D. Stalker (Ed.), Grue! The New Riddle of Induction. Chicago, IL: Open Court, 1994, pp. 263-272.

Learning and Projectibility

14

Patrick Suppes

It is surprising, and perhaps a reflection of a certain provincialism in philosophy, that the problem of induction is so seldom linked to learning. On the face of it, an animal in a changing environment faces problems no different in general principle from those that we as ordinary humans or as specialized scientists face in trying to make predictions about the future. But requests for a solution to the problem of induction, I would very much agree with Goodman, are usually formulated in ways that are inappropriate. A purely deductive justification is surely out of the question. We might as well ask for a logical solution to the problem of surviving forever.

The two basic ideas I will explore here are these. First, much of learning is nonverbal, and consequently a statement-type formulation of inductive problems is inappropriate. Second, scientists are no different from other mammals trying to survive in the wilderness in terms of the basic problems of projectibility, or put in another standard way, prediction. I do not mean to say that scientists do not have better methodologies in many cases, but that the fundamental problem is little changed when we confront such troublesome problems as prediction of the weather, of earthquakes, of ozone depletion, and the like, compared to what must have been a relatively stationary environment from the standpoint of dinosaurs, for instance, for many millions of years. I am not suggesting that dinosaurs were better predictors, but that the enormous advantage of a relatively stationary environment, from the standpoint of the features in which one is interested, is much more important than scientific methodology. If one is interested in predicting and using features of the world that are highly nonstationary then more than methodology, mainly a lot of good luck, is needed to do very well at prediction. In the worse kind of environment, no significant event, good or bad, has a high probability, and the best we can do is try to optimize, on the basis of fallible ideas of projectibility, prediction of the most likely, even if improbable, occurrence. That induction should deal with the probable is one of the most common mistakes of philosophical discussion. For a hungry animal finding a delectable prey may be the most improbable of events and there may be no prey, delectable or not, that has a high probability of being caught. Animals in such an environment learn to be persistent, patient, and attentive to very subtle perceptual cues if they are to survive. Stationarity, exchangeability, and many other assumptions of much modern statistics do not hold for environments of this kind.

I have organized my analysis into four sections: animal learning, human learning, experimental science, and induction revisited.

1. Animal Learning

I include in animal learning much human learning because at all levels of animal behavior, nonverbal learning is central. It is the only kind of learning for almost all animals, but its essential presence in human learning cannot be emphasized too strongly. For the child learning to walk downstairs, to recognize familiar faces, or to engage in a thousand other developmental projects of learning, it is evident enough, but later the adult walking, running, lifting, grasping, etc., also requires a range of nonverbal skills. It is nonverbal skills that we are very poor at verbalizing at all. Try describing with any physical accuracy how you write your name, hit a backhand in tennis, or walk into a room without falling over chairs.

Artificial predicates like *grue* and *grund* are no problems for animals, because they are not considered, but projectibility is. Much learning occurs in relation to features of the environment that are partly stationary, and those stationary features have nice properties of projection. What is important about learning, however, is dynamic adaptability. Animals that can only survive in a rigidly stationary world will not make it long in most places. The adaptability of learning is in the sharpest possible contrast to most conceptions of scientific experimentation under sharply defined circumstances. Examples of this adaptability are to be seen in every aspect of behavior, of course in hunting prey or avoiding becoming prey, but also in just the mundane task of getting around in the world. An animal that is raised initially in an environment of physical objects in fixed positions, as for example in a cage or a maze, can have considerable difficulty initially adapting to what we would call a wilderness environment.

For animals that want to survive, there is a problem of projectibility, but it is not exactly Goodman's problem. Can what they have learned in the past be flexible enough to permit them to adapt to a changing environment in the future? The answer is, of course, sometimes *yes* and sometimes *no*. Moreover, at the heart of the problems of projectibility for animals in a natural environment is the shortage of the pretty symmetries dear to the hearts of statisticians. For example, the symmetry of exchangeability pressed into such good service by de Finetti and other Bayesians is one that will not work well in a natural environment. In fact, it is generally a criticism of learning models if they satisfy a strong exchangeability assumption because they are not sensitive to new developments. One way of putting the matter is they do not adapt to the current moving average. Another is that the learning operators are commutative.

