

10

How to Draw Learning Curves: Their Use and Justification

Robert E. Shaw
Thomas R. Alley
Department of Psychology
University of Connecticut

"true rigor is productive, being distinguished from another rigor which is purely formal and tiresome, casting a shadow over the problems it touches"

—Emile Picard (cited in Kline, 1972, p. 1025)

INTRODUCTION: LEARNING AS A CUMULATIVE FUNCTION

Although learning theorists often disagree about what learning is, they agree that whatever the process is its effects are clearly cumulative and, therefore, may be plotted as a curve. By *cumulative* we mean that somehow the effects of experience carry over to aid later performance. This property is fundamental to the construction of "learning curves." The justification for drawing learning curves is the belief that such a cumulative function exists; that, mathematically speaking, the function is *well defined*. But is it? It is one thing to assume a function exists because you need it, and quite another to define it rigorously and justify its use.

This chapter examines this question, clarifies the definition of learning, and proposes a function whose existence justifies drawing learning curves. Finally, we show how this function may represent learning as a lawful process. This is important if we wish to justify the method for drawing learning curves rather than merely take it for granted.

The Cartesian Method for Drawing Learning Curves

All traditional techniques for drawing "learning curves" presume that successive data-points can be connected by a (smooth) curve. Typically, we use graphs to display such results because they provide a clear, succinct, and accurate way to reveal any cumulative effects over successive "trials" that may exist. What must we assume about learning to draw the corresponding curve?

Although there are a variety of methods for constructing "learning curves," they all assume that successive trials or episodes in a learning series may be plotted along the abscissa (x -axis), response characteristics along the ordinate (y -axis), and that the data-points distributed in the xy -plane may be legitimately connected by a curve. This is the Cartesian method. Everyone who draws learning curves tacitly assumes its validity (Kling & Riggs, 1971, p. 609). Using the Cartesian method to construct learning curves implies that learning is a function that maps values on the y -axis into values on the x -axis, and also that the mapping is at least continuous over the x -axis.

We say "at least" because, even though the function must be continuous over the x -axis if a curve is to be drawn, it may not also be continuous over the y -axis. Instead, it may be argued that learning is discontinuous over the y -axis (e.g., Greeno, 1974; Krechevsky, 1938). Under such an assumption, the function involved would still be represented by a continuously connected curve but the curve would be "stepped" over successive trials; that is, it would be a step function. Consequently, to appreciate what is at issue, one must not confuse stepwise-discontinuity over the response measure with topological continuity (connectivity) over successive trials. The latter assumption rather than the former underlies the Cartesian method for drawing learning curves of any sort—including those that fit an all-or-none hypothesis about learning (see Restle, 1965).

If learning curves are to be justified, there must be reason to believe that the x -axis represents a continuum, and trials a series of samples of that continuum. For if it is not a continuum, then how do we know whether the trials truly represent samples of the same phenomenon—whether there is a dimension called "learning" over which measurements might legitimately be taken? This is implied when we assume that successive episodes belong to the same *series*, as when we assume that a series of trials constitutes *one* experiment rather than a succession of distinct experiments. This also underlies our designation of experimental (independent) variables.

What guarantees that the successive samples do comprise a series? This is the crucial question for deciding whether or not the Cartesian method of curve drawing is justified for learning data. There are two answers to this question: One answer is that the successive samples may be connected because there is a *lawful* relationship among the samples (or data-points); a second answer is that the samples may be connected because there is some operationally defined *rule*

for doing so. Such a law or rule for defining a series of relationships for connecting values (e.g., xy -values) is exactly what we mean by a *function*.¹ The mathematical justification of the Cartesian method requires only that a rule be found for defining the learning function, but much more is required to justify its scientific use.

Given the mathematical rule by which samples *may* be connected to form a curve, the question remains *should* they be so connected? Are the constraints on the system responsible for the learning data a reasonable expression of the rule used by the system, or of a law governing it? An answer to this question moves us one step closer to understanding how an animal "recognizes" that successive situations are sufficiently similar to belong to the same series, a series over which "savings" might accumulate in the manner revealed by a learning curve. What the "mechanism" of learning is remains a mystery until we discover how the learner perceives the relation that connects the successive trials into a continuous series (cf. Koffka, 1935).

We must be careful not to jump to the conclusion that the scientist's view of the task, from an *external* perspective, is necessarily the same as the organism's

¹Historically, the most revolutionary proposal for how functions should be defined was made by the French mathematician Dirichlet in the 18th century. After witnessing many failures by others to provide completely satisfactory analytic or algebraic accounts of functions, he opted for liberalizing the concept so that it no longer lay in the exclusive province of mathematics but moved to the more general province of logic. He proposed that any value y be considered a function of a variable x whenever a precise principle of correspondence between x and y could be clearly stated. He no longer regarded it as indispensable that the principle of correspondence be defined by mathematical operations alone; rather, a purely verbal definition could suffice if it were logically precise. This compromise in rigor opened the door to the eventual development of a purely logical theory of functions, recursive function theory, on which the theory of computable functions is based, and by which a new interpretation is given to the meaning of "rigorous." The theory of computable functions provides an important alternative to the classical technique for rigorously formulating functions—an alternative, however, that exceeds the limits of traditional natural sciences such as physics.

Perhaps the notorious principle of complementarity owes its existence to this breach in the tradition of physics being the sole judge of how functions should be rigorously defined. This principle asserts that no description of phenomena in nature can be given completely using natural (dynamic) law but also will require the use of rate-independent linguistic variables. Hence all phenomena, especially those involving living systems, will require explanations that draw on both rule-defined (linguistic mode) and law-defined (dynamic mode) functions (Jammer, 1974; Pattee, 1979, in press).

Rule-defined functions provide the foundations for computer science, mathematical linguistics, and most major attempts in cognitive science to model the mechanisms responsible for psychological functions, such as perceiving, remembering, and learning. The tacit acceptance of Dirichlet's proposal is, most likely, a major reason for the close alliance we now witness being forged among these kindred disciplines.

The ecological approach to psychology is a notable exception that resists the rule characterization of function and, hence, of psychological functions. This approach prefers to redefine what is meant by law while conserving the original meaning of a function as the expression of some underlying law of nature. The purpose served by this strategy is that continuity between psychology and physics, biology and physics, and psychology and biology might be better preserved, and the paradoxes of mind-matter and mind-body dualisms might be avoided (see Turvey, Shaw, Reed, & Mace, 1981).

internal perspective. In designing an experiment, an experimenter can assume a vague, intuitive rule presumed to connect the trials into a series even if no such rule exists for the learner to follow, or if a different rule is actually used by the learner in perceiving the relationship over trials that links them to the same series. Because of this possibility, both theorists and experimenters may impute a simpler basis to the continuity over successive trials than is actually the case. Indeed, this may be the reason that so few questions have been raised about the continuity tacitly assumed to justify the drawing of learning curves. It is not enough simply to define operationally a function by which a curve might be drawn to fit observed data-points; one must also show why that particular function is a reasonable candidate for modeling the law- or rule-governed constraints actually responsible for the learning accomplished by the organism in a given context. Although we address the formal basis for relating members of learning series, explanation of the perceptual grounds by which certain temporally distributed events and not others join together to form a learning series is beyond the scope of this chapter.

Let us pause to summarize the steps that justify the use of the Cartesian method in drawing learning curves:

1. Plot the sampled data-points in an xy -coordinate space where the x -axis represents the independent variable and the y -axis the dependent response measure. This determines the discrete data-points to be connected.
2. Find a rule for operationally defining a function, $y=f(x)$, that describes a smooth curve that passes through every data-point. This is the function whose existence needs to be justified, for without it only discontinuous plots (e.g., bar graphs) would be justified.

Note that this second step is not a license for indiscriminately connecting adjacent data-points; rather it is a primary demand that before we connect the data-points we must first justify that the function assumed does in fact exist and, by existing, serves as a constraint that guides one's pen along a predetermined course for joining each point in a manner analogous to the law or mechanism that guides the learner through related learning episodes.

It is a popular but insidious mistake, a fallacy of method, to treat the learning curve (or any other curve for that matter) as having an existence independent of the function it expresses, or the function as having any reality apart from the law or mechanism that exhibits it. This is a dangerous tactic for science, for it risks mistaking a fiction for a fact. At best, the premature drawing of the curve serves only as the concretion of your hope or hypothesis, as the observer-turned-theorist, that the function will be found and scientifically justified—if not by you, then by someone else. Thus the final step toward justifying the production of learning curves is:

3. Specify the scientific constraints (e.g., the law or mechanism) that the learning function putatively expresses that make it more than a mathematical fiction.

