This paper contends that human organisms are not qualitatively different from infra-human ones. The same principles apply to the acquisition, maintenance and weakening of complex verbal skills as mathematics as to rats pressing bars or pigeons pecking windows. Two aspects of child research are assessment and training. Assessment involves cross-sectional sampling of the child's activities, while training involves more longitudinal repeated presentation of a task and its counterpoints. The focus of this paper is training. There are many systematic positions that attempt to account for the products of learning operations. The author's position leans entirely toward associative conditioning of external (rather than internal, movement-produced) stimuli with behavior. Drive reduction is not sufficient although it seems to be a sub-set solution to the overall problem of reinforcement--feeding the deprived rat works, but probably for reasons other than drive reduction. The terminal response is a given situation is that recurs when a large enough proportion of the original cue is represented. (Author/HMV)
GENERALIZATION AND TRAINING

CENTER FOR URBAN EDUCATION

W. O. JENKINS

APRIL 1, 1966
April 1, 1966

GENERALIZATION AND TRAINING

Notes on the role of generalization and other matters in learning and training experiments with children.

W. O. Jenkins

Center for Urban Education

There are old saws floating around the culture and even extant in psychological circles that man is different from beast and that human organisms are qualitatively different from infra-human ones. This is not so. Even the quantitative argument that man is inherently more complex - can do more things - is open to question. Witness Skinner's (1960) war-time employment of pigeons to guide missiles more accurately than did people, and Verhave's (1966) recent use of them as quality control inspectors. The fact of the matter is that the same principles apply to the acquisition, maintenance and weakening of complex verbal skills as mathematics as to rats pressing bars or pigeons pecking windows. The difference lies, not in principles, but in complexities of measurement. One confounding matter here is that investigators change models like models change clothes when they go from relatively simple organisms like the rat to people. For parsimony's sake (if no other) it seems reasonable to push one model to its limits before fading it out for a different one.

There is quite a different point relating to the utility as well as the versatility of infra-human organisms. We might be interested in the clinical problem of the female who reacts to all adult males as if they were "father". She avoids males and reports an increased "anxiety" level when in their presence. It turns out that during her developmental years her
father was the All-American business man who had little time for his children and was so actively engrossed in other matters that he almost completely withheld attention and affection from them. This situation immediately suggests a systematic research program that can be initiated in the laboratory, namely, the relationship between various kinds of deprivation and generalization of behavior across a wide stimulus spectrum. We began the program by investigating the effects of food deprivation in the pigeon on generalization of pecking behavior to a range of visual stimulus changes (Jenkins, et al 1958). The results were dramatic in showing increased generalization with increased deprivation. The next step in the program was to stay in the laboratory with human subjects and proceed to deprive them of attention (the investigators' controlled behavior) and increase their anxiety level to different degrees by varying intensity of electric shock in the presence of certain kinds of people, peers, younger and older. This conditioning would be followed by generalization tests with the several people stimuli. The potential extrapolation of this research to general stimulus and economic deprivation is obvious. For example, does economic and cultural deprivation produce increased prejudice? These are experimental matters.

Another case in point happened, accidentally, to deal with anti-impulsivity training in pigeons (Jenkins, 1955). It was an experimental exploration of generalization across conditioning and extinction in which conditioning was made as much like extinction as possible. The study is an analogue to teaching patience, reflectivity or non-impulsivity in a human organism. Since one of the main characteristics of extinction is infrequent responding, it was necessary to build in this behavior in
conditioning. It was not easy because the pecking response in the pigeon is a ballistic-type movement involving throwing the head forward, making contact with the pecking window so that the head is thrown back against the neck muscles and thus forward again and so forth. The response can occur a number of times a second. In any event, the first response or two were followed by food presentation and then the birds were taught to pause between pecks. At first the pause was brief, of the order of a couple of seconds. If a response occurred during the pause interval it went unrewarded. Only responses after the specified interval were followed by food presentation. By a kind of "crowding-the-threshold", the interval of non-responding was increased to the order of 30 seconds. (There were other experimental refinements that need not concern us here.) The treatment had the desired effect, so that 30 hours after the extinction proceedings were instituted, the birds were still responding at about one-third of their original rate as contrasted with the usual near operant level after a handful of hours. Attempts at suppressing behavior in this fashion were conducted on a large scale varying motivation (food deprivation) and certain characteristics of the external environment.

These small examples capitalize the efficacy of a back-and-forth give and take between the laboratory and the couch or the classroom.

