Reflective Abstraction and Mathematics Education: The Genetic Decomposition of Induction and Compactness

Ed Dubinsky and Philip Lewin
Clarkson University

I. INTRODUCTION

It is widely recognized by many educators that students do not master the post-secondary mathematics curriculum—concepts such as composition of functions, limits, continuity, compactness of sets, and proof by induction (Moise, 1984). Efforts to facilitate learning generally take the form of revising curricula, utilizing new technology in teaching, or trying to find ways in which teachers can better convey course material through classroom lectures (Strasbourg Document de Travail, 1985). What is usually lacking is an attempt to orient instruction around the actual process of learning (see, however, ADAPT, 1976; and DOORS, 1979). Instead of hoping that students will learn new material by imitating the teacher’s behavior or listening to a lecture, it may be more fruitful to consider the mental processes by which new abstract concepts are acquired. This latter approach requires an understanding or at least a theory of the psychology of concept acquisition.

In our efforts to develop a psychology of instruction for post-secondary mathematics we have found the genetic epistemology of Jean Piaget to be of use. Most attempts to apply Piaget’s ideas to teaching have been restricted to primary and secondary school (e.g., Duckworth, 1979; Schwebel & Raph, 1978). Piaget, however, argued that cognitive development could be conceived as a life-long process, consisting of an ever-ascending hierarchy of interdependent concepts (Beth & Piaget, 1966). Thus, an educator who is attempting to teach new conceptual material, as in mathematics, is in effect trying to induce cognitive development.

At the center of Piaget’s work is a fundamental cognitive process which he termed “equilibration.” Speaking informally, we can describe equilibration as the process by which a knower attempts to understand a given item of information by situating that item in the knower’s overall cognitive system. Such situating occurs as the knower cognitively constructs an understanding of the item through a process called “reflective abstraction” (to be described in detail below). We believe Piaget’s epistemology is rich enough to provide a framework
for a general psychology of instruction, not only for mathematics. However, our concern in this essay is with post-secondary mathematics, in particular with two fundamental concepts of the undergraduate mathematics curriculum, proof by induction and the compactness of sets.

In this paper, we use Piaget’s epistemology to explore how students either learn, or fail to learn, these two concepts. Through an application of the model of equilibration to a series of student interviews, we have been able, we believe, to generate an account of the arrangements of component concepts and cognitive connections prerequisite to the acquisition of these concepts. These arrangements, which we call “genetic decompositions,” do not necessarily represent the way in which trained mathematicians understand these concepts; rather, they map the way in which students empirically formulate their understandings for the first time. The implications of such decompositions for pedagogy are, we feel, quite important.

The structure of the paper, then, will be as follows. Given the importance of Piaget’s work to our own, we begin with an interpretation of his epistemology, focusing on the model of equilibration and the concept of reflective abstraction. We then, in the second part, offer a detailed analysis of our empirical work, apply Piaget’s epistemology to it, and describe our genetic decompositions. In a final part of the essay, we explore some of the implications for pedagogy of our findings.

The last section stops short of offering detailed suggestions for classroom implementations of our ideas. We do, however, propose an approach to teaching concepts such as induction and compactness that can be translated into specific activities with students. Work along these lines is in progress and will be reported elsewhere. It includes both the development of teaching methods that are based on the epistemology described in this paper and the results of using those methods in classroom as well as experimental situations.

It might be useful for the reader to know how we conceive our project with respect to the vast amount of theoretical and empirical work focusing on Piaget over the last two decades. We are especially concerned that our understanding of the concept of stages be clear. We believe that we take a fairly orthodox position in conceiving of the concept of stages as heuristic. (Though in order to avoid possible misunderstanding, we prefer to call the phases of comprehension which we found “levels.”) As Piaget noted in 1955:

One tries to construct stages because this is an indispensable instrument for the analysis of formative processes. Genetic psychology attempts to envisage the construction of mental functions, and stages are a necessary instrument for the analysis of these formative processes. But I must vigorously insist on the fact that stages do not constitute an aim in their own right. (Piaget, 1977a, p. 817)

The focus of Piaget’s thinking has always been on the dynamics of epistemology, and the stage conception was simply a means of provisionally organizing a
massive amount of data. "I would compare them to zoological or botanical
classification in biology which is an instrument that must precede analysis" (p.
817). This perspective has become increasingly clear since the mid-1970s, with a
body of research by Piaget and his Geneva associates focussed explicitly on
reflective abstraction and equilibration.

On the other hand, equilibration, even when recognized as a factor of central
importance for epistemology, requires further specification than that given by
Piaget (Boden, 1982). A variety of neo-Piagetian analyses have been developed
as a consequence. These have taken several forms, including attempts to provide
more careful articulations of the process of reflective abstraction itself (e.g.,
Campbell & Bickhard, 1983; Moshman & Timmons, 1982); attempts to include
related variables such as cognitive style (Neimark, 1981), memory load, and
organization of strategies (Case, 1980; Pascual-Leone, 1984); attempts to articulate
stage concepts more complex than those entailed by formal operations (Com-
mons, Richards, & Armon, 1984); and re-formulations of cognitive development
in terms of information-processing models (Greeno, 1976; Klahr & Wallace,
1976; Resnick & Ford, 1981). We have found all this work to be useful, and see
this essay as a further contribution to the exciting task of specifying the implica-
tions of Piaget's work for epistemology and education. Perhaps the most impor-
tant of these implications, and one we wholeheartedly affirm, is that the Piaget-
tian perspective leaves final authority for an individual's cognitive growth with
the individual, not with the peer, parent, teacher, or any one else. It is the
student's responsibility to construct knowledge for herself or himself, not to
duplicate the knowledge structure of someone else, however expert they may be.

However, we should also say what all this work, including our own, does not
and probably cannot do. These various attempts specify, not the structure of the
cognitive moment in which the reflective abstraction itself occurs, but the variety
of pre-disposing structures and influences which come into play at that moment.
In other words, an optimal specification for equilibration seems only able to
explicate all the prerequisite structures, both necessary and sufficient, for the
cognitive act to occur. It can provide the readiness, but the act itself remains
inaccessible and idiosyncratic, dependent on the particular way in which a given
subject notices and organizes her/his experience. It would seem one never has
direct access to cognitive processes—thought is an unconscious activity of
mind—but, at best, only to what an individual can articulate or demonstrate at
the moment of insight itself. Precisely what occurs at that moment seems as
inaccessible as it is essential.

II. PIAGET'S EPISTEMOLOGY
Piaget's epistemology is genetic in that it is concerned with describing how
cognitive structures develop over time according to general rules. Since his
epipistemology is explicitly concerned with the individual only insofar as that
individual is a knower (and not, for instance, as a member of a society or
family), it is convenient to refer to this abstracted entity as an "epistemic subject." Implicit in Piaget's analysis is the idea that knowledge and understanding are acquired only as the epistemic subject applies its existing cognitive structures to that of which it becomes aware, which we will refer to as "cognitive aliments." Aliments are integrated by making the appropriate modifications in one or more cognitive structures. Piaget has consistently argued that we must look at knowledge not in terms of the primacy of the knower (which would imply some version of rationalism or innatism) nor in terms of the primacy of that which lies outside the knower (which would imply simple empiricism or associationism, or a Platonism), but in terms of the primacy of the interaction of the knower with cognitive aliments, as mediated through cognitive structures.

It may be difficult to appreciate how radical Piaget's conception is, wedded as we tend to be to the naive realism of the natural attitude. Few would have trouble agreeing that knowledge is not exclusively a product of our minds, that there is a "real world" out there. But many are less willing to acknowledge that such knowledge as they believe they possess does not correspond to how things "really" are in a fundamental way, but only coheres systematically so that the experience of self and world seems integrated and whole. The difficulty of seeing how extensively our knowledge of the world is a construction can mislead us in our apprehension and attempted applications of Piagetian theory (and much else besides). Descriptors such as "object," "event," "situation," "world," "reality," "environment," and even such apparently neutral terms as "information" or "energy"—everything, in short, of which we claim to have "experience"—must be understood, insofar as we discuss Piaget, to refer to the outcome of cognitive activity. It is the conclusion, the result, of operations, not their starting point. (This reading of Piaget as a radical constructivist is developed by von Glasersfeld, 1985.)

Piaget's epistemology shares with Kant's the central insight that knowers construct their knowledge of the world. Let us clarify what "construction" means in this instance. What is not meant is that knowers create the world solipsistically, so that there is believed to be no reality outside that which a knower has constructed. Rather, while there is a world that exists independently of the knower (the Kantian noumenon) it is argued that it cannot be known "in itself," with pure and direct immediacy, but is known only mediately, through the various capacities—biological, neurological, cognitive—of the knower. Knowledge of the world, then, is "phenomenal," only representing how things are constructed by the knower.

What we directly experience is (in Kant's terms) a sensory manifold, or (in our terms) cognitive aliments, that is, "objects," "events," "experiences," and so on which in the instant of their recognition are already unified, accorded a locus in time and space, and situated with respect to a variety of prior "experiences" in the mind of the knower. The task Kant set for himself was to try to understand the conditions under which such unified experience came about,
which he believed he had found in a variety of hypothesized structures—the schemata of space and time, the Categories of the Understanding, the Transcendental Unity of Apperception—which collectively guaranteed the organized quality of lived-experience.