Exactly how one should do this is a choice of method. Behaviorists will do it one way, cognitive psychologists another, and ecological psychologists still another. Where the behaviorist might search for the invariant relationship among reinforcement contingencies common over trials (a stimulus dimension), the cognitive psychologist might search for a rule associating common features (a mediating construct). The two approaches differ regarding the role that observables are believed to play in characterizing the dimension of "belongingness" by which successive trials form a connected series: Behaviorism emphasizes "external" states of the learning situation, whereas cognitive psychology emphasizes the "internal" states of the learner.

Learning as an Ecosystem Function

In sharp contrast to both of these approaches, ecological psychology seeks the required continuity in the covariation of two dimensions: one having its footing in the environment, the other in the organism. The fundamental postulate of the ecological approach is that these two dimensions are invariantly related by a pair of functions: perception and action. Gibson (1979) refers to this as the principle of organism-environment mutuality; a mutuality that might be modeled as a mathematical duality² (Shaw & Turvey, 1981; Shaw, Turvey, & Mace, 1982; Turvey & Shaw, 1979; Turvey, Shaw, & Mace, 1977). Let us consider briefly the relevance of the principle of mutuality, or duality, for the problem of characterizing learning.

Roughly speaking, perception is a mapping from the series of values taken by the environmental variable into the series of values taken by the organism variable, whereas action consists of the inverse mapping. There is, however, a

²A mathematical duality D is an operation that establishes a special isomorphic correspondence between one structure X (e.g., a series) and another structure Y (e.g., another series), so that for any function f that establishes a value in X there exists another function g that establishes a corresponding (dual) value in Y . Furthermore, a duality between the structures is not transitive, for if there exists another function h that putatively carries the image of X into another structure Z , then Z must equal X .

When D is its own inverse, it is sometimes said to establish "double" duals between X and Z . The ecological approach we are espousing postulates a doubly dual relationship between the values of X , taken as environmental properties, and corresponding values of Y , taken as organismic properties. We call the duality operation $f: X \rightarrow Y$, *perception*, and its values, *affordances*. We call the inverse duality operation $g: Y \rightarrow X$, *action*, and its values, *effectivities*. The double duality, D , consisting of the operations f and g , $D: X \leftrightarrow Y$, designates a system of constraints composing an *ecosystem*.

special relationship between the perceptual function and the action function that guarantees their covariation whenever the organism is successfully guided by perception through a series of felicitous regulatory acts (e.g., muscular adjustments) that achieve an intended goal (e.g., the grasping of an object). Under such felicitous circumstances, the two functions must become *duals* so that the course of values assumed by one constrains the course of values assumed by the other (see footnote 2) in an ongoing and mutual process:

$$P(t_0) \rightarrow A(t_1), A(t_1) \rightarrow P(t_2), P(t_2) \rightarrow A(t_3), \dots$$

Thus perception and action operate on each other reciprocally, in a "closed looping" fashion rather than in an open-ended, causal chaining fashion. The covariation that results is properly termed *ecological* because it involves both environmental and organismic variables.

Finally, the perceptual and action series spiral through space-time intertwined like a double helix, tracing out a path determined by a logic of circular constraint. In this way, ecological events involving perceiving and acting determine "fat" world-lines in space-time (four-dimensional) geometry (see Kugler, Turvey, Carello, & Shaw, in press).

In this scheme, learning is a lawful operation that increases the coordination between perception and action series. Metaphorically, learning is a function that tightens the constraints on the double helix, bringing the perceptual "helix" (or series) into closer alignment with the action "helix" (or series). This view of learning assumes the existence of two series whose values can be coupled such that certain values in one series are potentially duals of certain values in the other series. "Affordances" and "effectivities" are just such duals.

Affordances and Effectivities as Duals. An object that *affords* grasping by some organism is said to have the affordance property of "graspability" for that organism. The affordance property is a value that will appear in the perceptual series of a properly "attuned" organism (i.e., of an organism that has a "grasper" properly designed and controlled to accommodate the object in question). When an organism learns to use its grasper to achieve grasping of a given object, the affordance property also takes on a corresponding (dual) value in the action series. We call this dual value in the action series an *effectivity*.

These series are by no means simple; instead, they consist of higher order invariant properties of environmental situations and organismic states. The perceptual series is a series of affordances linked by the actions of an organism toward environmental objects. The action series is a series of effectivities linked by an organism's ongoing perceptions of environmental properties. Like affordances, effectivity values are complex, referring to the fact that for an organism to act, some appropriate effector organ must be connected to a repertoire of

control constraints (e.g., muscular adjustments) suitable for determining an act (e.g., grasping of x) that realizes some affordance goal (the graspability of x).

Because of our endorsement of the principle of organism-environment mutuality, we, as ecological psychologists, require two of everything listed previously: two sets of data-points—one showing how the learner's perception of the task variables changed over time and one showing how the learner's action (response characteristics) changed as a function of the task variables; two learning curves must be justified—one showing perceptual learning in the task and the other action learning; two mathematical rules must be found for determining how the data-points are to be connected to form a series; two laws must be discovered to justify the scientific use of the pair of functions; and finally, in keeping with the concept of affordance, the two laws must be shown to "fuse" into a *dual* pair of reciprocal equations consisting of complementary variables—one equation to describe changes in the perceptual series and one to describe the dual changes in the action series. Here again the concept of dual is the mathematical one rather than the philosophical one and refers to a special relationship between two equations (or functions), so that the solution to (definition of) one specifies the solution to (definition of) the other.

We return to this discussion of duality in the ecological approach to learning toward the end of this chapter. In what follows now, we show how one might address each of the three points discussed earlier. Although our illustration is a serious attempt to provide a first pass on an ecological approach to learning theory, we hope that readers who disagree with that approach will, nevertheless, find some benefit for their own theories from an examination of the method used.

Step 1. Plotting Data-Points Sampled from a Continuum

Generally, we might define learning as a cumulative function, L , that determines a mapping between two series: a *perceptual* (stimulus) series and an *action* (response) series. The perceptual series consists of episodes (e.g., trials) on which the learning function applies to increase the value of some response characteristic of the organism over time. This means that we can analyze the global function L into a series of lesser functions that apply locally to give the exact increment to the y -values (response characteristics) at each x -value (trial) as follows:

$$L = f(x)_1 + f(x)_2 + f(x)_3 + \dots = y_1 + y_2 + y_3 + \dots, \quad (1)$$

$$\text{or } L = \int_a^b f(x)dx,$$

where a and b are first and last trials, respectively, in the series of trials run.

There is, however, a subtlety in this characterization of learning that should not be overlooked: The "trial" variable must be considered to be imbedded in a continuum or else no curve can be drawn to connect the data-points in the Cartesian manner (where each distinct data-point is an (x, y) value in the plane of the graph $y=f(x)$). The continuity requirement is satisfied, however, by the assumption that L is a function that is continuously summable (integrable) over x . This is tantamount to the claim that x is a continuous variable and, therefore, qualified to be an axis over which a time-varying process might be well defined. This assumption that the x -intercepts of all adjacent data-points are at least integrable over time guarantees, in principle, the connectability of the data-points into the desired learning curve. What kind of function must L be for this assumption to be justified?

In the next section we propose an answer to this question that accords with some of the most general facts known about learning.

Step 2. Defining Learning by Analogy to Dissipative Functions

Learning is not merely a simple accretion of a response tendency with repeated experience; often, if not always, specific and nonspecific changes in the general disposition to learn also accompany these changes in response characteristics. Therefore, any function used to represent learning must consist of two distinct parts: a "response" variable and a "state" variable (irrespective of whether the state variable is taken to be behaviorally, cognitively, or ecologically defined). The *response* variable expresses an "observable": the observed change in the behavior of the system. The *state* variable expresses a "dispositional": specific or nonspecific changes in the disposition of the system to learn. When this change in disposition is nonspecific, fosters learning, and persists over an extensive interval of time, we refer to it as the formation of a "learning set" (e.g., Harlow, 1949) or, more generally, as "learning to learn" (e.g., Bransford, Stein, Shelton, & Owings, 1981). Conversely, when the nonspecific transfer effect is increasingly negative (inhibiting learning), we call it "fatigue." Both of these generalized changes in the disposition of organisms to learn are *dissipative* parameters: Learning to learn can be considered the dissipation of inhibition (e.g., of distractions) and fatigue the dissipation of facilitation (e.g., of interest or energy).