2. Human Learning

Much of what I have said about animal learning applies to nonverbal human learning. Philosophers through the ages have been beguiled by our ability to talk and reason with words and have not recognized how much of human behavior is inaccessible to accurate verbal description in ordinary circumstances. It is of the greatest importance to recognize that walking, talking, etc., cannot be adequately described by ordinary verbal concepts, and even with the best scientific apparatus only partially. There is an absence of symmetries, an absence of explicit theory to guide experiment, and a paucity of adequate scientific knowledge even now. In the meantime, learning rolls along, robust and dynamic, providing us like other animals with a host of ways of dealing with the world as long as it is not too radically changed. Moreover, this absence of anything like accurate verbal description excludes the artificial problems of *grue* and *grund,* as in the case of other animals, from even arising for this whole domain of experience.

The absence of symmetries of a serious kind in our dealings with the ordinary world is evident in our ordinary descriptions of spatial and temporal phenomena (for more extended discussion of this point see Suppes 1991). It is empty space, not space filled with the bric-a-brac of ordinary existence, that exhibits the symmetries so important in the history of modern geometry. Abstraction and simplification in other areas of experience are similarly required to move from the framework of ordinary language from which symmetries are mainly lacking to obtain any of the symmetries fundamental to modern science.

3. Experimental Science

The ideal environment for projectibility, as has already been emphasized, is one that is stationary. This same ideal, and for closely connected reasons, holds for the design and execution of experiments in any domain of science. The Bayesian who wants to apply exchangeability in the spirit of de Finetti's fundamental theorem can only do so if he is persuaded that the experimental environment is one of stationarity. De Finetti's well-known lectures (1937/1964) are aimed at prevision, or what we could term in the present context, a form of projectibility. Just as exchangeability is an assumption stronger than stationarity—for example, stationary Markov chains can exhibit partial exchangeability, but not exchangeability—so the Bayesian, operating with conditioning as a form of learning with commuting operator properties based on the assumption of exchangeability, does not adapt well to changing environments.

It is important, however, to emphasize that the methodology implied by stationarity assumptions is not necessary for good experimental work. It is, for example, exactly the study of nonstationary processes that is characteristic of learning and of problems of control and adaptability. Moreover, there are a variety of standard statistical methods for analyzing the fit of a given learning model to given experimental data.

On the other hand, it should be emphasized that in many learning situations, but not all, one expects the learning and the associated learning models or theories to be ergodic in the following sense. The stochastic process that embodies the learning model has a unique stationary distribution, and it is the asymptotic distribution of the process under any initial distribution. Put another way, we expect most learning models to have the property that the initial probabilities of responding are not the same as the final ones, but the asymptotic stationary distribution of responses is independent of whatever distribution of responses we may start with. To achieve the ergodicity expected, the learning process need not be a finite Markov process, but it should have a distance-diminishing property that corresponds, roughly speaking, to a geometric fading away of the influence of the past (for technical details of such conditions and how they lead to a proof of ergodicity I just defined, see Lamperti and Suppes 1959). An important class of learning models that do not have such an ergodic property is studied in Karlin (1953). Organisms that for one reason or another have a genetic disposition or an environment that makes them behave in a way to satisfy nonergodic constraints will inevitably have a limited range of adaptability.

A wide class of phenomena, much studied experimentally but ordinarily not under the heading of learning, is the biological development of species and the modification of that development by various traditional and modern methods of changing the genetic structure. However, in a broader context and certainly in the context of modern theories of learning, the Darwinian model of survival is not only a kind of learning associated with species but has also been converted into a standard class of models of learning in artificial intelligence. These models, like other learning models, are not stationary processes except in very special cases.