Thinking of Piaget’s epistemology as a development within the Kantian tradition is a useful first step in understanding it. Piaget too, saw that the unity of experience was a phenomenon to be explained rather than naively assumed. However, a necessary second step is to see in what ways Piaget differs from Kant. He combined an epistemological perspective compatible with Kant with an explicitly biological orientation, and thus expanded Kant’s problem from that of accounting for the possibility of experience through hypothesized mechanisms to that of explaining the genesis of these mechanisms themselves. Thus, Piaget turned toward empirical psychology as a means of investigating epistemology. The task of his epistemology becomes that of describing the operation and construction of those cognitive structures through the application of which experience itself is constituted. In understanding the development of knowledge from Piaget’s perspective, we are concerned with how the knower’s interaction with the world leads it to construct such categories as “object” and “event,” on the one hand, and with how this interaction leads it to re-construct and modify these same cognitive structures, on the other hand. As Piaget observed some 30 years ago, “It is by adapting to things that thought organizes itself and it is by organizing itself that it structures things” (Piaget, 1963, p. 8).

The important differences between Kant and Piaget can be stated succinctly. First, the Kantian schemata and categories are innate and fixed, unchanging over time, while the Piagetian cognitive structures are constructed from the outset and undergo systematic changes of increasing differentiation and hierarchic integration. (Hence, Piaget’s epistemology can be described as a “genetic constructivism”.) Second, the Kantian structures are better conceived in analogy to neurophysiology than psychology. That is, they organize the activity of the knower into experience, but do not themselves perform actively. They simply structure the sensory manifold in accordance with their nature. (While Kant did not trace this nature into evolutionary biology, contemporary neurophysiologists, such as Warren McCulloch and Humberto Maturana, whose work is compatible with that of both Piaget and Kant, have.) Piagetian cognitive structures, in contrast, are most analogous to psychological ones. They are modes of action on the world which function as organized systems of transformation. They are progressively constructed through interaction with cognitive aliments and progressively re-constructed as a result of continued interactions with later cognitive aliments in a process called “reflective abstraction.” It is, in fact, the process of reflective abstraction, and the related concept of equilibration, that lies at the heart of Piaget’s epistemology. In the following paragraphs, we will explore equilibration and reflective abstraction, and describe what we see as their key implications for epistemology and psychology.
III. EQUILIBRICATION AND REFLECTIVE ABSTRACTION

Piaget distinguishes four factors affecting cognitive development: maturation, experience of the physical world, influence from the social world, and equilibration. The first three of these are widely recognized and are given a place within virtually all developmental theories. The fourth factor, equilibration, represents Piaget’s unique contribution to developmental theory. He accords equilibration a position of central importance in his theory because it is the organizing principle of cognitive development. Without it, the other three factors would be ineffective—present, but not integrated. The necessity of acknowledging a factor like equilibration follows from Piaget’s constructivism: If the epistemic subject constructs its knowledge of the world, then equilibration expresses the means by which it does so.

Having created a place for equilibration in the cognitive system, we can specify its meaning more precisely. Equilibration refers to a series of cognitive actions performed by a knower seeking to understand cognitive aliments. The aliment as such, is experienced as novel, resistant, perturbing, disequilibrating—terms which are roughly synonymous—to the cognitive system of the knower. This experience of disequilibration motivates the knower to attempt to re-equilibrate; it is like an itch that must be scratched. In a process called “assimilation,” the knower will apply to the aliment the set of cognitive operations which the knower has previously constructed. In general, the aliment will offer resistance to assimilation, leading the knower to modify its cognitive structures (a process called “accommodation”) until the aliment is no longer experienced as resistant. At this point, the knower has “understood” the aliment, and the cognitive system has re-equilibrated through having re-constructed and re-organized itself. In any given equilibration, there will be a greater or lesser degree of assimilation and accommodation, though both will always be present to some extent. There must always be assimilation, for all construction begins with existing cognitive structures, even if they prove inadequate for the task at hand. And there is always accommodation, for even in the most trivial and automatic cognitive acts, the particular aliment under consideration is understood only as cognitive structures are applied to it. Even in cognitive acts as simple as recognition, that which is recognized must be “re-cognized,” that is, situated within an already existing system of cognitive identifications.

What are some of the implications of equilibration for epistemology?

Since cognition begins with the epistemic subject, equilibration is only motivated when the subject apperceives that a cognitive aliment requires assimilation; it does not occur automatically. Once an aliment is recognized as disequilibrating, the success of re-equilibration will depend, at least in part, on how adequate already existing cognitive structures are to accommodate the aliment. Thus, the integration of novelty can be achieved with greater or lesser degrees of internal reorganization and reconstruction. For instance, if a student’s existing system of
mathematical structures includes quantification of logical propositions and a strong notion of function, then it may be possible to construct the concept of continuity. If these prerequisite structures are not present, however, then the learning experience may have to begin with their construction. This latter situation often leads to long periods of development during which the student may have difficulty in accepting the relevance of this prerequisite academic activity.

Further, since the experience of being disequilibrated and of being re-equilibrated resides in the knower, not in the aliment, it is possible for the knower to err in that he or she might simply fail to perceive the aliment, or to perceive that attempts to integrate the aliment are not succeeding. Moreover, the knower can also be "wrong" in believing he or she has successfully re-equilibrated, while from the perspective of an external observer, it may be clear that he or she has not. That is, a knower can perform the epistemic activity of reflective abstraction, and still be "incorrect" in terms of an observer's "real-world" knowledge. This is indeed a curious result, but it is one that follows directly from the proposition that knower constructs their knowledge for themselves.

Attempts at re-equilibration can be more or less successful, depending on a variety of factors. For our purposes, there are two factors in particular that need to be isolated: (1) the adequacy of already existing cognitive structures to adapt to the novel aliment; and (2) the degree to which the subject recognizes her/his adequacy or inadequacy, and thus the extent of the effort which the subject is prepared to make in order to equilibrate, if existing structures are not sufficiently advanced. (It is precisely at this point that influence from the social milieu of teachers and peers can have the most impact, whether positively or negatively. But note that social influence becomes an important variable only in the context of the attempt to equilibrate; it has no independent operative value.)

The most powerful, and cognitively, the most interesting form of equilibration is that in which particular cognitive structures re-equilibrate to a disturbance by undergoing a greater or lesser degree of re-construction, a process known as "reflective abstraction." We would argue that in those cases in which successful learning occurs, reflective abstraction has taken place. Since this concept is so central to our explanation of the cognitive behavior of our students in trying to learn induction and compactness, we would like to offer a fuller discussion of it.

There are two facets to reflective abstraction. The first is a reflection of one or more structures onto a higher plane in which the structures function in greater generality by being applied to new aliment which can even be structures functioning on lower planes. The second is a reconstruction of these reflected structures into new structures that are distinct from the old ones, although important similarities may continue to be apparent. As we discuss various ways in which reflective abstraction takes place it will be seen that each form of this process is some combination of these two facets. We will mention some examples here, but the illustrations of these ideas that are most important for us in considering
advanced mathematical concepts will occur later in the paper when we discuss induction and compactness.

A simple form of reflective abstraction consists of restriction to essential attributes. This occurs when the subject realizes that a new aliment, initially considered to be disequilibrating, can in fact be accommodated. The key factor is a recognition that in applying particular existing structures, it is necessary that the aliment have only certain attributes, no matter how different the aliment seems otherwise. This form of reflective abstraction can be called "generalization" or "extension." For example, a child's understanding of commutativity of addition could be easily extended to multiplication. Later it can be generalized to include operations on sets such as union and intersection. However, generalization is not an invariably correct strategy. For example, this form of reflective abstraction allows the structure of commutativity to be applied to matrix addition, but not to matrix multiplication.

When this form of reflective abstraction is applied to aliments that are physical objects, as opposed to operations, the process is called empirical abstraction. This distinction between objects and actions is often taken as expressing the essence of the difference between empirical and reflective abstraction, respectively. However, this is not the full story because, as we shall see, there is much more to reflective abstraction than generalization.

A second form of reflective abstraction is the coordination of two or more existing structures to obtain a new structure. This construction can be made in several ways. When a baby learns to obtain a desired object by first reaching and then grasping, two schemas are coordinated in a series. When a child constructs the concept of number by coordination of seriation and classification (in particular, inclusion), the linkage is more in parallel. When we turn to our examples of induction and compactness, we will see much more complicated ways in which structures can be coordinated.

It is often the case that different forms of reflective abstraction can be combined. For instance, one or more structures can be reflected onto a higher plane through generalization and then coordinated to form a new structure whose components are recognizable as, but different from, previously existing structures. Again, we will see examples of this in induction and compactness.

Perhaps the most important and powerful form of reflective abstraction involves a process of encapsulation. At any point in time, an epistemic subject possesses a number of structures which perform processes implicitly, in the sense that inputs are processed to obtain the appropriate output, but the subject is not able to explain what is being done, much less understand the total process. This is what is going on when children use the "INRC group" (as described by Piaget) long before they are capable of understanding mathematical groups. That is, the child begins by developing the ability to perform acts which require each of the four actions I, N, R, and C (i.e., identity, negation, reciprocity, and correlation) separately. Next, they are coordinated into a single system which,
with its two kinds of reversibility, forms an operation. This operation is in fact a 
mathematical group (the Klein 4-group), but the child is not aware of it as such. 
It is only later (and it may not happen for everyone) that the epistemic subject 
sees the operation as a total structure. Reflective abstraction includes the act of 
reflecting on one’s cognitive actions and coming to perceive a collection of 
thoughts as a structured whole. As a result, the subject can now encapsulate 
the structure, and can see it as an alimen for other structures.