Historically, the mathematical characterization of functions involving dissipative parameters has proven problematic. For instance, the classical treatment of the behavior of a spring under differential loadings assumed the validity of Hooke's famous law asserting that strain (stretching or compressing) is directly proportional to stress (restorative force). Unfortunately, the problem is more complicated than this because the coefficient of elasticity of the spring, a dispositional variable, tends (like the disposition to learn) to change with repeated use. Hooke's simple law fails to take this change into account. In Hooke's day,

it was not mathematically possible to formulate a function that included the effects on the spring's behavior of both the stress force and the dissipation of elasticity arising from stress. In fact, the mathematics needed to provide a generalized version of Hooke's law did not become available until nearly two centuries later when Volterra, the great Italian mathematician, turned his attention to problems involving "hereditary" influences (Kramer, 1970). Dissipative changes in the disposition of a system to respond is but one among a class of influences that may operate persistently to alter dynamically the response characteristics of a system, be it living or not.

A formal analogy holds between the formulation of laws required to relate observable to dispositional variables in physics and those required to serve a similar function in learning theory. (Some of the historical and technical details of this analogy are contained in Appendix A.) We believe that the task of formulating learning functions will be made easier if we examine how physics solved the problem of designing a general form of Hooke's law for springs: one that included dispositional variables that change as a function of use. With this as a guide, we might formulate analogous functions for learning with nested dispositional variables that also lawfully change their values as a function of use (practice). Such a formulation would satisfy Step 3 of our procedure for justifying the drawing of learning curves by showing how the underlying function could be continuous over the x -axis.

Common abstract solutions to the two classes of problems arise because learning systems, like physical systems with dissipative parameters, are governed by *hereditary* laws: laws that express the effects of hereditary influences on the state (dispositional) variables of a system. If this is so, the problem of formulating rigorous laws of learning becomes an aspect of the more general problem of discovering the laws of what Picard (1907) called "hereditary mechanics." Picard coined this name for a new discipline to emphasize the fact that no existing form of mechanics yet included the study of laws involving hereditary influences; not classical mechanics, variational mechanics, quantum mechanics, or relativistic mechanics, for none of these approaches, in principle, can explain the behavior of systems that are governed even in part by dispositional variables that serve a "record-keeping" function (Jammer, 1974; Pattee, 1979, in press).

Psychology is reputed to be the science of systems that do keep "records" that influence current states, yet it has not developed the techniques needed for characterizing functions that incorporate dispositional variables: new functions whose courses of values are steered by hereditary influences. Instead, the field has concentrated on developing "mechanisms" founded either on old functions borrowed from the classical period of mathematical physics or on rules so intuitive that the functions they entail escape analysis.

Much help can be obtained by exploiting analogic connections to abstractly similar problems in older fields, like physics, where formal techniques and scientific methods may be better developed. Thus our problem may be better

understood by analogy to the struggle of physics to define rigorously dissipative functions, a close analogue of our problem. Consequently, let us treat physics as a resource field from which certain formal tools might be borrowed. No reductionism is necessarily implied by such a strategy because the variables to be included in the learning function, the response variable and the dispositional variable, are psychological rather than physical. It is only the logical scheme, or syntax, of the law that we intend to borrow to help determine the type of function needed to model learning. We beg the reader's indulgence as we delve into unfamiliar matters drawn from the history of physics and mathematics. Although one half of the analogy, the resource field of hereditary mechanics, may be unfamiliar, the second half, the test field of learning theory, is not. Our intention is not to import psychology into physics nor physics into psychology, but to clarify the function by which learning might be defined by justifying learning curves.

Learning as a Problem for Hereditary Mechanics

A function constrained by a hereditary influence (e.g., a dissipative parameter) cannot be captured by the ordinary means available to traditional physics for expressing functions, or by the laws they may represent. The means available have traditionally been either differential or partial differential equations and their integral counterparts. But none of these will do. (For discussion of the reasons, see Appendix A.) Because the laws of classical (variational) mechanics depend exclusively on rigorous expression in differential (integral) form, dissipative phenomena (such as the effects of fatigue on the behavior of an elastic system), although perfectly lawful, failed to fit the form of any known laws. This can be seen in the case of elastic systems (e.g., springs) alluded to earlier.

Experimentation revealed that the behavior of elastic systems was not merely a function of its last initial condition, but also of the entire history of its initial conditions.³ For instance, as pointed out earlier, a spring's behavior is not

³Traditionally, two forms of physical laws are recognized: those that map initial conditions (forward in time) into final conditions, and those that map final conditions (backward in time) into initial conditions. The mathematical form of the "forward" acting law was that of a differential equation, whereas that of the "backward" acting law was that of an integral equation (see Appendix A). Classical mechanics developed laws that were exclusively differential in form, whereas *variational mechanics* was the name of the discipline that was devoted to the development of laws that were integral in form. These two approaches to mechanics expressed two distinct philosophical, even theological, orientations. Briefly, their differing motivations can be explained as follows: Because causes necessarily precede effects, the differential form, as used by classical (Lagrangian) mechanics, was deemed the appropriate function for expressing "causal" laws. By contrast, because goal-directed systems are constrained by the final (often optimal) state toward which they tend, the integral form, as used by variational (Hamiltonian) mechanics, was deemed the appropriate function for expressing "teleological" laws. As discussed in Appendix A, however, the defense of this distinction is not so straightforward as might first appear.

predictable from Hooke's law, which asserts that *stress* (restorative force) is directly proportional to the *strain* (change in length) undergone; instead, careful experimentation showed stress to be a function of the entire history of the strains undergone by the spring on all previous occasions (Lindsay & Margenau, 1957). Hence, strain acts as a hereditary influence on an elastic system, exerting a cumulative fatigue effect on its later behavior. What makes systems that operate under hereditary influences impossible to explain by the ordinary laws of mechanics?

Hooke's law can be represented by the equation

$$y = -kx \quad (2)$$

where x is the strain behavior of the spring (i.e., how far it will stretch), y is the stress (the tendency of the spring to restore itself to its original length when stretched), and k is the coefficient of elasticity of the system. From this equation we see that elastic systems with high coefficients have greater restorative forces than those with lower ones. Therefore, after a spring with a high k -value is loaded, we would expect it to stretch much less than a comparably loaded spring with a low k -value. Furthermore, according to this equation, we should also expect springs with the same k always to stretch by the same amount under identical loadings, irrespective of the distance or number of times they had been stretched in the past. As noted previously, this was shown not to be the case. Frequency of use and style of use have cumulative (hereditary) effects analogous to practice in the case of learning. Elasticity (k) decreases according to some function not contained in Hooke's simple law. Hooke's law treats k as a constant when it is really a variable that exhibits a rather orderly change in value over time proportional to its history of use. This proportion also requires a function for its rigorous expression.

Stated more generally, we recognize that what is needed is a law statement involving two kinds of functions: an observable or behavioral function, and a dispositional or state function. Hooke provided us with a law statement involving only one function and a constant, but clearly there can be no generality to such an equation with respect to historical effects. Even if we make k a variable instead of a fixed value, we would have to determine its exact value empirically just before loading the spring for a new test of the law. Moreover, we would be unable to generalize to future uses of the spring because the test itself would have required loading of the spring, which would again alter the value of k : a vicious circle!

In learning theory this would be analogous to trying to express learning over a series of trials as the function of an unspecified variable whose behavior is not known, but which cannot be determined without running the subject in still another learning trial, which itself alters the value of the variable: another vicious circle! This unspecified variable, like the k in Hooke's law, is a dispositional term referring to an unspecified state of the system: the capacity to learn under

the stipulated circumstances. If this unspecified variable can, like k , be treated as a fixed value (i.e., a constant), the equation predicting learning will be well behaved when the appropriate value is plugged in. On the other hand, should this value change as some linear or nonlinear function of exposure to the experimental variables, then, to that extent, the equation will fail to predict and will be an incomplete characterization. More importantly, learning will appear to be unlawful even though it is actually lawful. Thus, we may conclude that learning either is or is not law governed only when we are sure that our equation is complete and that the behavior of *all* its terms is functionally specified.

