PIAGET AND ARTIFICIAL INTELLIGENCE

Jarrett K. Rosenberg
Department of Psychology
University of California, Berkeley 94720

ABSTRACT

Piaget's Genetic Epistemology and Artificial Intelligence can be of great use to each other, since the strengths of each approach complement the weaknesses of the other. Two ways of bringing the two approaches together are suggested: elaborating the parallels between them (such as the concept of schemata), and building AI models based directly on Piaget's theory. How this might benefit AI is illustrated by examining how Piagetian research on problem solving can suggest new ways of building programs that learn.

I. INTRODUCTION.

My thesis in this paper is that two such superficially disparate areas as Piaget's Genetic Epistemology and Artificial Intelligence can inform each other immensely. This fact has already been noticed, as long ago as Seymour Papert's work at Geneva [1] and as recently as Margaret Boden's writings on Piaget [2, 3]. Nevertheless, I think it still needs to be pointed out how the two fields both parallel and complement each other, and how one might go about trying to connect them. Piaget's work is immense, stretching over half a century and dozens of volumes, and so only a glimpse of it can be given here, but hopefully it will be enough to stimulate some AI people (and maybe psychologists too).

II. WHY SHOULD PIAGET AND AI BE MUTUALLY RELEVANT?

To address the problem of artificial intelligence is to presuppose what natural intelligence is, but AI work in general doesn't really treat the latter in any comprehensive way; it's assumed that playing chess is intelligent, or planning, or ordering a hamburger. But can we really hope to understand intelligent behavior, natural or artificial, without deciding what criteria we use for intelligence? If the answer is No, then it seems natural to look to psychology for a theory of intelligence. The only theory I find satisfying, and I suspect most AI people will as well, is Piaget's.

Piaget's theory is relevant to AI because it is a fully developed, *motivated* theory of intelligence which is general enough to subsume not only human and animal intelligence, but purely artificial intelligence as well. By motivated I mean that the form of Piaget's theory of knowledge is based on both the *logical* requirements of epistemology, and on the *biological* requirements of behavioral evolution. Hence there are good reasons for accepting his approach

to intelligence, just as there are few, if any, compelling reasons to believe the (often implicit) definitions given in AI (or psychology, for that matter). Part of my argument, then, will be that AI researchers might be able to use this foundation as a basis for their own work.

Piaget, on the other hand, can learn a lot from AI, since it has developed ways of looking at the detailed structures and processes of complex behavior, thus suggesting ways of elaborating Piaget's theory into one with a more "psychological" flavor. Piaget's theory has always been more epistemologically than psychologically oriented (in the sense of information-processing psychology), and it has only been recently that Genevan researchers have addressed the question of the *procedures* involved in the acquisition and use of knowledge-structures.

Table 1 gives a general (and biased!) comparison of the strengths and weaknesses of Piaget's theory and AI work taken as a whole. The important thing to note is that the strengths and weaknesses of the two approaches complement each other nicely. Let me discuss them briefly.

The advantages of Piaget's theory are:

▶ It is both epistemologically and biologically motivated, that is, it starts from the most central issues of a theory of knowledge (e.g., how space, time, and causality are constructed), and places this analysis within the biological context of human evolution. This contrasts with most AI work, which considers intelligence outside of both its philosophical and biological context.

▶ It is highly theoretical and comprehensive. As a biologist with philosophical interests, Piaget resolved to study the "natural history of knowledge", and so developed a general theoretical framework which he then applied to a number of areas, particularly those of physical and logico-mathematical knowledge. This contrasts with the general lack of interest in AI for large-scale theories or applications.

▶ It has a large amount of empirical validation; virtually everything he has proposed has been extensively replicated and consistently extended.

▶ It proposes a "fundamental unit" for cognition, the *scheme*, from which all knowledge-structures are constructed. This supplies a uniform analytic framework for all research in cognition.

▶ It has a formalism for the structures of knowledge. This formalism is an algebraic, rather than a computational one, and is only in its beginnings, but is nevertheless valuable.