When combined with the previous two forms of reflective abstraction, this 
leads to a highly complex and very non-linear organization of structures (e.g., 
see Chart I). A structure is, in some sense, a form, acting on various aliments as 
content. After encapsulation this form can become content for other structures 
which, when generalized, can act upon the encapsulated structure as an alimen. 
The encapsulated structure still exists on a lower plane, and can act on aliments 
appropriate for it when necessary. In this way, the process of reflective abstrac-
tion can construct the very complicated structures required to handle extremely 
abstract logico-mathematical entities.

We should mention, to conclude our description of reflective abstraction, that 
this process takes time. It can take time before a subject realizes that the existing 
structures are not really sufficient to assimilate an alimen and the disequilibrat-
ion has not been removed. More time is spent in generalizing and reconstruc-
ting an existing structure or in coming to see it consciously as a structured whole 
(i.e., encapsulating it). Coordinating different structures may not be in-
stantaneous. Finally, once a new structure has been constructed and the system 
can re-equilibrate by assimilating the disequilibrating alimen, this experience 
may have to be repeated a large number of times with variations of the alimen 
before the new structure becomes stable and permanent.

Returning now to our discussion of equilibration, we describe a detailed 
model which Piaget developed in the mid-1970s (Piaget, 1977b). Although 
imperfect, it is a valuable first attempt to specify exactly what happens in cog-
nitive development. In this model, he distinguished three distinct types of cog-
nitive behavior—called alpha, beta, and gamma behavior—employed by the 
knower as it seeks to comprehend novelty.

At the most naive level, cognitively speaking, a subject can simply deny that 
its already existing cognitive system cannot integrate the novelty in question, or 
alternatively, can recognize that the alimen is challenging, but integrate it in an 
unstable or inadequate way which cannot withstand criticism, or which shifts 
from moment to moment. This type of cognitive activity is referred to as "alpha 
behavior"; in it, no reflective abstraction occurs, for there is only the ap-
pearance, but not the reality, of cognitive re-construction. One of our findings 
was the frequency with which alpha behavior occurs. That is, we found with 
disheartening regularity instances in which students thought they had understood 
and mastered a particular concept that they in fact had not.

There are many examples of the results of instruction in the literature that can
be described as alpha behavior. One is the well-known "reversal problem" (see, e.g., Clement, Lochhead, & Monk, 1981). In elementary algebra, Rosnick and Clement (1980) report that student misconceptions regarding the use of variables for unspecified numbers are deep seated and strongly resistant to several different teaching methods. They even document situations in which college students will distort "objective reality" (such as relative amounts of oil and vinegar in a bottle) in order to assimilate aliments without revising their existing, inadequate structures. The difficulties persist even with more advanced topics in mathematics, such as limits and continuity. Tall and Vinner (1981) observed cases in which the formal definition of a concept was very different from the internal cognitive structures which students seemed to have available for working with it. By the simple expedient of ignoring the resulting conflict, students repeatedly produced incorrect answers which they considered satisfactory. Again, it was found that traditional teaching methods did not correct these errors.

Alpha behavior is not limited to mathematics education, however. The use of phenomenological primitives among physics students (diSessa, 1985) appears to be another instance of this phenomenon, and we suspect that alpha behavior in general is widespread across the curriculum.

The most recent work of the Geneva School (Inhelder, Sinclair, & Bovet, 1974; Karmiloff-Smith & Inhelder, 1975; Piaget, 1978) has suggested why alpha behavior should be so common. Their work with preschool and school age children has shown that it seems to be the natural tendency in the construction of knowledge to focus first on the most obvious perceptual features of a phenomenon, quite literally, to notice in a naive way "what happens." But merely focussing on the positive aspect—what Piaget calls "affirmations"—is not sufficient for attaining operational knowledge, for such a focus does not entail reversibility. It is not until the child compensates for affirmations by mentally constructing what Piaget calls "negations," or an anticipated negative which would restore the situation to its original state, that an equilibration has taken place. The work of the Geneva School has shown how slowly such an awareness develops (and again, it is at this point that social interaction can be crucial) for understanding frequently lags behind successful performance (Piaget, 1978). We believe we have found analogous phenomena in our work with mathematics concepts, where students are only able to perform explicit calculations in well rehearsed situations (imitative behavior) but are not capable of seeing the calculation as an implementation of an idea which can be applied as well in other, unfamiliar situations (autonomous behavior).

A second type of cognitive behavior occurs when the subject succeeds in assimilating a novel aliments through cognitive construction. This is "beta behavior"; in it, reflective abstraction occurs as (1) the cognitive system as a whole is enriched by being re-organized to accommodate the new structure, and (2) the novel aliments is understood in that it is located within the re-constructed domain of concepts. Beta behavior is the paradigm case of successful learning.
In the third type, called "gamma behavior," the cognitive system of the epistemic subject is already sufficiently rich to integrate the novel aliment without constructing new cognitive structures. Nonetheless, reflective abstraction occurs as the new aliment induces the existing system of concepts to accommodate it through extending itself. The aliment is incorporated through applying to it the existing system of concepts, and new relations are formed between it and the existing system of concepts. The facility with which the epistemic subject seems to understand the novel aliment in cases of gamma behavior might lead one to conclude that it is only abstracting properties from the aliment in a simple, associationist fashion. It actually entails, however, the application of cognitive structures which do not reside in the object, but which have been constructed earlier by the knower.

We will analyze the construction of the concepts of induction and compactness according to this comprehensive model of equilibration.

IV. INDUCTION AND COMPACTNESS

We now apply the foregoing general considerations to two specific logico-mathematical concepts taken from the undergraduate curriculum, mathematical induction and compactness. We will use these examples to elucidate the ideas we have expressed and also show that it is possible, through observation of students in the process of learning these concepts, to arrive at a coherent genetic decomposition of fairly sophisticated concepts. In making these observations, moreover, we will see many examples of the behavior which our theory attempts to explain.

First it is necessary to explain these two concepts. We apologize to the non-mathematician for the technical nature of this discussion, which seems to be unavoidable.

Mathematical induction is a method of proof. It applies to a proposition P which depends on the value of a positive integer n so we may write P(n). For example, P(n) could represent the statement, "the sum of the interior angles of a convex polygon with n(≥3) sides is equal to 2(n−2) right angles." Thus if N signifies the set of positive integers from some initial integer \( n_0 \) on, the function of induction is to prove that this holds for all \( n \) in N, that is, to establish the statement: for all \( n \) in N, P(n). The method proceeds as follows. First P(n) must be established for the initial value, \( n = n_0 \) (in our example, \( n_0 = 3 \) which is the case of a triangle). Then the implication P(n) ⇒ P(n+1) must be proved for an arbitrary \( n \) in N. The principle of mathematical induction states that once these two steps have been taken, it follows that P(n) holds for all \( n \) in N. This method can be used in the same way if N is replaced by any set which can be placed in one-to-one correspondence with the positive integers, such as the positive multiples of 3 or the negative integers.

Compactness is a more sophisticated concept applicable in a wide variety of contexts and of fundamental importance in that branch of mathematics known as
analysis. We shall restrict our context to sets of points in the plane or two-dimensional space. Compactness is a property that such a set may or may not have. A type of disequilibrating aliment to be faced, once the student can explain compactness, is any particular two-dimensional set, for which it must be determined whether or not it is compact.

The definition says, roughly, that such a set, let us call it \( S \), is compact if, whenever it is possible to cover \( S \) with an infinite set of rectangles, then \( S \) is already covered by finitely many of these rectangles. The term "cover" is used here in the ordinary sense. That is, imagine the set \( S \) as consisting of points marked on a thin sheet of material, and then place rectangles made of opaque material on top of \( S \) until no part of \( S \) can be seen. One way of thinking of this definition, then, is that it asserts that no matter what infinite set of rectangles and their positions is prepared in advance, if they would suffice to hide \( S \) then \( S \) is already hidden by finitely many of them.

These two concepts were taught, using traditional classroom methods, to an advanced calculus class of 22 mathematics majors at Clarkson University. After each unit was completed, each student was interviewed individually (for about 15–20 min) in order to determine the nature of the students' understanding of the concept. They were asked to explain the concept in their own words and were questioned about the meaning of various aspects of it. In case the student came to a reasonable explanation, he or she was asked to apply it by proving a particular statement by induction or determining the compactness of given sets. An attempt was made to determine what (if any) mental pictures the student had relative to the concept and how such images were used.

The atmosphere of the interview was generally supportive with free use of prompting. In cases where the student had not yet acquired the concept, every effort was made to find out what he or she did know about the particular concept or various parts of it, and how far the student could go with prompting. Most, but not all, students showed up for both interviews.