Two fundamental aspects of learning, present in everyday experience and much studied experimentally, are transfer and generalization. Learning one task and then another can in many cases lead to positive or negative transfer. A well-known anecdote illustrating negative transfer is that of the Southern mules who, having been taught to plow up the weeds growing between the rows of young corn or cotton plants, could not then be taught to plow up the plants as required after World War I by government measures to deal with an agricultural surplus. Transfer is negative when prior learning gets in the way of new learning. There is truth in the adage that you cannot teach an old dog new tricks.

Similar problems arise for generalization. An intense focus on a narrow range of problems can make subsequent generalization of the same skills to new problems difficult. But without generalization reasonable behavior in a natural environment would be nearly impossible. Positive generalization and transfer are forms of projectibility required for survival.

More important than asymptotic stationarity for a learner, or a mathematical model of the learner, is adaptability, which is the essence of transfer and represents a further move from stationarity. It is why learning models that one way or another use estimators of parameters that are like good statistical estimators for stationary environments often do poorly in fitting transfer data. For example, the maximum-likelihood estimate of a numerical parameter for stationary processes is just the mean of the number of observations, little changing after a large number of observations. But this ever slower rate of change over time can be disastrous for adaptability. Learning models aimed at experiments to study initial learning but not transfer often exhibit this failure, probably much more often than the animals they are meant to model.

The experimenter as learner.

What I have said earlier about experimentation sounds much too formal and theoretical. We run carefully designed experiments and then apply carefully thought-out statistical analysis to see what the quality of the learning is, and especially to see how well the actual data of learning fit our theoretical model. But there is another sense in which learning is fundamental to experimentation and is of quite another sort. This is the learning that takes place on the part of the experimenter. For any experiments of any complexity, there are a number of technical things that must be done, ranging from extremely elaborate technological preparation in the case of modern physics experiments, to learning experiments with humans that involve carefully thought-out instructions and psychological settings. Whether it is a matter of physics or psychology, the activities of the experimenter cannot be described with any completeness and corresponding accuracy in ordinary language. Moreover, the detailed execution of the physical actions required of the experimenter are to a large extent not under conscious control. The physicist handling delicate equipment cannot describe the subtle manner in which his perceptual and motor control systems operate to do just what he wants them to. The psychologist instructing subjects does not have under his own explicit conscious command the variety of cognitive and emotional cues that surround the handling of experimental subjects. The most unmanageable and most unruly part of experimental science is the fact that the learning of experimenters is not itself an easy subject for experimental study.

Scientific reports of experiments are like sports reporting, unintelligible to the uninitiated and sketchy and incomplete in description. Moreover, much of this understood background knowledge is not codified or written down somewhere else that can be looked up as occasion demands. It is rather like looking up the rules of chess and being disappointed that they tell you very little about how to play the game. No philosopher who holds some strong propositional theory of knowledge or is much caught up with the content of thought has yet had anything very interesting or subtle to say about complex scientific experiments and how experimenters learn to conduct them. I claim to understand why this is so. Too much of the experimenter's essential behavior cannot be accurately described with verbal concepts currently in use.

4. Induction Revisited

Much that Goodman (1955) has to say about induction and his new riddle continues to seem eminently sensible and pertinent. I especially agree with him that the traditional view has been that the central problem of induction is justification. He says, for instance,

> The typical writer begins by insisting that some way of justifying predictions must be found; proceeds to argue that for this purpose we need some resounding universal law of the Uniformity of Nature, and then inquires how this universal principle itself can be justified (p. 65).

(The ghost of John Stuart Mill is rightly stalked in this passage.)

Validity.

Goodman goes on to ask what we mean by a valid prediction. He says:

> predictions are justified if they conform to valid canons of induction; and the canons are valid if they accurately codify accepted inductive practice (p. 67).