▶ It is a *developmental* theory of knowledge, promising a way to understand how knowledge-structures can radically change and yet preserve their integrity. It can thus provide an additional, developmental, set of constraints on AI theories: both structural ones (certain logical properties of cognition, with a logical and developmental ordering on them), and functional/procedural ones on the way those structures are constructed and used (e.g., forms of coordination). This should be one of its greatest attractions for AI.

The disadvantages of Piaget's theory are:

▶ It is not very detailed about how knowledge is represented, or how it is processed; to someone looking for a theory at that level, Piaget's seems mostly descriptive.

▶ It says little about the *procedures* used in cognition; only recently has this issue been addressed at Geneva.

▶ Its emphasis is on physical and logico-mathematical knowledge, rather than other kinds such as social or linguistic.

Here the field of AI has to offer precisely what Piagetian research lacks, and a glance at the table shows that I consider the converse to be true.

III. HOW CAN WE PUT THEM TOGETHER?

Given that we want to bring these two areas into contact, two possible ways of doing it come immediately to mind.

First, there are a number of parallels between Piagetian concepts and ones introduced in AI, the most obvious being that between Piaget's notion of a scheme and the schema/frame/script ideas in vogue the past few years in AI. The most important such parallels are:

▶ Schemata. For Piaget, the fundamental unit of knowledge is the *scheme*, a generalizable action-pattern.

▶ Schematically-based processing. As a consequence of his scheme-based theory, Piaget views cognitive processes as primarily top-down.

▶ Knowledge as action. In Piaget's epistemology, no knowledge is gained without acting and its consequent transformation of experience.

These parallels can be used not only to suggest new ways of developing the AI notions, but as ways of formulating the Piagetian concepts in more detail.

Second, we can attempt to create AI models of Piaget's *theory* (not tasks, as has sometimes been done). This will introduce another paradigm into AI research, no worse than the others and hopefully better. Simultaneously, such work can serve to rigorously develop the processing aspects of Piaget's theory, using all the insights of AI research. Let me give a brief example.

One of the ways that Piagetian theory is relevant to AI is that it can give both theoretical and empirical suggestions for how to design programs that learn, i.e., develop new cognitive structures in order to solve problems. Current AI learning programs work by assuming some initial body of knowledge and learning techniques, and then showing how additional knowledge and techniques could be learned. However, this creates a regress: how the starting knowledge and techniques are acquired is left for someone else to explain. In addition, the initial knowledge used is often much more advanced than what is to be acquired; for example, Sussman's HACKER uses fairly sophisticated debugging techniques to learn something children acquire by age two. Piaget's theory, on the other hand, tries to account for all of cognitive development starting with only the simplest of structures.

Piagetian research reveals a number of important characteristics of problem solving that are not fully achieved until adolescence, being gradually constructed starting around age five (ignoring the sensori-motor precursors). They are:

▶ The ability to construct and overgeneralize theories, and as a corollary, to interpret phenomena in terms of confirmation or disconfirmation of those theories.

▶ The ability to construct *refutations* as tests, as well as confirmations.

▶ The presence of articulated part-whole relations, e.g., as reflected in an appropriate decomposition of the problem.

Table 1. A Comparison of Piagetian Theory and Artificial Intelligence.

Piaget		AI	
Strengths	Weaknesses	Strengths	Weaknesses
1. Motivated (biologically & epistemologically)	1. Not detailed	1. Precise/formal	1. Unmotivated
2. Comprehensive theory	2. Not procedural	2. Process-oriented	2. Not theoretical
3. Lots of empirical validation		3. Detailed	3. Little empirical validation
4. Fundamental unit of knowledge		4. Body of techniques	4. Little agreement
5. Formal model of knowledge structures			5. Narrow
6. Developmental theory			6. No model of development

▶ The ability to reflect on procedures in order to switch or modify them.

▶ The ability to coordinate goals and subgoals.