The first thing that we observed in looking at the protocols of the interviews on induction and compactness is that in both cases the subjects could be divided into three levels, which represented their understanding of the concept. In what we called Level I, they did not understand the concept. They could explain it only incompletely and could not relate it to specific situations. At Level II, the subject could give a coherent explanation of the concept, including its formal definition, was able to explain various statements about it, and also could discuss small variations of definition. The subject could not, however, apply the concept to specific situations in order to make a proof or test for a property. Finally, at what we called Level III, the subject had a completed structure that could be used to assimilate a theorem (i.e., prove it by induction) or a set (i.e., decide if it is compact). These attempted applications succeeded unless the particular problem drew upon additional concepts which the subject had not yet mastered (cf. Ausubel, Novak, & Hanesian, 1978; Novak, 1977).
Although this trichotomy served as a useful heuristic for a first description of what the protocols showed about how students learn these concepts, a closer analysis indicated that the distinctions could be encompassed in terms of Piaget’s theory of equilibration (including reflective abstraction and the various types of cognitive behaviors).

For instance, there were many forms of alpha behavior displayed by the students at Level I. In some cases, their responses were totally irrelevant or incoherent. Sometimes the student mentioned correct phrases about the concept, but in a fragmentary and uncoordinated way. Their words appeared to be little more than remembered strings and symbols, not related to any relevant cognitive structures. A particularly striking example of alpha behavior would occur when a student would give a logically coherent statement involving phrases that do appear in the concept’s definition, but the statement as a whole was completely incorrect. This can be seen, for instance, in Excerpts 9 and 10 which follow.

Typically, the student would offer a response and look to the interviewer for approval or rejection. If the latter was not forthcoming, the student would blithely offer an alternative, continuing this process in a sort of random search through memory for that of which the authority figure would approve. In such cases, it seemed clear that the student did not possess, or at least did not use, any form of these concepts which had been previously constructed.

But alpha behavior was not the only thing which we observed in Level I students. They also on occasion exhibited some of the antecedent structures required for understanding compactness and induction. From this we inferred they were in the process of constructing these concepts, an example of beta behavior. See, for example, Excerpt 3.

Perhaps the most exciting examples were provided by explicit instances of beta behavior, when a student in Level II was in the process of making the transition to Level III, and successfully made the transition, with prompting, during the course of the interview. For instance, in the following excerpt (which we will call Excerpt 1) from an induction interview, the student has correctly defined the concept, but is at first on the wrong track in attempting to apply it. The problem which the student has been given is to prove by induction that every value in the sequence \( \{x_n; n=1,2,\ldots\} \) is irrational where \( x_1 = \sqrt{2} \) and subsequent values are determined by the relation,

\[ x_{n+1} = \sqrt{1 + x_n}. \]

PR  That’s fine in great generalities and so forth. Let’s see how that gets applied to a specific problem. Like, for instance, there’s one on the board. I’m not saying we’re going to go through all the details of this problem, but just give me a general idea how you would prove that that statement is true.

S1  This is a kind of nasty one.
It's not nearly as nasty as it looks, by the way. Just explain how you would go about it, and then maybe you'll get stopped at certain points, but I want the general form of your approach.

S1: Well, \( x_1 \) is irrational.

PR: That we did in class, in fact. What are you trying to do?

S1: If the number inside there is a perfect square, you've got a rational.

PR: Then it would be rational. That's right.

S1: So you want to avoid that, or what?

PR: The problem is to prove that you don't have a rational number.

S1: Oh. It seems to me that one case would show that that's false.

PR: But you'll never find that case cause this is a true statement.

S1: Is it?

PR: Yes.

Further prompting is conceptualized as disequilibrating by the student (see below, "Okay. Would you ever get to eight?"), and gradually leads him to progressively construct the appropriate concept (see below "Oh no? Oh, I see, right. Starting with an irrational number," and, "You know that \( x_n \) is irrational.")

S1: Well, how about \( x_n \) is 8.

PR: Okay.

S1: \( x_{n+1} \) is the square root of 9.

PR: Okay. Would you ever get to eight?

S1: No.

PR: You do know something about \( x_n \), don't you?

S1: \( x_n \) is n.

PR: No.

S1: Oh no? Oh, I see, right. Starting with an irrational number.

PR: When you look at the equation \( x_{n+1} = \) the square root of \( 1 + x_n \), what are you trying to show?

S1: You want to show that it's irrational.

PR: And what do you know?

S1: You know that \( x_n \) is irrational.

PR: Is irrational?

S1: Yeah.

PR: It's not 8.

S1: The statement for \( x_n \) is \( 1 + x_{n-1} \), mm, that's more confusing than a recursive type of definition.

PR: But after all, if you know that \( x_n \) is irrational, what about \( 1 + x_n \)?

S1: Well, it's also irrational.

PR: And then what happens when you take its square root?
Then that's irrational too.

With any luck, right. You have any idea why?

It could be proven by referring to . . . well you assume that you cannot express $1 + x_n$ as a product of relatively prime numbers.

Rationals.

Rationals. Certainly, the square root of those is not rational.

So let's suppose you did that and your conclusion is that the square root of $1 + x_n$ is irrational. What then?

That's it.

Here the student has successfully constructed the concept.

I agree with you that that's it, but just wrap it up for me.

Since you've shown that $x_1$ is irrational, and . . . well, okay, you assumed that $x_n$ was irrational and then you immediately could be able to prove that the square root of $1 + x_n$ is irrational and therefore $x_{n+1}$ is irrational.

The most important information obtained from the protocols was the many explicit examples of the forms of reflective abstraction, which provide the basis for our derivation of the genetic decomposition of induction and compactness. For instance, the transition made by the student in the transcript above from Level II to Level III also provides an example of a particular reflective abstraction, that of encapsulation.

Consider Chart I (which we will analyze more fully later). There is a prerequisite structure, "method of proof," and the process of induction, which is constructed in attaining Level II. This process must be encapsulated so that it can be an aliment for "method of proof," which then produces an algorithm to be used at Level III when proving a theorem by induction.

Because of the key role played by the protocols, we will present excerpts from them illustrating (1) the three levels, (2) alpha, beta, and gamma behaviors, and (3) examples of reflective abstraction. These excerpts are arranged, for both induction and compactness, in an order that demonstrates the progressive steps which we observed that students undergo in order to construct the concepts. In other words, later excerpts for each concept are taken from students who had indicated either mastery of the lower level concept or the necessary reflective abstraction which we illustrate in detail in the earlier excerpt. In the interest of brevity and focus, however, we chose to not reproduce entire transcripts. It should also be noted that we make a critical assumption about the validity of our genetic decomposition, namely, we assume that the cross-sectional analyses of transcripts across students (i.e., what we did) corresponds to what we would find if the same analyses were done longitudinally with the same students over time.
(i.e., what we did not do). That this assumption is often made in developmental research does not obviate the problem. Obviously, such longitudinal investigations will need to be done to corroborate our findings.

We begin with induction. In the following, Excerpt 2, the student is being asked to explain proof by induction and the question is whether his concept of function is general enough to include a proposition which depends on a positive integer. His initial response refers to a vague "something."

S2 When you're trying to prove something, you would start off first by proving something that's either obvious with one element and you're going on making it a little more general, and then work yourself into a position where its got to be true for whatever you're proving.
The interviewer now tries to suggest something like P(n), but the subject does not pick it up.

PR In induction usually you want to prove certain statements true for every integer n. So in terms of those integers what would you do?
S2 First prove it for one integer by proving it directly.
PR All right, good.
S2 Then, it would depend a lot on what you were trying to prove, I guess.
PR Well, the overall format of proof does it. The overall format of proof is the same every time, but what you actually do to implement that format may change. What I'm more interested in today is, in the extent to which you understand the overall format. I don't care about the details of actually putting it into practice.
S2 Then I guess you would prove it for one, directly, then you might prove it for a set of integers, or a group of them, I guess. Between this and this, make it, and then I guess doing theorems and postulates you have, you try to make it more general.

The entire response is an example of Stage I alpha behavior relative to proof by induction. Contrast this with Excerpt 3 in which the student is definitely thinking about a logical proposition depending on n and having a different truth value, or at least requiring a different proof as n varies.

S3 You're given an equation for n . . . You prove it true for the first n, you put the 2 in there and you prove that's true. Now you want to prove that it's true for each subsequent n . . . and you prove that's true.
PR When you prove it's true for n + 1, what is the thing that you use? What is the most important thing that you use?
S3 The equation.
PR But why did you prove it then for the first one?
S3 You prove it true for the first value.
PR Right. And then what do you do?
S3 Add 1 to that. Prove that's true.
PR Isn't there an implication in all that?
S3 That's true for all n.
PR But isn't there a kind of a next step implication?
S3 (No response).

Here the subject's concept of function has been generalized to include a proposition valued function of positive integers. Another difficulty has arisen, however, in that the subject's concept of logical necessity has not been encap-
lated, which would allow an implication $P \Rightarrow Q$ to be a proposition that can be used in a proposition-valued function. The following exchanges tend to confirm the absence of any awareness of implication as a cognitive object.

PR  How would you prove it true for 4? You've got it proved for 3, right. Now how would you prove its true for 4?

S3  I just always thought that once you proved it true for one number, and 2, 3, 4, 5, and proved it true for $n + 1$ that it was true for all the numbers.

At another moment:

S3  Just because you prove it true for the first statement, $n = 2$, just means it's true for 2. It doesn't mean for any other positive integer.

PR  That's right. If you prove its true for 2, then it means that it's true for 2.

S3  Okay. So what am I missing?

PR  What you're missing here is that the step here is that what you prove is that if it's true for $n$, then it's true for $n + 1$.