The existence of learning to learn demands a search for the appropriate formal means of expressing the dispositional term as a well-behaved function. It will not do simply to treat the capacity of an organism to learn as fixed, for then any lawfulness that might exist will escape detection. Nor will it do to try to circumvent the problem by treating learning in terms of difference equations rather than the more demanding technique of functional analysis required in hereditary mechanics, for such difference equations are themselves not well behaved under circumstances where terms refer to functions of functions. This occurs, for example, where dispositional functions change their course as a function of behavioral functions (Greeno, 1974). Functions that are the function of other functions rather than of variables are called *functionals*. Strictly speaking, all hereditary functionals are in a sense nonlinear and, therefore, cannot be adequately expressed by linear equations; they can only be approximated. Moreover, as others (e.g., Herrnstein, 1979) have recognized, learning is better described by nonlinear than linear models. (See Appendix B.)

If, as we have argued, learning is a *functional* rather than merely a function, the linear equations typically used (as in traditional physics) will fail to capture the significant and necessary role played by hereditary influences. This suggests that a better understanding of the nature of this limitation may prove helpful in understanding the form that realistic formulations of learning must take if any lawfulness present is to be expressed.

In the parlance of mechanics, we express this limitation of traditional law statements by saying that laws must be "initialized"; that is, the *initial conditions* for a law are not themselves an intrinsic part of the equation of the law but must be added before the law can be applied.⁴ This seems to make initial conditions

⁴Initial conditions are the values that must be given to the free parameters in the equation for a law statement in order to be able to predict final conditions (outcomes) dynamically. For instance, applying the laws of mechanics to predict the final resting places of all the billiard balls struck in a game of billiards requires that we specify the momentum of each ball in the chain reaction. But because momentum is equal to the mass of a ball times its velocity, and the momentum of each subsequently struck ball is determined ultimately by the impulse force (force \times time) applied to the first ball, the mass and velocity terms are initial conditions that must be specified before the equations for the applicable laws of motion can be evaluated to predict the final locations of each ball after the chain of collisions.

an indispensable afterthought, something sorely needed for the practical application of the law but unencompassed by it. Then why not simply build laws that incorporate their initial conditions as an intrinsic component of the law statement? Why not simply construct equations for learning or mechanical laws out of functionals that incorporate functions of hereditary functions? Unfortunately, this is easier said than done.

Classical mechanics (and all other branches of science that choose differential equations for expressing laws) is prohibited from taking this course of action because there is formally no way to place initial conditions within a differential form and still have a solvable equation.⁵ This is because a hereditary influence requires a separate and distinct form of expression. Thus to formulate a law that incorporates hereditary influences like learning, we must liberalize our concept of the mathematical form that law statements might take. We must look beyond simple differential or integral forms and entertain some other possibility. What might this new form for a law equation be?

To answer by way of a recapitulation, note that three properties of learning functions have been discussed:

1. Learning functions are *cumulative* (i.e., on the average, they increase monotonically).
2. The cumulative nature of such functions is, in general, positively (or negatively for aversive conditioning) *constrained by hereditary influences* on analogy to dissipative parameters.
3. The generic learning functional, most likely, will prove to be *nonlinear*, with linearity being a special, if important, case.

Taken together, these three properties suggest that learning functions must be characterized by a *nonlinear integro-differential form*. In the next section we

⁵Mathematically speaking, integral forms are opposite, or complementary to, differential forms: Whereas, under different initial values a differential equation specifies the values of individual points lying on a curve, an integral equation provides a general solution to the whole set of distinctly initialized differential equations, treated collectively and simultaneously, and thereby yields the curve itself. Thus, differential forms sample the curve piecemeal whereas integral forms exhibit the whole curve so that it might be sampled. For this reason, the solving of a differential equation requires integration whereas the solving of integral equations requires differentiation. The attempt to nest all the initial conditions needed to evaluate fully a law statement couched in differential form would be ludicrous, because this would require the law equation to be solved prior to its application; it would require solutions to a potentially infinite set of differential equations before any application could be made: an impossible demand.

The strategy of attempting to embed the final conditions under the law function in order to postdict the initial conditions, in the manner of (Hamilton's) variational mechanics, is similarly ill fated, for this would also require a priori knowledge of an inverse nonlinear hereditary functional on k : the very thing being sought. The proper strategy for avoiding this dilemma was taken by Volterra (see Appendix A).

discuss some of the problems that must be overcome before integro-differential equations can be made completely appropriate for characterizing learning.

Step 3. A Hereditary Law Form for Learning

Like the law forms of classical physics that ignored the dissipative parameters of elastic systems, the law forms adopted by classical learning theory, stochastic learning models, and cognitive approaches also fail to provide a principled place for the hereditary aspect of learning. This is not to say that learning equations have not been written. They have, but to our knowledge, no innovation in the construction of law forms has resulted from such attempts.

For example, neither reinforcement (causal) nor expectancy (teleological) theories succeeded in providing new law forms that can explain all types of learning. Their failure is analogous to the failure of their corresponding law forms in mechanics. As discussed earlier, because of their dependence on differential and integral law forms, neither classical ("causal") nor variational ("teleological") mechanics could provide a strictly lawful explanation of elastic systems. Apparently, some law form more powerful than either of these is needed.

It is not our purpose to develop the mathematics of hereditary systems, but we do wish to provide the abstract form of the functions defining the behavior of such systems because this, presumably, is the form that learning functions must take if the analogy proposed is valid. Consequently, here we present only the abstract equivalent to the actual mathematical expressions required. We would also like to contrast the form of functions that involve hereditary influences with those that do not.

INTEGRO-DIFFERENTIAL EQUATIONS AS THE LAW FORM FOR LEARNING

The abstract form of all hereditary laws, including learning, is that of an *integro-differential* equation. The general form of this equation is

$$y(t) = kx(t) + \int_a^b K(t,s) x(s) ds \quad (3)$$

The hereditary influence is defined over the temporal interval from time s to time t , as associated with the *occasions* (a to b) when samples of the x -series are taken (e.g., from Trial a to Trial b). Here $y(t)$ expresses the value of the *behavioral variable* (response characteristic) y observed at time t , whereas $x(t)$ is the value of the *dispositional variable* x also at time t . The constant k scales the

initial value of the dispositional (state) variable (such as the elasticity coefficient or capacity to learn). Finally, $K(t,s)$ is an integral transform called the *coefficient of hereditary influence* which acts over the interval from s to t (of the trials continuum) and determines the series of values x assumes over that sampling interval. This change in the x variable is measured in units $x(s)ds$. Such equations may describe a *learning* function when $y(t)$ is, on the average, increasing, and a *decay* function (for habituation or fatigue) when $y(t)$ is, on the average, decreasing.

To simplify our presentation, we treat the integro-differential form algebraically in "operator" notation. An *operator* denotes a mathematical operation that converts one function, say $x(t)$, into another function, say $w(t)$. (Note that an operator is not itself a function; instead, functions are the "objects" of operators.) We use k as the special symbol for an operator. Using operator notation we can portray, abstractly, the major difference between the classical formulation of Hooke's law (Eq. 2) and the more general hereditary formulation known as the *generalized* version of Hooke's law. The major terms of this equation are represented by the algebraic operators that act as place holders for integral and differential forms as follows:

$$y = kx + K \quad (4)$$

In this equation, K can be called a "hereditary transform of k ." Equation 4 asserts that strain x (of, say, a simple spring) is not simply proportional to stress, as claimed in Hooke's original formulation of his law (Eq. 2), but instead depends on all the values that x assumed from the time of application of initial stress to the time now being considered. In other words, K is a series of values (perhaps nonlinear; see Appendix B) that must be added to k to change its course of values over time; this, of course, changes the series denoted by the independent variable x .

With the hereditary formulations of Equations 3 and 4 in hand, the process of learning can at last be translated, through analogy, into a hereditary functional. For this analogy, learning situations may be viewed as placing a "stress" on an organism's general learning capacity (k), thereby causing a relatively permanent change in response tendency (y), as reflected in "strain" (average monotonic increase) of the relevant response characteristic (x), graphically shown as the amplitude of the learning curve. The hereditary influence, designated by the transform K , represents the cumulative effect of learning to learn over the series of trials; an effect not explained by the facilitative stress applied on each individual trial.