To return to the main theme of measurement, admittedly, measuring thinking or creative behavior is a tougher problem than recording errors in maze learning, but it is not insurmountable. In any case the first stage of scientific operation is to delineate the phenomenon under experimental scrutiny, i.e., measure it. We may pick an insensitive or even inappropriate measure, e.g., looking at the feet or hair texture in
studying the common cold, but measurements must be continued until the experimenter is satisfied that the phenomenon is pinned down for his purposes - and he may be wrong or have flipped the coin so it stood on edge. Across the board some stab must be made at a working definition of the antecedents and consequents before the experimental program can be started -- even one of the fishing expedition variety.

Tying in with this point is another straw man, namely, the intriguing and implausible state of affairs that we have more basic information about our laboratory rats (and maybe even dolphins) than we do about our children. It would seem far easier to teach a rat to bar-press than a child to work with numbers. It is a reasonable assumption that most behavioral scientists are not interested in rats per se, but rather in principles of behavior. Experimental extrapolation seems easy, but has not been accomplished on a large, systematic scale. One drawback is lack of accessibility of subjects characteristic of administrative red tape due to lack of education as to the nature of research. Other problems exist. Society, including parents, place rigid limits on the kinds of experimental treatments to be applied to their children and to themselves. A plausible hypothesis on the behavior complexity matter is that parameter will be added but not multiplied with phylogenetically more complicated organisms than the rat and pigeon. This is an experimental question.

Returning to the main stream, the two obvious aspects of child research are assessment and training. These shade over into one another, but can be separated for working purposes. Assessment is cross-sectional sampling of the child's activities while training is the more longitudinal
matter of repeated presentation of a task and its accouterments. Assessment has been briefly touched on elsewhere (Jenkins, 1966) and the focus of this note is training.

The paradigm for training experiments is straightforward. A stimulus situation is presented that sets the stage for a certain class of behavior (including "motivation"), the presence or absence (or degree) of the specified behavior is recorded, stimuli terminating the sequence are introduced ("reinforcers") and the whole set-up is repeated one or more times. These are the classic operations for any behavioral change study. If the behavioral change is in the direction of increased response strength, it is classed as "learning". Thus, we deprive the rat of food, smear wet mash on it, he presses and gets a pellet and then returns to repeatedly pressing the bar. We show the child a ball and say "ball", the child says "baw", we say "good", "ball" and thus shape up the word behavior.

The situation is obviously more complex than this and the details will be anchored in later paragraphs. Experimentally in the simplest case, two matched groups are selected on one or more dimensions (diminishing returns are rapidly reached) presumably related to the task under consideration, one is given training, the other something else and a test of acquisition or effect of training applied to both. Far more elaborate designs may be employed involving special groups without the pre-test and other variations on the theme, several treatments may be applied "simultaneously", and so forth. The essence of the matter is selection - treatment - criterion.

In passing, it should be noted that all learning or training
experiments reduce in bedrock to transfer of training investigations. Our organisms are not blank sheets of paper on which experimental experience writes. Rather they bring a rich supply of behavior to the situation in which we measure and observe them. The basic question is what behavior classes are in their repertoire at the time of testing and how these available responses (and their associated stimuli) relate to the corresponding items of the experimental set-up. In brief, the point is the extent to which the store of behaviors competes or interferes with the learning to be acquired or the training to be applied. Facilitation obviously occurs in some instances, but the more common event when the subject is exposed to a relatively novel situation, is absence of behaviors to similar stimuli or the occurrence of at hand competing responses.

The present note deals with the propaedeutic role of generalization-stimulus change and constancy and degree of response decrement in learning and training investigations.

A Little Theory

There are many systematic positions that attempt to account for the products of learning operations ranging from the broad sweep of the Tolmanian view (1949) of six factors and their combinations (six factorial) that probably covers more principles than there are facts to the relatively stripped down positions of Hull's (1943) drive reductionism and Guthrie's (1935) kinesthetic contiguity.

The present position is a contiguity one, but unlike Guthrie's, leans entirely on associative conditioning of external (rather than internal, movement-produced) stimuli with behavior. Drive reduction is not
enough although it seems to be a sub-set solution to the overall problem of reinforcement: feeding the deprived rat works, but probably for reasons other than drive reduction.