Finally,

S3  I think I knew that if you prove it true for $n$, then it's true for $n + 1$, and $n + 1$ would bring it to 3, and now if it's true for 3, and it's true for $n + 1$, that would bring you to the 4, if it's true for 4, and it would be true for the $n + 1$, which brings you to 5. I couldn't say that though, I suppose.

This subject is still in Level I relative to induction, but the response is certainly not alpha behavior because disequilibration is present, and an awareness of an implication is emerging. We can infer that the response is an example of beta behavior.

In the next example, Excerpt 4, the subject has the concept of proposition-valued function of integers.

S4  First of all you prove that something is true for 1, for $n = 1$, and then it's true for $n$, then if it's true for $n + 1$, then it's true for all $n$ because that's all the natural numbers.

And, in the following exchange, the use of the phrase "same thing" to refer to two separate instances of the cognitive action, "if $P$ then $Q$" suggests that $P \Rightarrow Q$ is, indeed, a cognitive object for this subject.
S4 It's true for 1, and it's true for n+1, and 9+1 is 10, and 9 is a natural number, and it's true for 9, so it's true for 10.

PR How do you know it's true for 9?

S4 I could say the same thing with 8.

PR Say it.

S4 Okay. It's true for 1 and it's true for 8, so it's true for 8+1, which is n+1, so it's true for 9.

However, this subject is not able to go one step farther and coordinate proposition-valued function of integers with implication as proposition to obtain an implication-valued function of integers, as indicated by the inability to speak of proving, for a given n, the implication, P(n) \( \Rightarrow \) P(n+1):

PR Okay, let's try to get at it. First you prove it's true for 1. Now what's the next thing that you do after you prove it's true for 1?

S4 You prove it's true for any n.

PR So why did you bother proving that it's true for 1? Since that's what you want to do anyhow.

S4 Cause if you want to show it's true for n+1, you have to know it's true for n first.

PR What I'm saying is, if you proved it's true for any n, then of course it's true for n+1, because that's just another n.

S4 I suppose.

PR There's something you're missing there, right?

S4 I guess so.

PR Do you actually prove it's true for n?

S4 Well, you show it's true, for any n. If you prove it's true for n+1, then you're showing it for any n.

This difficulty is not present in the next example, Excerpt 5. The subject first gives a correct statement of induction. The response to the question about n = a million indicates that the subject is evaluating an implication-valued function for n=1,2, ... up to a million. Then the explicit mention of a pattern suggests an awareness of the function as a cognitive object.

S5 You have to show that it's valid for one, then you show that if you have a special n in it, that that would work for that expression and you try to go for the next n, so that's an element of the expression, and then the next one. In other words you try to show it for all n+1 that's in that expression.

PR Well, that's right and of course convinces me and so forth, but suppose you were talking to people who didn't know very much about
this thing, and suppose they wanted you to explain that to them, that that's how you prove it for all n, and they said, "But I don't under-
stand." How does what you did guarantee that it's true, for instance, for \( n = a \) million? How do we know that?

S5 If you want to show it for a million, you'd say, if it is equal for 1, then you go on and say, it's equal for 2, and then you go on and say it's equal for 3, and so forth.

PR Right.

S5 You're starting to develop a pattern, once you get that pattern going, then you can kind of see ahead, and even if you go further than a million, you can see that \( n \) would still be an element of that expression.

Thus we have an example of a subject in Level II with respect to induction. He is not in Level III, however. When faced with a particular theorem to prove by induction, he does not apply the method that he has described, because he has not yet encapsulated the description above as a specific method within the structure of "methods of proof." The following Excerpt 6 is an example of another subject in the same situation. She has explained induction adequately but her attempts to apply it to a specific situation results in what can only be described as alpha behavior.

PR When you're making a proof by induction, what's the first thing you do?

S6 You define your function and you determine if the first element of the set belongs in that function.

PR Okay. And what's the next thing you do?

S6 Then you take an arbitrary element \( k \) and you assume that it's true for \( k \). And then you replace \( k \) with the element which will be following it, that is \( k + 1 \) and you plug that in, and you determine now that it's true for that, and since \( k \) is an arbitrary element, if it was true for \( k \), and it's true for \( k + 1 \), it would have to be true for everything in the set.

PR Now this is all very general. Let's see if we can't try to tie it down to a specific problem.

S6 Okay

PR There's a problem on the board. I what I'd like you to do is to explain to me, you don't necessarily have to solve the whole problem, or you might, explain how you would go about trying to solve that problem, using proof by induction.

S6 First thing, you've already got \( n + 1 \) up there, I don't really have to break it down, get it to the \( x_n \ldots \)

---

1This problem is stated on page 67.
PR  Sorry, you’ve already got it to the . . .
S6  Right here. This is already to 2 + 1.
PR  Right.
S6  So you probably have to manipulate this algebraically so this is $x_n =$
    the sum of n.
PR  Back off a little bit. We’ll get to that in a minute. How about the first
    one. What’s the first thing you do when you make a proof by
    induction?
S6  You prove it’s true for this.
PR  And there you’ve got $x_1$ is equal to what?
S6  n + 2.
PR  Now what about that . . . What is it that you’re trying to prove about
    $x_1$?
S6  That it works.
PR  That it works. What does “it works” mean?
S6  That if you plug it in you get out a value which is defined.
PR  What’s the problem?
S6  I’m not even sure.
PR  Read it. It’s a set.
S6  So that $x_n$ is irrational . . .
PR  That’s right. Now what is it that you desire to show for $x_1$?
S6  That it’s irrational.
PR  That it’s irrational, right. And you’ve got written there what $x_1$ is
    equal to, right?
S6  Cause it’s irrational.
PR  Right, cause we did that in class. Now what?
S6  You’ve got this. So if you want to prove that if you set $x_n$ equal to $x_k$,
    that $x_{k+1}$, if you use this formula, you plug it in and get an irrational.
PR  Right, and what is it that you’re assuming?
S6  That $x_k$ is irrational.
PR  So you’ve got that over there and you’ve shown that $x_k$ is irrational.
    You want to prove that $x_{k+1}$ is irrational. You see how I’m doing that
    up there?
S6  I’m not sure. I always have trouble with these.

Although they were not able to apply it to specific problems, subjects S5 and
S6 were able to explain the induction process adequately. A key ingredient seems
to be the ability to coordinate an implication valued function with *modus ponens*
which consists of the assertion of Q, given both P and the implication $P \Rightarrow Q$.
The use of *modus ponens* is the result of encapsulating and becoming aware of
the structure of logical necessity so that it can be applied repeatedly. Thus if one
has a function which assigns to each integer n the implication $P(n) \Rightarrow P(n+1)$
and one knows that $P(1)$ is true, then by repeated evaluation of this function
and application of *modus ponens*, one obtains the validity of \( P(1) \), \( P(2) \), \( P(3) \), \ldots This is particularly clear in the following excerpt, Excerpt 7.

**PR** You proved \( P(1) \). And then you also proved that if you knew \( P(n) \) and then \( P(n+1) \) you could do the rest of the problem.

**S7** So given \( P(1) \) we plug that into our little equation and show that since \( P(1) \) is true, \( P(2) \) is true. Then since \( P(2) \) is true, \( P(3) \) is true, and so forth and so on up to if \( P(9) \) is true then \( P(10) \) is true. I could demonstrate that for any \( n \), although it would take me a while for up to a million.

Finally, in Excerpt 8 we have a subject who gives a correct explanation of induction and then, using this explanation as an aliment to the structure, "method of proof," gives a clear and succinct recipe for making a particular proof by induction. This subject is at Level III with respect to induction and his response to the specific problem is an example of gamma behavior.

**S8** It means that you prove it for one particular \( n \), normally you start with 1 and then prove that for any \( n \) you can prove it for the next one. And since you already proved the first one.

**PR** You prove it for the next one.

**S8** Provided that the first one is true.

**PR** The first one?

**S8** The \( n \) is true.

**PR** The \( n \) is true.

**S8** So if you can prove that the next one is true, when one is true, then from the first one you can prove the second, and the third, and the fourth, and so on.

**PR** That's very good. Now let's see if you can apply that to a particular induction problem. Here's one on the board. Now I don't necessarily want you to solve the whole problem. I just want to know how you're going to go about it.

**S8** Okay. We know that \( x_1 \) is irrational, because we've proven that \( \sqrt{2} \) is irrational already.

**PR** Right.

**S8** And then I would say, assume \( x_n \) is irrational.

**PR** Okay.

**S8** And then prove that \( \sqrt{1 + x_n} \) is irrational, because \( \sqrt{1} + \text{an irrational} \) is irrational.

**PR** Right.

Now we will derive the genetic decomposition of the concept of induction, that is, we will describe this structure and explain how reflective abstraction is used to construct it. Our derivation is based on the protocols. We begin by
summarizing the examples of reflective abstraction which appeared in Excerpts 3–8.

1. In Excerpts 3 and 4, the subject has generalized the concept of function to include a proposition-valued function of the integers.
2. In Excerpt 4, the cognitive operation "if P then Q" has been encapsulated to form an object which can be an example of a proposition.
3. In Excerpt 5 the object P → Q is now an aliment to the structure "proposition-valued function" which is thereby generalized to assimilate "implication-valued function."
4. In Excerpts 5 and 7, an implication-valued function (in particular, its evaluation) has been coordinated with the structure, *modus ponens*. The result is the detailed statement of how one uses the idea of induction to know that the statement is true for n equal to 10 or for n equal to a million.
5. In Excerpt 8 the process of induction has been encapsulated into a cognitive object that can serve as an aliment to the structure, "method of proof" which is thereby generalized to include "proof by induction."