We urge learning theorists to consider adopting a hereditary law form for learning functions. No one can yet tell whether hereditary law forms will prove more adequate than their predecessors. We believe they will, because they can express the rather peculiar circularity (nonlinearity) of dispositional variables,

like k , being dependent on the behavioral variables that they help to determine. Such "circular logic" (Turvey, Shaw, & Mace, 1977) for nonlinear learning laws stands in sharp contrast to the more traditional, "one-way" transitive logic of the classical attempts to formulate linear laws for learning. The potential power of the circular transitivity of the laws of hereditary mechanics is much like the recursive power of rule-defined learning functions favored by cognitive approaches that attempt to model learning formally, or by computer-driven simulation programs. We favor the hereditary functional formulation over the recursive rule formulation, because it permits the search for laws of learning to continue as an extension of the classical tradition while maintaining close ties, without reductionism, to other sciences (e.g., physics) that may continue to function as useful resource fields for our test field of psychology.

In the next section we discuss some more specific aspects of treating learning as a hereditary functional from the perspective of ecological psychology.

THE ROLE OF COMPLEMENTARITY AND DUALITY IN AN ECOLOGICAL PSYCHOLOGY OF LEARNING

An essential ingredient of an ecological theory of psychology, following the principle of organism-environment mutuality (Gibson, 1979; Turvey & Shaw, 1979), is the existence of a perception-action "loop" by which an organism is related to its environment. This mutuality principle asserts that an organism's perception of the environment provides control constraints for its actions, and, conversely, that an organism's controlled encounters with its environment (via its action system) provide constraints on the perceptual information that unfolds as a function of such encounters.

In the language of control systems engineering (see Nagrath & Gopal, 1982), such a loop between two variable-state systems constitutes a *duality* (i.e., a mutuality of constraints) of controllability and observability. *Controllability* is a property of a system when, given any initial state and any designated final state, there exists a time interval and an input that determine how the system moves from the initial state to the final state during the specified interval of time. *Observability* is a property of a system when, given any final state and any designated initial state, there exists a time interval and an output that determine how the system moves from the final state to the initial state during the specified time interval.

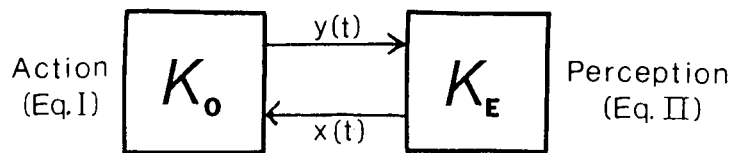
We have met this duality principle before in our discussion of the relationship between laws stated in the differential ("causal") forms of classical mechanics and in the integral ("teleological") forms of variational mechanics. The concept of duality is now sufficiently ubiquitous to call it a *fundamental* principle of all mathematical science. (See Lautmann, 1971; Patten, 1982; Shaw & Mingolla, 1982; Shaw & Todd, 1980; Shaw & Turvey, 1981 for examples of its applicability in fields ranging from physics to ecology and psychology.)

Duality and Ecosystems. Because an ecosystem consists of both energy interactions and information transactions between organisms and their environment, one might argue that the principle of complementarity applies in psychology as well as in physics and biology. On the contrary, we suggest that in an ecosystem where perception and action are equal, concomitant, and intrinsic components, the principle of complementarity is usurped by the principle of duality. The principle of duality is the manifest constraint that permits an ecosystem model to circumvent the principle of complementarity of energy and information by construing these two concepts as a duality rather than a dualism. The principle of complementarity asserts that living systems may not be adequately explained by the dynamical (rate-dependent) laws but also require information (rate-independent) rules as well. We see the need for these two modes of description most emphatically in the apparent dualism of the enzyme-folding (dynamical) process by which DNA molecules replicate as contrasted with the double-helical code that provides the informational constraints that guide the replication process (Pattee, 1979, in press).

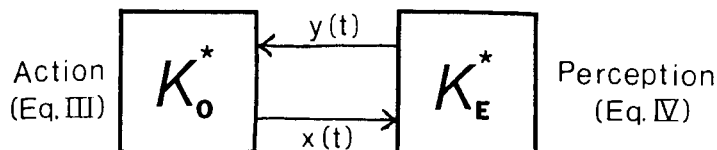
The concept of *control*, interpreted for ecosystems, corresponds to the *means* by which actions and perceptions are determined. Such means are the operators by which the organism processes energy to produce actions, on the one hand, and the operators by which the environment structures the energy distributions that are perceived, on the other. Let K_0 and K_E designate the organism and environment operators, respectively, and x and y their respective energy inputs. The corresponding control equations are ($y = K_0x$) and ($x = K_Ey$).

The concept of *observable*, interpreted for ecosystems, corresponds to the *goals* toward which actions are controlled and that perceptual information specifies. An *effectivity* is a goal-directed function that an organism might, in principle, realize. When the goal-directed function is also determined by the environment, it is called an *affordance*. The "observable" properties of an ecosystem are effectivities and affordances. Using this same operator notation, the information that renders each of these properties observable can be given by ($x = K_0^*y$) for an effectivity goal and ($y = K_E^*x$) for an affordance goal, when the energy control equations are also given. Figure 10.1 illustrates how the energy support for perception and action (control) and the information specification of affordances and effectivities (observables) come together to define an ecosystem as a *closed loop* over a pair of dual external states.

An ecosystem requires four equations for its complete specification: a pair of dual energy equations that are dual to a pair of dual information equations. Table 10.1 presents these equations in their generic integro-differential form (in contrast to their operator form as in Fig. 10.1). Because these equations involve hereditary transforms (K_0, K_E, K_0^*, K_E^*), they express the generic form of the hereditary law governing an ecosystem's history of energy interactions and information transactions. The operators designated by the asterisks denote the hereditary transforms of integrals that specify information transactions, and that run temporally counter to the causal direction of energy interactions (represented by the integral



I. ENERGY (CONTROL) DUALS



II. INFORMATION (OBSERVATION) DUALS

FIG.10.1 Diagram of the energy (control) and information (observation) duals shown in operator notation. The corresponding equations (denoted in parentheses) for perception and action are given in Table 10.1.

TABLE 10.1
Table of Duals

		Perspective Duals	
		Organism-Perspective (Action)	Environment-Perspective (Perception)
Temporal Duals	Energy (Control)	$y(t) = \int_0^t K_O(t,s) x(s) ds$ (Equation I)	$x(t) = \int_0^t K_E(t,s) y(s) ds$ (Equation II)
	Information (Observation)	$x(t) = \int_0^t K_O^*(s,t) y(s) ds$ (Equation III)	$y(t) = \int_0^t K_E^*(s,t) x(s) ds$ (Equation IV)

Note: The symbols in these equations are explained in the text in connection with Eq. 3. The transforms whose kernels take the form $K(t,s)$ designate "past pending" integrals that sum energy interactions from some past state (s) to some future state (t). By contrast, those with kernels of the form $K^*(s,t)$ designate "future tending" integrals that sum information transactions from some potential future goal state (t) "backwards" in time to some current state (s).

equations whose operators have no asterisks). These dual information and energy integrals express a temporal symmetry analogous to the temporal contrast expressed by the equations of variational and classical mechanics, for there are both future- and past-pending forms. Where the energy integrals (Eqs. I and II in Table 10.1) express the *means* by which energy interactions between an environment and an organism provide the "drive" for the action and perception systems, the information integrals (Eqs. III and IV) specify the effectivity and affordance "goals" that the information transactions signify.

Nature must allow organisms to solve this predicament: Given a certain need state, how can an organism being showered only with the present energy (time-forward) interactions reach a stipulated future goal state in an optimal fashion? The answer is that it cannot! Intentional behaviors cannot be reduced to causal relations. Rather, the organism must learn to use hereditary influences and current information (time-backward) transactions to guide it optimally toward the required goal state. The arguments raised by supporters of teleological views against the causal approaches of classical mechanics and reinforcement theories were not so much off the mark as merely one-sided. Neither variational nor expectancy theories alone can be complete: A theory that incorporates the causal and intentional, energy and information, as dual forms of the same law, however, might be complete or, at least, completable in principle.

Because the four equations in Table 10.1 are intrinsically related by duality operations, the ecosystem can be said to be *self-controllable* and *self-observable*. These four equations in concert provide us with the two coordinated series needed for learning. Hence, if these series have solvable equations, then, contrary to the complementarity perspective, the learning functional coordinating them exists. If we identify the learning functional with the dual equations (Eqs. I and III) having K_O as an operator, the composite functional provides a rigorous characterization of action (response) learning: With the dual equations (Eqs. II and IV) having K_E as an operator, it provides a rigorous characterization of perceptual learning.

