptual check for colour, size, and shape is made at each step. The trace is as follows:

Object 1 has interest at time 1

.
.
.

Object 1 has interest at time 15
and is expected at place (60,4,10).
The sphere called Object 1 is not occluded at time 1.

Object 1 has interest at time 16
and is expected at place (64,4,10).
Object 1 changes shape between times 15 and 16,
call the result Object 2.
Look back for previous Object 1.
The cube called Object 2 is not occluded at time 16.

.
.
.

The cube called Object 2 is not occluded at time 30.

There are no further objects in view at time 31.