We believe this series of reflective abstractions could be used to construct the concept of mathematical induction. The interrelation between structures which we found is shown in Chart 1. However, we should note at this point that the decomposition illustrated here is not necessarily the only way to decompose this concept; it is simply the decomposition which we empirically found to be used by our students. Neither is it the way in which trained mathematicians would necessarily decompose the concept.

It seems that three structures must be present to begin with: logical necessity, which is fairly elementary and probably possessed by most college students (see Moshman & Timmons, 1982, for an interesting discussion of how it might be constructed); function, which is usually present, but only as a process relative to numbers and algebraic formulae; and method of proof, that is, the notion that a particular algorithm (e.g., proof by contradiction) can be applied to various propositions in order to prove them. These three structures are shown on the top line of Chart 1.

The first step is to generalize the structure of function to obtain functions of the form n → P(n) (see 1. above). Next, logical necessity is encapsulated to obtain the cognitive object P → Q (see 2.) which becomes an aliment to the generalized structure of function to obtain functions of the form n → (P(n) → P(n+1)) (see 3.).

At the same time, logical necessity is encapsulated to form *modus ponens*. As indicated in Chart 1, all subjects up to this point are still considered to be at Level I relative to induction because they cannot yet explain induction. The next step, coordinating implication-valued function with *modus ponens* (see 4.) brings the subject to the explanation of induction and Level II. Finally, by encapsulating induction from a cognitive process into an object that can serve as an aliment to
method of proof (see 5.), the subject moves to Level III and is able to apply induction in solving various problems.

Next we turn to compactness and apply the same analysis as we made for induction. In these interviews, the students were asked to give the definition of compact and not compact, and to apply these definitions to two examples. The examples consisted of two sets $A$, $B$ in the plane. $A$ is a sequence of points converging to the origin but not including the origin. It is not compact. $B$ consists of the origin together with a sequence of points converging to the origin. It is compact.

We begin with excerpts from the protocols. In the first, Excerpt 9, the response is connected with the definition of compactness only in that some of the same terms are used, but the relations are all wrong and the total statement is meaningless. This is an example of alpha behavior in Level I relative to compactness. The trouble seems to be that the subject is unable to think about elementary manipulations of (finite or infinite) collections of sets.

**PR** Now can you tell me what it means for a set, $S$, in the plane, to be compact.

**S9** That means that there will be a finite number of open sets that cover $S$. It's a subcovering that will contain the whole set $S$.

**PR** What do you mean by subcovering?

**S9** An infinite number of open sets is a subcovering. That will cover the whole set $S$.

**PR** I guess you're not making clear to me yet what is your idea of subcovering. What's going on? What's the situation?

**S9** A finite number of sets, let's see, of open sets, that are contained in the whole set of $S$. Those open sets will contain, will cover that whole set in the plane. It will cover the whole thing, a subset of set $S$ in the open set will cover that whole set.

In Excerpt 10 we see a subject who does possess a structure for manipulation of collections of sets but is unable to coordinate it with universal quantification of propositions. Note that even when the interviewer offers the correct statement, and the subject accepts it ("ok"), the next response indicates very little understanding and no possibility of negating the statement given by the interviewer. This is another example of alpha behavior.

**PR** So let's start by seeing if you can tell me what it means for a set $S$ in the plane, in $R^2$, to be compact.

**S10** A set has to be compact in the plane, if it can be covered by a collection of open sets, so that the union of these finitely many open sets contains $S$.

**PR** You certainly are talking about the notions that are involved here. A collection of open sets, and finitely many of them, and so forth. But
you said that what you need to be able to do is to cover them once so that a collection of open sets, finitely many of them cover it. Is that what you really mean?

S10  Could you repeat that?

PR  What you said is that it's compact if you can cover it with a collection of open sets so that finitely many of them cover it. Is that what you meant to say? Because you can always do that. Just by covering it by a single open set. If you cover it by one open set . . .

S10  There has to be more than one.

PR  So you cover it by one open set and then you also throw in any number of other open sets, and that one you covered it with still covers it. So your statement is always satisfied, isn't it?

S10  It has to be unique open sets. In order to get at what you're saying, you have one large open set covering the whole set $S$, and you have open sets inside that set, they have to be distinct, that these open sets, none of the two are the same.

PR  That's not what I'm trying to get at. What I'm trying to get at in that is that that's one way of doing it, and that way will work, but maybe some other way won't. What you have to say is that no matter how you cover it with open sets, you could have done it with finitely many of the ones you used.

S10  Ok.

PR  So given that, what would it mean for a set to be not compact?

S10  When you cannot, in any case, cover it with a finitely many of open sets.

In Excerpt 11 we have a subject whose structure of set manipulation has been coordinated with universal quantification of propositions to permit a first understanding of the definition of compactness. We consider this subject to be still in Level I relative to the concept of compactness, however, because he cannot explain what it means for a set to fail to be compact. He obviously has the structure of negation but it has not been sufficiently generalized to be applicable to universal quantification which could then be coordinated with set manipulation, or to the total definition of compactness. Either form of reflective abstraction would lead to the definition of "not compact" but this subject is not capable of either, even after considerable prompting. Note also that although we had several subjects who could give the definition of compactness but not the definition of non-compactness, we had no examples of the reverse. This is consistent with Piaget's principle that affirmations are constructed before negations (Piaget, 1978).

PR  Can you tell me what would be the definition of what it means for a set to be compact.

S11  A subset $S$ of $R^2$ in the plane, is considered to be compact if for each collection of open sets which covers the subset $S$, there exists a finite
subcollection of open sets from the . . . which will again cover the subset S. If you want I can write it down more clearly.

PR It's okay. So then what would it mean for a set to be not compact?
S11 A set would not be compact if for each collection of open sets that, for each infinite collection of open sets that cover it, there is no finite subcollection which will also cover it, so that only an infinite collection of open sets can cover it.

PR Wait. In your definition of compact you said that every time you had a covering by a collection of open sets you could take a finite subcollection and cover it. So you must be able to do this with every cover. So if you say that it's not compact . . .

S11 I see what you mean. OK. You might be able to do it for some coverings but . . .

PR Now let's say this again. Explain to me what does it mean for a set to be not compact.
S11 A set, a subset of the plane R2 is considered not to be compact if we cannot for each subcollection, for each infinite collection of open sets that cover . . . you cannot find . . .

PR No, that's still not right. You said for compact it means what?
S11 For every subcollection, for every collection of infinite sets you can find a subcollection that covers the set.

PR Not compact means?
S11 Not compact means even if you can find some collections, some infinite collections of open sets that covers S, the subset from which the finite subcollection will also cover it, we cannot do this for any subcollection of infinite sets.

In Excerpt 12, the subject has progressed to the point of being able to negate the definition of compactness (with a little prompting) so his concept of compactness is the coordination of the affirmation and the negation. This brings him to Level II relative to compactness. He is not fully at Level III, however, because his attempts to apply the concept to specific sets are halting, uncertain and a mixture of irrelevancies together with correct statements. He relies heavily on prompting and looks to the interviewer for validation. He does, in the end, get both examples correct and so we can consider this an instance of beta behavior in the transition from Level II to Level III.

PR Can you tell me what it means for a set S in R2 to be compact?
S12 In R2. Well if you take . . . A set is compact if and only if for any collection of infinite open sets you can find a subset of that collection, and if the open sets covers your set, you could find a subset, a finite subset that also would cover that set S.

PR Good. So then what would it mean for a set to be not compact?
For it to be not compact, it would be say, for instance, take any infinite collection such that their union also contains S, you couldn’t find a finite set that also contains S.

You’d have to come across that situation how often? In order to find if a set is not compact.

Just find one example. That’s all you have to do for that.

I think you said it perfectly correctly. Let’s see if you can relate it to concrete examples. Here’s two sets in a plane, described on that card. The question about them is whether they are compact or not.

OK. (pause) Yes, I’d say they’re both compact. Because of the fact that if you make a ball out here, this is going to be more or less finite. You come to a point, say be infinite within a close space, to y = 0.

What does this have to do with your definition, which was that if you had an open covering you could find finitely many of them that covered. So you have to show that that’s true for that set. I don’t want to mess with a formal proof, but just an indication of how it would go.

Generally, I think finding a ball that is any interval you take, you look at an interior point, any ball you take whose center point would be at say 0, 1, in this case, or 0, −1.

Suppose that ball wasn’t there. Why does that ball have to be there? What about the second one?

(pause) It’s the same type of thing except both are approaching. (pause) Let me try to say this again. In this case here you have, set A, you can find a case where there’s an infinite number of open balls such that they would exactly cover one point each.

Show me by picture how you do that?

Here. (Shows picture of an infinite cover that cannot be reduced to a finite subcover).

What happens as you get closer and closer to the y axis?

They get smaller and smaller. But I need every single one of those. If I take one away then it won’t work.

So what does that say about the set A?

It’s not compact.

Now what about set B.

Set B is a little different.

What’s different about set B.