Presumably, the overall learning series requires the sum of the hereditary influences of both perceptual learning and action learning: a summation performed by the learning functional. We might consider learning over the coordinated series to be complementary such that:

$$(K_O x + K_E y = 1) \text{ for } 0 \leq K_O < K_E \text{ or } 0 \leq K_E < K_O. \quad (5)$$

Learning with respect to Equation I might be interpreted as improvement in biomechanical coordination so that energy resources are optimized. By contrast, learning with respect to Equation II might be interpreted as improvement in perceptual strategies producing a refined use of the energy patterns required to orient and detect. Similarly, learning with respect to Equations III and IV pertains

to the selection of optimal paths from potentially attainable effectivity and affordance goals through the state-space of subgoals that define the current action state and perceptual state of the system, respectively. These four kinds of learning might contribute differentially to the overall change in response characteristics.

The amount of variance explained by each of these series can be experimentally ascertained by *within-subjects* designs with repeated measures where the subjects are brought to criterion on the perceptual displays incorporated in the learning, as contrasted with a similar design where subjects are permitted practice on the response measure. Overall, then, the process of learning will include three fundamental sources of variance: the dual (energy and information) perceptual learning series, the dual (energy and information) action learning series, and the learning-to-learn contribution over each of these series.

SUMMARY AND CONCLUSIONS

Learning theorists were once content to adopt the linearly causal law forms developed by traditional physics. Later, after seeing the failure of such law forms to explain all kinds of learning, and under a variety of incisive criticisms (see, e.g., Hinde & Stevenson-Hinde, 1973; Johnston, 1981; Seligman and Hager, 1972), many learning theorists abandoned altogether the quest for laws in favor of rule descriptions for learning functions. In this chapter we have offered reasons for reconsidering the case for learning laws. Let us close with an assessment of the formal analogy between our test field (learning) and our resource field (mechanics).

Our historical junket (Appendix A) has revealed considerable support for the proposed analogy between physics and psychology. The validity of the analogy is seen to be a consequence of a shared belief that laws are desirable, attainable, and capable of rigorous formulation. We have argued that strong analogical connections arose because of the mutual desire to formulate law-based equations in both fields. Both fields tacitly assumed that such laws must take on the analytic forms developed by classical physics: the familiar differential or integral forms. This analogy was strengthened by noticing that reinforcement theory stands to classical mechanics on the "causal" interpretation of differential forms as expectancy theory stands to variational mechanics on the "teleological" interpretation of integral forms of law (see Appendix A). The discovery of hereditary (dissipative) phenomena, not expressible in either of these forms alone, led us to consider the hybrid integro-differential forms required to express the laws of hereditary mechanics: laws that incorporate the historical series over which initial conditions are "updated" to reflect hereditary influences.

Acceptance of the proposed analogy between learning theory as our test field and hereditary mechanics as a resource field entails serious consideration of the

integro-differential form as the law form appropriate for modeling learning functions. Because learning acts as a hereditary functional over trials, as evidenced by learning to learn, rather than as a simple linear function, only this law form can fully justify the drawing of learning curves.

ACKNOWLEDGMENTS

We are grateful to J. J. Jenkins, T. D. Johnston, P. N. Kugler, W. M. Mace, E. Mingolla, and M. T. Turvey for their helpful comments and discussion during the preparation of this chapter. The second author was supported by a Postdoctoral Fellowship from the National Institute of Dental Research (Grant DEO-7047).

APPENDIX A

Analogical Law Forms in Physics and Psychology

The classical period of mathematical analysis, the 17th and 18th centuries, gave birth to the optimistic view that physical science could soon be completed with mathematical rigor and its foundations secured with axiomatic certainty. Near the end of this period, Lagrange (1788) was confident enough to boast in his highly influential book, *Mécanique Analytique*, that he had developed Newton's mechanics to such a high state of mathematical rigor that he had banned forever from science the need for diagrams and (therefore) geometric intuition.

In the mid-20th century a similar optimism regarding the role of mathematical rigor in defining deterministic laws was entertained by learning theorists. Was there good reason for this similarity in scientific attitudes shared by physicists during the classical period of mechanics and psychologists during the classical period of learning theory? Why do scientific determinism and mathematical absolutism go hand in hand?

Analytic Functions as the Classic Law Form

Mathematicians of the 17th and 18th centuries knew only of functions that were *analytic*, a property of being represented by a convergent Taylor series: a particularly simple kind of series, known as the power series, which always converges (has a limit) over a continuous range. Consider the infinite series,

$$a_0 + a_1(x - k)^1 + a_2(x - k)^2 + a_3(x - k)^3 + \dots,$$

in which a_0, a_1, a_2, \dots and k are arbitrary, fixed finite numbers, and x is variable. Note that the series is based upon successive integral powers of $(x - k)^n$ —hence the name "power series." The power series always converges for the value $x = k$

because it then reduces to its first term, a_0 . The series may also converge for other values of x , thereby defining a continuous range of values of x with the point $x = k$ occupying the center of the range. Within this range, the power series defines a continuous function whose derivatives are all continuous.

Because all functions dealt with by mathematicians of this era were expressible in terms of such series, they erroneously concluded that all mathematical functions were necessarily analytic (i.e., were expressible in terms of infinite series that are both convergent and continuous). It scarcely need be noted that geometric curves can be drawn using the Cartesian method to represent such functions graphically rather than numerically.

For the classical physicist, the belief that series of observations or measurements could be presented graphically via the Cartesian method, and that all such functions so represented could be expressed by Taylor series, gave mathematical expression to a strict determinism. Given a small set of observations, predictions and postdictions could be made by means of geometric intuitions or (later) by the rigorous application of the technique of analytic continuation. Hence the very concept of being *determined by law* became synonymous with being *expressible by analytic function*.

It should be obvious that an implicit, if little discussed, assumption that learning functions must be analytic functions (i.e., that learning series are Taylor series) lies behind the construction of learning curves: For how else could one justify extrapolating from discrete data points to a smooth continuous curve that converges on some criterial value as its limits? Little wonder then that a similar attitude of strict determinism prevailed among learning theorists of the mid-20th century, and that discovering laws of learning became synonymous with formulating the learning function in an analogous analytic manner.

For the process of drawing continuous curves from discrete data points to have greater justification than simply geometric intuition or ill-founded convention, another property of analytic functions must hold for learning: the property of *analytic continuation*. Gestalt psychologists recognized this property as a perceptually based, geometric intuition they called "good continuation," whereby any curve or line "carries its own law within itself" (Koffka, 1935, p. 175). This geometric intuition can be understood in a more rigorous fashion when expressed in terms of analytic functions.

The perceptual tendency to continue a curve from an arc can be attributed to the fact that a small arc generatively specifies a longer curve. An arc is mathematically equivalent to an interval of all values of x for some analytic function $f(x)$ from which the derivatives specifying a longer Taylor series may be extracted (d'Abro, 1951). Thus the Gestalt law of good continuation, intuitively used in the extrapolation of learning curves, has a legitimate basis so long as the curves are analytic and the series upon which the curves are based determine Taylor series. The analogy developed here between classical mechanics and learning

theory is neither superficial nor accidental but is grounded in a common assumption that laws are deterministic because they involve functions that are analytic.

Causal and Teleological Determinism in Physics and Psychology

In classical mechanics, the laws of motion and change may be expressed in either of two ways: as *differential* equations that describe the process as it unfolds over time, or as *integral* equations that describe the complete time course of the process. The differential form of a law assumes that the process unfolds with infinitesimal continuity from a preceding state to its next successive state. For this reason, it is assumed to be expressing the dynamic laws of a system in terms of *efficient* cause (cause and effect). The integral form of the same process, because it defines each intermediate state in terms of the total process, can be seen as expressing the dynamic laws in terms of *final* cause. Mathematically speaking, these two forms of expressing dynamic laws yield identical quantitative results. Furthermore, the two methods are just complementary expressions of the fact that a system's continuous change or motions trace a continuous curve in space over time so that, given any intermediate arc of the curve, the procedure of analytic continuation (in principle, but not always in practice) allows the arc to be extended either backward or forward in time.