In spite of my sympathy for much that Goodman has to say in expanding on the passage just quoted, I am skeptical that we can hope to find a unique sense of validity applicable to all environments in which we or others make predictions. First, as already stressed, I take the case of animal behavior to be important, but I find it very awkward to talk about valid animal predictions. Certainly it is uncontroversial to talk about animals making *correct* predictions when hunting prey or being hunted, but correctness, unlike validity, is in this context purely result, not rule, oriented.

Second, there is the Bayesian answer that there can in general be no universal criterion of validity, for inductive inferences or behavior are for a good Bayesian first of all a subjective matter, and we can hope for intersubjective convergence of belief or inference only in certain restricted classes of cases about which I have more to say below. Although I am also skeptical of the range of the positive Bayesian program, which I consider below, I am wholly sympathetic with this negative point. Third, Goodman himself gives a good argument for skepticism when he later stresses the importance of the background and context within which predictions are made.

Role of language.

I am also skeptical of the central role Goodman assigns to language and his claim that inductive validity arises from regularities which are a function of our linguistic practice (p. 117). Why I find this idea hard to accept is clear from what I have already said. Too much learning even by humans is utterly nonverbal in character and yet in much of our ordinary getting about in the world it is inductively highly successful.

Bayesian projectibility.

In another article in this volume, Brian Skyrms sets forth a clear and persuasive case for a Bayesian concept of projectibility. As might be expected, he stresses stationarity and the symmetries of various kinds of exchangeability. In the final paragraph, he has this to say:

> We see that—at increasingly general levels of abstraction—projectibility is a reflection of probabilistic symmetries. The logic of projectibility is probability logic. At the most abstract and general level, it is the mathematics of probability-symmetry structures.

I applaud such symmetries when they can be found or, more likely, be imposed by experiment. But such a regime of experimentation occupies a very special place in the world, not even satisfied by many experiments in physics, and certainly not at all in important nonexperimental sciences such as astrophysics, geology, and meteorology. (Skyrms rightly points out the great difficulty of proving the ergodicity of any natural physical systems.)

If animals of any species had had to count on such symmetries to develop workable projectible concepts, evolution would in all likelihood have never gotten off the ground. It is the development of robust learning mechanisms and ever better evolutionary features, which work well enough in environments without such symmetries, above all without stationarity, that lead to projectible concepts good enough to facilitate survival.

Learning and lawlikeness.

Much philosophical talk about induction has been aimed at characterizing genuine as opposed to not so genuine laws, but it seems to me there is good reason to be skeptical of the outcome of this effort—if only because of the negative results thus far. Goodman gives Hume his due for suggesting a mechanism, that of custom or habit, for making predictions. What Hume has to say in Book I of the *Treatise* (1739/1951), especially in Part III, is a brilliant early contribution to the theory of learning. His sketches of how the mind works are directed toward the acquisition of knowledge. No doubt he assumes an environment that is too stationary, but that is hardly surprising.

The central point I want to make by invoking Hume is that he was even more on the right track than Goodman. Hume replaced the necessary connections often thought essential for knowledge by psychological mechanisms of learning. As Hume emphasizes, learning is mostly unconscious and unreflective. We do not think before we walk or talk. We think while we are walking or talking, but ordinarily not about the walking or talking.

So what I urge is to change the focus of projectibility from concepts and laws to the more fundamental mechanisms of learning. I turn to the informal description of some results along the line suggested for various learning models.

Measuring projectibility.

There is no direct way to determine an optimal measure of projectibility of a given learning model relative to a set of tasks and a set of possible environments, also possibly changing from time to time. There is, in principle, a Bayesian approach to the problem of choosing an optimal measure, but it is too intricate to consider at this preliminary stage. Some Bayesian aspects of the problem are mentioned below. What is important is that for nonstationary environments and problems or tasks involving transfer it is inappropriate to use as a measure of success the cumulative relative frequency of success, a measure that accommodates badly to change, as already emphasized.