Because there are, you have zero in the set, and then as you get closer and closer you’re eventually going to have a covering that’s going to contain an infinite number.

How’s that going to be? What’s the difference between set A and set B? (pause)

Is there anything in set B that set A doesn’t have?
Yes, the zero. Yes, that’s true, so if you take zero, put any kind of covering around it, it will contain an infinite number of sets. Therefore, if you take any cover, any open ball, covering the origin, that open ball would contain an infinite number of points.

And what would be left out of your ball?
The finite. Therefore it would have to be compact.

Why?
Because no matter what infinite set of balls I took to cover ...

Finally we have Excerpt 13 in which the subject is in Level III and can apply a concept of compactness to two specific sets to determine if they satisfy the definition.

Can you give me a definition of what it means for a set to be compact?

If you have a collection of open sets that cover a compact set, then a finite subset of that collection would already cover the compact set.

OK, good. A cover means what?

A cover means all the elements of the compact set is in one of the open sets of the collection.

Good. Can you explain what would be the situation for a set which is not compact?

Well, it means you can find a sequence or just a collection of open sets where, an infinite collection where every single one of the open sets is required to, all of them are required to cover the compact set ...

That’s very close, but not quite. You don’t really need all of them.

But no finite subset of that would be all that’s required.

Good. Here are two sets which I would like you to take a look at, and my question is are those sets compact?

That doesn’t look compact.

Good. That happens to be the correct answer. What about B?

B I would say is compact.

OK. That’s also correct. Now can you give any reasons in either case?

I would say for the first one, in these two sets, the dots approach this point and this point, it should be $[0,1]$ and $[0, -1]$ so it can have open sets that cover, well the first ones cover the first point at $[1, 1]$ and the next one just covers one more, and then just covers one more, and you can set up a sequence of open sets that follow just one more set and you require the next one to attempt to cover more sets, so you have an infinite collection of the open sets that will cover the whole set but no finite subsets that will cover the whole set.
PR  Because what would happen if you took a finite set? Subset?
S13  Well, you can find a point that would be left out.
PR  Where would you find it?
S13  Right outside the one you left off.
PR  Right. Good. That's perfect. What about the second set?
S13  The second set I think the points would be like that. (Shows a drawing).
PR  That's exactly right.
S13  Both of them approach this point and that point is included in the set. Any collection of sets that cover the whole set must have an open set that covers \([0,0]\). And inside there, there is an infinite number of points that belong to the set, or actually outside it there is a finite number, so those could be covered by the finite number.
PR  Very good.

As we did with induction, we now summarize the examples of reflective abstraction that appeared in our excerpts from the protocols on compactness and then, based on these examples, show the genetic decomposition of compactness.

1. In Excerpt 10, the structure of set manipulation possessed by most college mathematics students in a fairly elementary form has been extended to include operations like infinite covers and extraction of finite subcovers.
2. In Excerpt 11, this extended structure of set manipulation is coordinated with universal quantification to obtain a statement of the definition of compactness.
3. In Excerpt 12, the negation of the definition of compactness has been obtained. However the specific forms of reflective abstraction that are required for this construction were not evident in any of the interviews. Nevertheless, at this point the definition and its negation are coordinated to obtain the concept of compactness.
4. Finally, in Excerpt 13 the concept of compactness has been encapsulated into a cognitive object that can serve as an aliment to the structure, "test definitions" which is thereby generalized to include compactness.

Thus it is not difficult to see how reflective abstraction could be used to construct the concept of compactness, and a possible decomposition of the concept is offered in Chart 2. There is one gap, however, in this decomposition, for we found no subjects in the group examined who were at intermediate stages between understanding the definition of compactness and understanding both the definition and its negation. Hence, although we can think of at least two ways in which this can be done with reflective abstraction, we do not have any genetic data on this point and more experiments will have to be performed in order to fill the gap. In the meantime, in Chart 2 we use broken lines to show two ways in
which the definition of "not compact" could be constructed. Interrelationships for which we do have genetic data are shown in Chart 2 with solid lines.

There are four structures which must be present in order to begin the construction of compactness: (1) manipulation of sets such as unions, intersections, and so forth; (2) universal quantification of propositions which is the logical form, for all x in some set P (a proposition possibly depending on x) is true; (3) negation which is the structure that assimilates a logical proposition as an aliment and transforms it into its negation (it is in fact the N that appears in Piaget's INRC group but at this elementary level it need only be able to assimilate aliments such as P \(\Rightarrow\) Q); and (4) test definitions, which is the notion that given a definition and an aliment one can check the statements in the definition to determine if that aliment satisfies the definition (presumably an outgrowth of the early classification structure).
The first step is to generalize the structure of manipulation of sets (see 1. above) and coordinate it with universal quantification (see 2.) to obtain the definition of compactness. We suggest two alternatives for constructing the negation of compactness. One is to generalize negation so that it can accept universal quantification, which would have to be encapsulated, as an aliment thereby obtaining existential quantification which is the logical form: there exists \( x \) in a certain set such that a proposition \( Q \) (possibly depending on \( x \)) is true. This could then be coordinated with manipulation of sets to obtain the definition of not compact. Another possibility is to generalize the structure of negation so that it is powerful enough to directly negate the logical statement in the definition of compactness. There could be other alternatives. In any case the next step is to coordinate the definition and its negation to obtain the concept of compactness (see 3.) which brings the subject to Level II with respect to compactness. Finally, by encapsulating compactness into a cognitive object that can serve as an aliment to “test definitions” (see 4.) the subject moves to Level III and is able to determine whether various sets are compact.

V. IMPLICATIONS FOR PEDAGOGY

The preceding epistemological analysis has pedagogical implications. It suggests to us as teachers that our pedagogy must begin with the ways in which the epistemic subject constructs knowledge. The best and most dynamic instruction will fail if it does not take into account the cognitive structures—both those possessed and those that must be acquired—of the knower, as well as the process (reflective abstraction) by which these constructions take place.

We have been particularly concerned in this paper with logico-mathematical concepts that typically appear (in North America) in post-secondary undergraduate mathematics courses. These concepts are significantly more sophisticated than those which are studied at earlier ages and there is no evidence that students tend to acquire them spontaneously, nor is there reason to believe that students discover them on their own. The concept of compactness, for instance, was invented less than 100 years ago and went through a period of development for several decades before it reached the form in which mathematicians understand it today. It is not reasonable to expect it to emerge spontaneously in the minds of students. Calculus provides an even sharper example. After 2000 years of scientific progress, it took the genius of Isaac Newton and Gottfried Liebniz (simultaneously, but independently) to discover the inverse relation between differentiation and integration. Not many of our students are likely to rediscover this idea without considerable orchestration of their learning activities.

Even with the full orchestration of a college education, however, it remains the case that, with the exception of those with special talents for mathematics, students do not learn the concepts in the undergraduate mathematics curriculum.
Comparing our epistemology with the usual methods of instruction in mathematics, one can begin to understand why.

When a professor stands in front of a class, trying to explain a certain concept, he or she possesses a particular structure which can assimilate all aliments appropriate to this concept. The words spoken, the pictures on the blackboard (or overhead projector), and the examples calculated all represent descriptions, valid for the lecturer, of a structure which already exists for her or him. The students, unfortunately, do not possess this structure and so can only try to assimilate these aliments with whatever structures that they do possess.

As we see in Charts 1 and 2 for induction and compactness, there is a very complex, highly non-linear arrangement of antecedent structures which must be present, or if not present must first be constructed before the student can take the last step and perform the necessary reflective abstraction to construct the concept. Consider the node, "explaining induction," in Chart 1. First the learner has to have the structure of implication-valued function as well as modus ponens. Next, the latter must be encapsulated and then coordinated with the former. Students who don’t have these prerequisite structures (and we have seen that such students exist) will not be able to assimilate any of the usual classroom explanations of induction. Words, pictures, worked examples are not powerful enough to induce the construction of these earlier structures and thus may be received as nothing other than words, pictures and worked examples. The student decides, with support from the overall educational system and in the absence of any structure that would permit autonomous activity, that learning consists of imitating the behavior of the teacher. In this way is alpha behavior promoted; the student believes that he or she understands when he or she does not. Thus, as any college teacher of mathematics knows, the best result that can be hoped for in a traditional class is that the students repeat the words they have heard.* draw the pictures they have seen, and solve the same problems on which they have laboriously practiced. There is little evidence of long-term retention of this material and no indication of the existence of structures that can perform any kind of assimilation that is not a repetition, or a minor variation of one.

Given this explanation of the failure of traditional methods of post-secondary education in mathematics concepts, the question arises whether Piaget’s genetic epistemology as we have described it can form the basis for developing a pedagogical approach that can lead to better results. We believe that this is the case and that the rudiments of such an approach which we now describe can be developed into effective methodologies for teaching various concepts including many for which traditional approaches have not been successful. Indeed, the results of preliminary experiments are encouraging. Some of these are referred to below in the general description of our pedagogical approach.

---

*The key word here is “traditional.” There are a few non-traditional classes that aim for, and achieve, deeper cognitive growth. — Editor
In brief, we propose three main steps in teaching a logico-mathematical concept: (1) determining the genetic decomposition of the concept; (2) inducing the students to perform the required reflective abstractions; and (3) explanation and practice. Let us explain these steps in some detail.