This property accounts for the reversibility of the laws of classical mechanics (cf. Overseth, 1969). These laws treat change in state of a mechanical system in terms of an infinite chain of states with infinitesimal links, where the direction of motion of the system with respect to time depends only on whether a past (initial) or future (final) state is specified. The same laws, therefore, permit two opposing philosophical interpretations of strict determinism: causality when the direction of change is fore, and teleology when the direction is aft. Either interpretation is mathematically legitimate because the laws are taken to embody change as an analytic function. Thus, the causal and purposive outlooks on physical reality are but complementary aspects of the mathematics selected to describe strict determinism.

That the philosophically complementary views are a consequence of the mathematics selected rather than the physics employed suggests that analogous consequences should be found in other sciences that (implicitly or explicitly) adopt similar mathematical methods for treating change as strictly deterministic. Classical learning theory offers just such an analogous case.

Classical theories of learning (e.g., those of Hull, Skinner, Thorndike, and Tolman) differed with respect to what conditions are required for learning to occur but tacitly agreed that the learning function can be graphed as a curve in the Cartesian manner. Behind this assumption sits the deeper assumption that learning curves are justified because they are analytic and, therefore, that strict

determinism under the stipulated laws follows. In addition, the laws of learning, whether based on the belief in the necessity of reinforcement or mere continuity, were believed to be universal in two senses: in holding (1) over all contexts, and (2) over all stimuli and responses. Regarding this latter universal property, sometimes referred to as the "assumption of equipotentiality," Bolles (1975) observes:

To be sure, it has always been recognized that some stimuli may be more salient or more vivid or have more ready access to the sensorium than others, and some responses may have had a higher initial rate of occurrence than others, but these considerations were always of minor theoretical importance. They only meant that in a given application *learning might start at a different point on the curve, but all learning followed the same basic curve.* (p. 253, our emphasis)

The implication that various types of learning may be abstractly equivalent in the sense of *following the same curve* clearly supports the contention that the learning series defines some mathematical function. That in spite of differences in application learning might be considered to start at different points *on the same curve* suggests that the function in question has the analytic property required for continuation. If so, then like classical physics, there should be two equally legitimate, mathematically complementary, alternative forms that the laws of learning might assume, each of which expresses a different temporal direction for determinism.

The two contrasting views are of course the *reinforcement* view and the *expectancy* view. Where Thorndike, Hull, Skinner, and other reinforcement theorists sought a system of empirical laws to explain later behavior in terms of past elements in a learning series, Tolman and other expectancy theorists sought laws to explain the association of prior elements in the series in terms of the "purposive" relationships to anticipated (future) elements in the series. Thus, as in the case of classical mechanics, the learning laws also took two forms: a differential (reinforcement) form that is *past pending* in that the future state of a learner depends on its past schedule of reinforcing events, and an integral (expectancy) form that is *future tending* in that the significance of experience is determined by the future usefulness of the expectancies created by this experience. For instance, in a contemporary version of the law of effect, Herrnstein (1979) has used a differential law form to predict the (behavioral) effects of reinforcement.

We have painstakingly drawn the analogy between classical mechanics and classical learning theory partially in order to prepare the way for understanding the current controversy between ecological psychologists with their penchant for law-governed learning, and the cognitive scientists with their rule-governed view of learning. Just as the attempts to give the same sort of rigorous formulation to laws can account for the mathematically complementary but philosophically

opposing views, we propose that the current debate in both physics (see Pattee, 1979, in press) and psychology (e.g., Fodor & Pylyshyn, 1981; Turvey et al., 1981) over rule-governed versus law-governed views of phenomena can be best understood as a continuation of the analogy. This thesis can be supported, and the analogy strengthened, if both physics and psychology not only encounter the same difficulty in adhering to strict determinism as spelled out with analytic precision but seek resolution of the difficulty by drawing on the same formal means. There is reason to believe this is the case.

The Problem of Nonanalytic and Discontinuous Functions

To the extent that dynamic law proved analytic, physicists could ignore the causal and teleological interpretations of law, because on purely quantitative grounds it made no difference. But if one of the two properties required for such functions to be considered analytic (i.e., continuity and convergence) failed to hold, how could the corresponding law be rigorously formulated? If the data series derives from a discontinuous function, how can curves be extrapolated to make either predictions or postdictions? Likewise, if the series is continuous but fails to converge, how can derivatives be taken or integrations formed?

Although the functions dealt with by 17th- and 18th-century physicists in mechanical problems were typically analytic, many nonanalytic functions are encountered in physics. A classic example is the vibrating string. There is no analytic series that allows the V-shaped form of the string when plucked to be continuously mapped on to the other curves the string exhibits during its vibration. (In the mid-18th century, Bernoulli proposed a nonanalytic series of curves to handle this problem.) Likewise, the classical physicists assumed that essentially all physical change might involve *continuous* change and, therefore, that all functions underlying laws of change might be modeled (piecewise) by continuous functions. This turns out not to be the case: Rubber bands stretch and break, beams buckle, ice expands and cracks, and so on. Continuous variation, if pursued too far, inevitably leads to discontinuous changes in state. The 19th-century physicists were well aware of the need to address all such phenomenon but were not equipped with the appropriate mathematical tools for doing so.

Later, it was shown that certain restricted classes of discontinuous curves could be represented by Fourier series. Also, a formal understanding of some discontinuous series not integrable by Newton's method was provided by Riemann and Lebesgue. Contemporary research in differential topology (e.g., catastrophe theory and bifurcation theory) provides even greater understanding of such functions.

The discovery of the limited applicability of analytic function theory to numerous physical changes forced mathematicians to recognize that the unity, symmetry, and completeness of relation between function and series were still wanting.

This forced them to liberalize the denotation of function along lines proposed by Dirichlet (see footnote 1). As argued earlier, adoption of his proposal that a function need not be analytic but might be a rule defined a branch-point in the history of attempts to give rigorous characterization to all functions. Rules may be formally rigorous, but they are not mathematically rigorous in the analytic sense originally intended.

During Dirichlet's era the only analytic forms for functions were the differential and integral forms. Hence, these were the only tools available to model laws. Physics needed laws, however, that went beyond these two forms. Thus, from necessity (the mother of invention) they became willing to entertain, for the first time, rule-defined laws (e.g., the so-called "delta" operators of Kronecker and, later, Dirac). With this innovation, three forms emerged for laws: the differential form, the integral form, and now the (recursive) rule form.

It is not too surprising, therefore, that with the recognition of discontinuities in learning series, such as in learning with prolonged delay of "reinforcement" (Andrews & Braveman, 1975; Garcia, Ervin, & Koelling, 1966), psychologists, like physicists, should also entertain nonanalytic characterizations for functions and should realize that learning rules should vie for dominance over learning laws. What analytic alternative was there?

Actually, there is an analytic alternative by which certain discontinuities in series might be modeled in a lawful manner without invoking nonanalytic rules; or, at least, by which such a breach in tradition might be forestalled: the integro-differential form compounded from the other two analytic forms; that is, perhaps laws should be considered as functionals rather than functions. This was Volterra's strategy, and the one we are endorsing. Whether it will prove to be only a means for postponing the inevitable remains to be seen, but at least the strategy deserves serious consideration because the rule strategy aborts the search for laws in favor of the search for mechanism. These are not equivalent theoretical goals, for rules require embodiments whereas laws do not (Feynman, 1965). The search for the mechanistic embodiments of laws has been fruitless in the extreme, as witness the history of failures to find convincing evidence for the mechanistic embodiments of laws of the electromagnetic field (ether), combustion (phlogiston), and motion (impetus).

APPENDIX B

Nonlinearity in the Learning Functional

Two independent variables are nonlinearly related if the course of values of one is constrained by the course of values assumed by the other; that is, of one of them appears in the argument of some more general function, whereas the other appears as a value in the range of that more general function. The subtlety is that whereas the pair of functions may be linear, the functional that relates them

may be either linear or nonlinear. For instance, each value of k plays a linear role in determining the magnitude of the stress (restorative force), x , for a given strain, y , of a spring, and each stress value (x) plays a linear role in determining the magnitude of the strain. This is in accordance with the linear form of Hooke's law where k is treated as a constant because it is assumed to be independent of strain (y). Still, over the history of its use the strain variables (frequency and magnitude of strain) constituting y determine a nonlinear influence on the future course of k values and require, therefore, a generalized form of Hooke's law that is nonlinear. Thus, two linear functions may add up to a nonlinear functional!