Let L be a finite nonempty set of learning models whose performance on a finite sequence of prediction tasks $\tau_1, \ldots, \tau_N$ we evaluate for projectibility using the following recursive linear model. For every model i in L, let the projectibility measure $p_{i,1}$, i.e., at the beginning of trial 1 be 0.5—this choice of a particular number between 0 and 1 is a parameter that can be changed. Then for any learning model i in L the recursive rule is:

$$p_{i,n+1} = \begin{cases} (1-\Theta)\, p_{i,n} + \Theta & \text{if a correct prediction was made on trial } n, \\ (1-\Theta)\, p_{i,n} & \text{otherwise.} \end{cases}$$

Again, the particular number Θ is a projectibility parameter which may be changed as long as it lies between 0 and 1. (A good Bayesian problem is to derive optimal estimates of $p_{i,1}$, and Θ.)

As in good horse races, we do not expect one learning model to have the largest projectibility measure for all trials. In extensive but as yet unpublished studies of a variety of neural net learning models, Lin Liang and I have found that a multivariate normal learning model, i.e., a model which recursively estimates a multivariate normal classification model on each trial, does very well on a variety of classification tasks, but only after a substantial number of trials. At the beginning and in the short run several other models formulated only in terms of individual feature inputs or a few combinations often do better, i.e., have a higher projectibility measure for an initial segment of trials. Something similar happens after a large number of trials on a given classification task. Transfer to a new task is often poor because of the relative frequency estimates of feature means relative to each class of the classification scheme.

Note that $p_{i,n}$ is meant to give a reasonable estimate of the probability that the prediction of model i on trial n is correct. We can obviously refine the measure to correctness for a given response. For example, a model may have little difficulty in predicting when an observed "object" belongs to class α, but great difficulty discriminating between objects in classes β and γ.

It may seem to some that I have, inadvertently or not, abandoned the problem of giving a principled account of projectibility for the weak solution of a purely phenomenological quantitative measure. It has been a deliberate advertent choice. A measure just of the blind success of predictions is limited, but here the measure is tied to individual mechanisms of learning. Beyond such mechanisms there is no meaningful concept of validity of predictions or deeper concepts of justification of induction. As Hume rightly said most learning in ordinary experience is unconscious and unreflective. There can be no logical or transcendental justification of the inherited biological mechanisms of learning beyond some measure of their success. Exactly which measure is a good topic for constructive academic dispute.

Furthermore, I would argue that the same is to be said for the systematic methods of statisticians, Bayesians or otherwise. The contribution of these methods to the design and evaluation of experiments in many scientific domains is one of the signal methodological achievements of this century, but their ultimate justification is at bottom very similar. The pragmatism of success is not replaceable by a religion of justification, not even by theistic invocations of the uniformity of nature.

Stanford University

REFERENCES

De Finetti, B. (1937/1964). "Foresight: Its Logical Laws, Its Subjective Sources," in *Studies in Subjective Probability,* ed. Henry E. Kyburg and Howard E. Smokler (New York: J. Wiley).

Goodman, Nelson. (1955). *Fact, Fiction and Forecast.* (Cambridge, Massachusetts: Harvard University Press).

Hume, D. (1739/1951). *A Treatise of Human Nature,* ed. L. A. Selby-Bigge (Oxford and New York: Oxford University Press).

Karlin, S. (1953). "Some Random Walks Arising in Learning Models I." *Pacific Journal of Mathematics 3:* 725–756.

Lamperti, J. and Suppes, P. (1959). "Chains of Infinite Order and their Application to Learning Theory." *Pacific Journal of Mathematics 9:* 739–754

Skyrms, Brian. "Bayesian Projectibility." This volume.

Suppes, P. (1991). "The Principle of Invariance with Special Reference to Perception," in J. Doignon and J. Falmagne, eds., *Mathematical Psychology: Current Developments* (New York: Springer), 35–53.