Determining the Genetic Decomposition
Two questions arise here. The first is: What is the relation between the genetic decomposition of a logical-mathematical structure, that is, the arrangement of antecedent structures which the learner must construct and organize during the learning process, and the corresponding arrangement which the knower possesses, later, as a completed concept? The second question is, how is the arrangement and the particular structures which it involves, to be discovered?

Regarding the first question, Piaget has given examples in which the genetic decomposition is very different from the decomposition of the knower. One striking example is the concept of number. Piaget (1941) has shown that the genetic decomposition in this case involves a coordination of seriation and inclusion whereas the knower’s decomposition involves things like one–one correspondence, successor functions, and Peano’s postulates (Beth & Piaget, 1966). Our investigations of induction and compactness provide additional examples of this kind of distinction. We do not think that many mathematicians would offer anything like Charts 1 or 2 as a description of how they think about these concepts although after looking at the charts, most would agree that these are logically valid decompositions.

One can also offer more theoretical explanations for the distinction. After a reflective abstraction has resulted in a new structure, there is no reason for a knower to consciously connect an antecedent structure (which is still present) with a completed structure. For example, once the epistemic subject has constructed the concept of compactness and can use it to test examples, there is no reason to remember that an earlier step was to coordinate universal quantification with manipulation of sets to obtain the definition. An encapsulated process such as *modus ponens* is thought of and used as an indivisible cognitive object. The epistemic subject may not always be aware of its antecedence as logical necessity. There are many such examples which contribute to the distinction between the genetic (learner’s) decomposition and the knower’s post hoc decomposition.

Obviously, our main concern as teachers must be with the genetic decomposition. We feel that it is a mistake for a teaching system (whether it be human or computer oriented) to base its design on the knower’s decomposition, that is, to “ask an expert.” Indeed, it is clear from the research reported in this paper that our position on the second question is that the way to discover the genetic decomposition of a concept is to observe learners as they go through the process of learning the concept.

There are two caveats of which we must be aware. First, we have already indicated above that it is only an assumption (which is generally accepted by
developmental researchers) that the cross-sectional analysis which we made observing several students at the same time is an accurate reflection of what happens with one student over a period of time. The second concern is that genetic decomposition may be different for different students; different cognitive styles, for instance, could lead to different ways of constructing a concept. Our hope is that variations in the style of decomposition will not be too numerous and thus will not overly complicate the use of genetic decompositions as a basis for teaching.

Inducing Reflective Abstraction
Ideally, as a preliminary to inducing the students to perform the reflective abstractions required to construct a particular concept, the teacher should determine, for each student, her or his exact position relative to the genetic decomposition that was derived according to the first step. Although this presents no intrinsic difficulties and could be done with observations and analyses similar to, but simpler than, those performed in the first step, it could be tedious if the number of students is large. It would be even more impractical if there are different genetic decompositions involved. It is not clear that such an exhaustive effort is necessary, however. It may be that (as is apparently the case for induction) the total number of nodes on the decomposition chart and connections between them is small. If the teacher (with the help of pedagogical research) designs effective learning experiences that help to move a student from one position on the chart to another, then it is not unreasonable to think of having most students go through most of these experiences. Each experience will be necessary for some and, for others, will serve to reinforce existing structures. The teacher can introduce as much individualization as is practical in a given situation. The main point is to proceed with the design and implementation of specific activities whose goal is to move a student along the genetic decomposition chart. There are, however, a number of further difficulties which present themselves.

A surprisingly serious obstacle to overcome—perhaps because it seems trivial, and is therefore easily overlooked—is to defeat the student’s tendency to substitute alpha behavior for real cognitive effort. That is, a student may persist in applying an incorrect understanding of a concept, even after a teacher has pointed out the error and tried to re-teach the concept. In deference to existing social relations (i.e., teachers assign grades), students will accept the criticism and instruction overtly, but revert to incorrect usage as soon as possible, somewhat bemused by the teacher’s apparent pedantry. We have mentioned several examples of such behaviors in our initial discussion of alpha behavior.

How can we understand this tendency for alpha behavior? We begin by suggesting that the student, in this situation, possesses one of two possible cognitive frameworks. Either he or she has no overall sense of the target concept,
but only has disorganized fragments of it (what diSessa calls "knowledge in pieces"), or there is an integrated concept, but it is incorrect. Bringing Piaget's epistemology to bear on this point suggests a pedagogic perspective. If an epistemic subject (here, a student) is motivated to attempt reflective abstraction only when it is disequilibrated, it is also the case that it can be disequilibrated only if it has a prior conception, however tentative and partial, from which to work. (This primitive model may well be another cognitive structure which has to be constructed.) We can call this primitive model a "theory" (Karmiloff-Smith & Inhelder, 1975), or an "anticipatory schema" (Kuhn, 1981), or an "intuitive epistemology" (diSessa, 1985), but the point is that only if the subject realizes that it is exposing a set of expectancies to possible disconfirmation (à la Popper's philosophy of science) can it be disequilibrated. If it has no organized expectancy, or if it does not apply that expectancy consciously, it will not be disequilibrated by disconfirmation. Instead it will simply continue to function in a cognitive fog, in which concept, definition, application, and feedback bear amorphous relations to each other. Thus effective pedagogy must not only "start" with what the student knows and affirm active cognitive construction, as important as these are, but it must also make the student consciously self-aware of its presuppositions, its implicit epistemology, in order that that which is disconfirming is also disequilibrating. Otherwise, alpha behavior of this type will continue.

Another general difficulty lies in the fact that much of this mental activity involves the manipulation of imagined objects and processes which are not things which have been experienced directly and, in fact, may only exist as logical possibilities and not as part of the physical world. The ability to deal with such situations is included in the stage of cognitive development which Piaget has called "formal operations." According to many studies only a small percentage of beginning undergraduates have achieved this stage (see Chiappetta [1976] for a description of several results) and so for many students, something will have to be done to help them deal with abstract processes as concrete objects.

Recent work has suggested the possibility that some of these difficulties can be overcome and appropriate reflective abstractions induced through computer experiences. Before going into a little bit of detail on this exciting development which we believe has potential for significant improvement in learning abstract concepts, we would like to emphasize that this use of computer experiences is not the usual form of computer aided instruction but is in some ways closer in spirit to the use of LOGO for more elementary concepts (Papert, 1980).

We have found that implementing an abstract process as a computer program can help students see that process as a cognitive object and learn to manipulate the process mentally. When we did this with the concept of functions and their composition, the subjects having computer experiences performed much better than a control group (Ayres, Davis, Dubinsky, & Lewin, 1986).
It also seems possible to use familiar programming constructs to induce specific reflective abstractions. In a prototype classroom experiment, the programming language SETL (Dewar, Schonberg, Schwartz, & Dubinsky, in press) was used to induce exactly the reflective abstractions which appear in our genetic decomposition of induction (see Chart 1). For example, working with arrays of boolean variables led students to think about proposition-valued functions of integers; a small program that checked for an implication between two successive entries in the array helped them generalize this structure to include implication-valued functions of integers; and simple if-then clauses helped to make students aware of modus ponens so that they could encapsulate it. Writing certain more complicated programs helped them to think about quantifications and their negations. This was not used for learning compactness however, but for studying the concept of limit. Reports on this work and follow-up studies will appear elsewhere.

It also seems possible that working with computers can help eliminate some forms of alpha behavior and make the subject more conscious of the need to re-equilibrate. Indeed, there is nothing more disequilibrating than a computer program that does the wrong thing.

As difficult as all of this is, inducing cognitive activity forms a major part of the role of the teacher and the fact that this is generally missing in our colleges helps explain why undergraduates do not learn mathematics concepts. The exceptions are those with special talent for mathematics. These students, once disequilibrated, do perform reflective abstractions on their own, spontaneously. But most students need help with this step and where methodology for providing this assistance is lacking, it is the role of educational research to propose and evaluate new methods. We have tried to show in this section how computer experiences can be helpful in this regard.

Explanation and Practice
The third step is similar in form to the traditional teaching method in which a description of the concept, either written in a text or presented verbally in class, is integrated with examples, applications, problems, and so forth. The main difference is that the explanation can be directly and consciously (for both teacher and student) related to existing structures which have been developed to the point where the student is ready to reflectively abstract them to form one new structure. Discussion should guide them through this procedure and here, discovery learning could prove quite effective.

It should be noted that no matter how the structure has been initially acquired it may not be firmly established. Practice with using this new structure to assimilate disequilibrating aliment over a period of time will probably be necessary before it becomes stabilized in the student's cognitive repertoire.

In conclusion, we believe that the process of cognitive construction through which students attempt to learn post-secondary mathematics needs to be exam-
ined much more closely than has traditionally been the case, and that Piaget's epistemology offers a rich guide for that examination. It is how students conceptualize material that determines their degree of success in mastering it. Consequently, it behooves us, as teachers, to incorporate into our pedagogy an approach that facilitates the construction of the concept. In this paper, we have tried to detail such an approach for the concepts of proof by induction and the compactness of sets by specifying the cognitive acts students actually perform during concept construction. It is our belief that similar detailed analyses, if carried out for other concepts in the post-secondary curriculum, offer promise for a reformation of mathematics education. Our continuing work is devoted toward this effort.

REFERENCES

ADAPT. A Piagetian-based program for College Freshmen. (1976). Lincoln: University of Nebraska.