We expect learning to be equally complex (i.e., a nonlinear hereditary functional). In fact, Grossberg (1980) has found it necessary to use nonlinear functionals in the design of neural nets that exhibit interesting forms of cooperativity and competition: a kind of learning. Because the hereditary influence (K) over learning series may not be linear, the learning functional (y) probably requires a hereditary law that is the sum of a linear function x and a nonlinear function K .

A simple test of linearity requires that

$$kx + K - y = 0. \quad (5)$$

If the learning-to-learn (or fatigue) effects introduce a nonlinear dissipative parameter such that

$$kx + K - y = d > 0 \quad (\text{or } d < 0 \text{ for fatigue}),$$

then the coefficient d provides a measure of the *linear deficiency* of the equation for the learning functional. Under such circumstances, the learning series would correspond to a *nonlinear* Volterra series defined over a set of simultaneous, multiple integral equations:

$$v(t) = kx(t) + \sum_{m=1}^i \int_0^t \cdots \int_0^t \underbrace{K(t, s_1)x(s_1)ds_1 \cdots K(t, s_m)x(s_m)ds_m}_{m \text{ times}} \quad (6)$$

In abbreviated operator notation Eq. 6 is represented by:

$$v = kx + K_1 + K_2 + \cdots + K_m. \quad (7)$$

From a computational viewpoint, this means that after solving for the learning function v on trial $n < m$ in a series of m trials, the entire computation must be repeated if the learning effect on $n + 1$ trials is desired. Hence, to determine the

hereditary influence of K_m , one must solve for $K_1, K_1 + K_2, \dots, K_1 + \dots + K_{m-1}$. Little wonder that the laws of hereditary mechanics and learning may be difficult to formulate, their corresponding functionals hard to express, and the exact nature of the underlying data series as discretely sampled not readily revealing.

Furthermore, the theory of nonlinear integro-differential equations is still incomplete and no general approach yet exists. However, approaches do exist for determining the integrability of simple nonlinear series (e.g., Volterra series). Nonlinear Volterra series have been required for adequate characterization of a wide range of systems that exhibit hereditary effects, including pupillary reflexes (Sandberg & Stark, 1968), irradiated tissues (Iberall, 1967), competing populations of neural cells (Grossberg, 1980), and automatic control systems (e.g., Tsytkin, 1973). Consequently, one should not be surprised that so ubiquitous a strategy might emerge again in human and animal learning. If so, then psychologists, like the physicists, would be wise to incorporate integro-differential equations into their theories of learning and to admit openly allegiance to hereditary mechanics.

REFERENCES

- Andrews, E. A., & Braveman, N. S. (1975). The combined effects of dosage level and interstimulus interval on the formation of one-trial poison-based aversions in rats. *Animal Learning and Behavior*, 3, 287-289.
- Bolles, R. C. (1975). *Learning theory*. New York: Holt, Rinehart, & Winston.
- Bransford, J. D., Stein, B. S., Shelton, T. S., & Owings, R. A. (1981). Cognition and adaptation: The importance of learning to learn. In J. H. Harvey (Ed.), *Cognition, social behavior, and the environment*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- d'Abro, A. (1951). *The rise of the new physics: Its mathematical and physical theories* (2nd ed.). New York: Dover.
- Feynman, R. (1965). *The character of physical laws*. Cambridge, MA: MIT Press.
- Fodor, J. A., & Pylyshyn, Z. (1981). How direct is visual perception? Some reflections on Gibson's "Ecological Approach." *Cognition*, 9, 139-196.
- Garcia, J., Ervin, F. R., & Koelling, R. A. (1966). Learning with prolonged delay of reinforcement. *Psychonomic Science*, 5, 121-122.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Greeno, J. G. (1974). Representation of learning as discrete transition in a finite state space. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 1): *Learning, memory, and thinking*. San Francisco: W. H. Freeman.
- Grossberg, S. (1980). Biological competition: Decision rules, pattern formation, and oscillations. *Proceedings of the National Academy of Science*, 77, 2338-2342.
- Harlow, H. F. (1949). The formation of learning sets. *Psychological Review*, 56, 51-65.
- Herrnstein, R. J. (1979). Derivatives of matching. *Psychological Review*, 86, 486-495.
- Hinde, R. A., & Stevenson-Hinde, J. (Eds.). (1973). *Constraints on learning*. New York: Academic Press.
- Iberall, A. S. (1967). Quantitative modelling of the physiological factors in radiation lethality. *Annals of the New York Academy of Sciences*, 147, 1-81.
- Jammer, M. (1974). *The philosophy of quantum mechanics*. New York: Wiley.
- Johnston, T. D. (1981). Contrasting approaches to a theory of learning. *Behavioral and Brain Sciences*, 4, 125-173.
- Kline, M. (1972). *Mathematical thought from ancient to modern times*. New York: Oxford University Press.
- Kling, J. W., & Riggs, L. A. (Eds.). (1971). *Experimental psychology* (3rd ed.). New York: Holt, Rinehart, & Winston.
- Koffka, K. (1935). *Principles of gestalt psychology*. New York: Harcourt, Brace.
- Kramer, E. E. (1970). *The nature and growth of modern mathematics*. New York: Hawthorn.
- Krechevsky, I. (1938). The study of continuity of the problem-solving process. *Psychological Review*, 45, 107-133.
- Kugler, P. N., Turvey, M. T., Carello, C., & Shaw, R. (in press). The physics of controlled collisions: A reverie about locomotion. In W. H. Warren Jr. & R. E. Shaw (Eds.), *Persistence and change*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lagrange, J.-L. (1788). *Mécanique analytique*. Paris: Chez la Veuve Desaint.
- Lautmann, A. (1971). Symmetry and dissymmetry in mathematics and physics. In F. LeLionnais (Ed.), *Great currents of mathematical thought* (Vol. 1). New York: Dover.
- Lindsay, R. B., & Margenau, H. (1957). *Foundations of physics*. New York: Dover.
- Nagrath, I. J., & Gopal, M. (1982). *Control systems engineering*. New York: Wiley.
- Overseth, O. E. (1969). Experiments in time reversal. *Scientific American*, 221(4), 88-101.
- Pattee, H. H. (1979). Complementarity vs. reduction as explanation of biological complexity. *American Journal of Physiology*, 236(5), R241-R246.
- Pattee, H. H. (in press). Cell psychology: An evolutionary approach to the symbol-matter problem. *Cognition and Brain Research*.
- Patten, B. C. (1982). Environments: Relativistic elementary particles for ecology. *American Naturalist*, 119, 179-219.
- Picard, E. (1907). La mécanique classique et ses approximations successives. *Revista di scienza*, 1, 4-15.
- Restle, F. (1965). Significance of all-or-none learning. *Psychological Bulletin*, 64, 313-325.
- Sandberg, A., & Stark, L. (1968). Functional analysis of pupil nonlinearization. In H. L. Oestreicher & D. R. Moore (Eds.), *Cybernetic problems in bionics*. New York: Gordon & Breach.
- Seligman, M. E. P., & Hager, J. L. (Eds.). (1972). *Biological boundaries of learning*. New York: Appleton-Century-Crofts.
- Shaw, R. E., & Mingolla, E. (1982). Ecologizing world graphs. *Behavioral and Brain Sciences*, 5, 648-650.
- Shaw, R. E., & Todd, J. (1980). Abstract machine theory and direct perception. *Behavioral and Brain Sciences*, 3, 400-401.
- Shaw, R. E., & Turvey, M. T. (1981). Coalitions as models for ecosystems: A realist perspective on perceptual organization. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Shaw, R. E., Turvey, M. T., & Mace, W. (1982). Ecological psychology: The consequence of a commitment to realism. In W. Weimer & D. S. Palermo (Eds.), *Cognition and the symbolic processes II*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Tsytkin, Ya. Z. (1973). *Foundations of the theory of learning systems*. New York: Academic Press.
- Turvey, M. T., & Shaw, R. E. (1979). The primacy of perceiving: An ecological reformulation of perception for understanding memory. In L.-G. Nilsson (Eds.), *Perspectives on memory research*. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Turvey, M. T., Shaw, R. E., & Mace, W. M. (1977). Issues in the theory of action: Degrees of freedom, coordinative structures and coalitions. In J. Requin (Ed.), *Attention and performance* (Vol. 7). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Turvey, M. T., Shaw, R. E., Reed, E. S., & Mace, W. M. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn. *Cognition*, 9, 237-304.