Piagetian studies of how these abilities are constructed can provide useful ideas for AI. Consider, for example, how children learn to balance blocks (see [4]).

For very young children, actions of objects are more or less random unless assimilated to their own actions and intentions. Their attempts to balance blocks ignore completely the relevant factors in favor of pushing down on the block to maintain a balance. Somewhat older children, after *much* trial and error, recognize a regularity in the outcomes of their efforts, namely, that most things balance at their geometrical centers. Once this idea (a primitive theory) is grasped, it is rapidly generalized, to the point where exceptions to it are dismissed as chance. The eventual overgeneralization of this notion of regularity, and consequent repeated failures to confirm it, give rise to the idea that the exceptions to the theory may have regularities themselves. At this point, children will retain their original theory where it works, but construct an additional one for the counterexamples (balance based on weight distribution). As their powers of conceptual coordination grow, they can combine the two theories into a single, more general, one. Then finally the idea of a refutation test usually appears (the preference for confirmations stemming from the same source as that of overgeneralizing in the first place).

What does this suggest for AI work on learning and problem solving? If Piaget is correct, then in order to be generally successful, AI programs should have the abilities mentioned, and they should be able to *develop* them in the manner given above. In particular, we can imagine a class of AI programs that approach problem solving in the following way. Faced with a new problem they would:

▶ engage extensively in trial-and-error behavior,

▶ strive to find regularities in the outcomes (including regularities in the procedures used),

▶ use the regularities to construct and then overgeneralize a theory,

▶ construct confirmation tests of the theory and note the exceptions,

▶ construct a theory for the exceptions, if they are numerous enough,

▶ try to coordinate the two theories into a larger one,

▶ test this resultant theory by confirmation and refutation tests.

Besides theories of problem situations, the programs would also construct theories of the problem-solving procedures themselves, i.e., they would study the relationships of procedures among themselves and with respect to their outcomes, rather than just theories of the relationships among states of the world.

Some of the above is partially captured in various AI programs (Sussman's HACKER, Winston's arch builder, etc.), but none of them are configured in exactly this way, and all of them assume fairly sophisticated techniques and knowledge as a basis for learning. The proposal above is obviously not the last word in learning and problem solving, but at least provides a useful place to start.

In addition, the vast amount of empirical data that Piagetians have collected can be invaluable as an empirical testing ground for AI models. Up to now, AI researchers have only attempted to model performance on Piagetian tasks, without realizing that it is both the empirical *and* the theoretical constraints that would make Piagetian work of use to AI: were a model to simulate performance on a Piagetian task, and do so in a way fully consistent with the theory, it would have met a very strong test of its validity. Moreover, the tasks themselves have a high degree of ecological validity and epistemological significance (for example, conservation of matter and number).

Thus, by making use of the existing parallels between the two approaches, and by recasting Piagetian concepts in AI terms, the two might produce a hybrid that would have all the advantages of a Piagetian foundation as well as all the benefits of AI's hard-won insights into knowledge-representation and information-processing.

For this to happen, however, it will be necessary for both Piagetians and AI researchers to learn more about each other's work; as mentioned above, previous AI work has neglected Piaget's theory in favor of his tasks, while Piagetians have only the vaguest notion of what a computational model is all about. And of course, Piaget's theory is not some sort of panacea—psychology's gift to AI. On the contrary, it needs to be much more developed, along the lines that AI work is pursuing. It is precisely because each could profit from the other that I've presented these points here.

REFERENCES

[1] Papert, S. "Etude comparée de l'intelligence chez l'enfant et chez le robot." in *La Filiation des Structures. Etudes d'Epistémologie Génétique.* vol. 15. Paris: P.U.F. 1963.

[2] Boden, M. "Artificial Intelligence and Piagetian Theory." *Synthese.* 38: 389–414. 1978.

[3] ———— *Piaget.* New York: Viking. 1979.

[4] Karmiloff-Smith, A. and B. Inhelder. "If you want to get ahead, get a theory." *Cognition.* 3: 195–212. 1975.