The main point seems to be that the terminal response in a given situation is the one that recurs when a large enough proportion of the original cue situation is re-presented. Identification of the terminal response is not always simple. When rats are dropped through a trap door at the end of a runway, they will exhibit increased running speed on later occasions if they were approaching when dropped. If they were backing up, they will show progressively increasing avoidance on later exposure. The implications of this position are many and obvious.

The basic ingredients of the present position are four:

1. **Associative Conditioning.** A set of cues present on a given occasion is conditioned to the last response made in their presence.

2. **Terminal Response.** Re-presentation of the stimulus compound (or an empirically determined major portion of it) brings about the same response that was terminal or postreme on the previous occasion.

3. **Removal.** A stimulus presented after a response that alters the original cue situation and thereby produces a change in behavior serves to leave the original response terminally associated with the original stimuli that accompanied it ("reinforcement").

4. **Cue Constancy.** The strength of a response on a given occasion is a direct function of the percentage of the
total stimulus compound currently present that was previously associated with that response.

Simple as these principles sound, complicated behavioral deductions follow from them (for details, see Jenkins, 1955). We will talk to the last one, traditionally known as generalization, in greater detail throughout this note.

Empirically, these matters are not too difficult to handle. We hunt and peck, search and find stimuli that are effective in bringing about the behavior, in "stamping" it in and maintaining it. Theory can greatly facilitate the procedure by pointing up, for example, the classes of stimuli ("reinforcers") that terminate a response sequence by generating a radical alteration in behavior, e.g., food for the hungry rat.

Still in the theory context, the matter of motivation must be faced. Empirically again, the definition of motivation is straightforward: A set of operations performed external to the organism that set the stage for the occurrence of the required behavior. (Inherent in this situation is the heuristic business of the role of the past history of the organism as an external stimulus.) When we say a child is motivated in something, what we mean, descriptively and behaviorally is that he participates in an activity with greater frequency or intensity than his peers. In this instance, the term "motivation" becomes highly redundant and useless.

Motivation is the form of attention behavior getting and sustaining is a core problem in child research. What is needed is a sys-
tematic exploration of stimulus properties that do the job of initiating and maintaining behavior.

A Touch of Methodology

Three matters need consideration here: the experimental treatment, the behavioral measure and control. The last can be dismissed briefly but not easily. Whether or not to apply control procedures to a variable is basically an experimental question: how much and what kind of an effect does the variable have on behavior. In experimental practice, we disregard almost all potential variables in a situation and respond to few. We employ the implicit criterion of relevance: if the dimension of variation has been shown to have an impact on the behavior or if we think it might have, we apply the control operations of minimizing or measuring it. The point is: when in doubt, control, or even better, manipulate the potential variable as an independent treatment.

The matter of behavioral measurement has been elaborated elsewhere (Pascal & Jenkins, 1961; Jenkins, 1966) and need only be touched on here. All behavior is measured in basic terms. All other considerations of measurement are philosophical. The measures are: frequency, rate, latency, intensity, duration, amount, variety, conditions, direction and quality. These apply in whole or in part to all behavioral measurement, regardless of the investigator's theoretical slant. A point not to be overlooked is that the investigator must fit the measure to the experimental treatment and select properties of response classes that are sensitive to and will reflect the experimental operations. In exploratory research it is usually better to use the shotgun early in the game and employ many measures rather than few.
The major point concerning the selection of experimental treatments has many overtones and undercurrents. We are looking for "whopper" effects, but these are assessed a posteriori. The subject is too complex to go into here, in any case. The picture, fortunately is much clearer when it comes to selecting values of a given experimental treatment. Here, there are obvious and concrete rules of thumb. In choosing values of the treatment, cover as wide a range as possible and within the available range, spread out the values. The panoramic view should be taken. A couple of examples may help. If we're studying the impact of food deprivation of food-procuring behavior a la the Skinner Box, it behooves us to consider the range of values available. The experimental approach is continuous deprivation coupled with observation for signs of behavioral, weakening and inanition effects. This provides a cut-off or terminal point. The other extreme is zero, no deprivation, as an anchor point in unconditioned or operant responding. The limits provided then are zero and roughly 96 hours in the typical laboratory rat. Values of the experimental treatment need to be fixed within these limits to provide feedback for a functional relationship to unfold. Assuming a constant environment, values of 0, 12, 24, 48, and 96 hours will do nicely. This selection process says nothing about the actual procedure. The hard way is to use large groups of independent rats. An easier way is to match on behavior at some intermediate "drive" level and switch matched groups to or rotate through different levels of deprivation so as to include a side study of the effects of drive change. The simplest way is to convert the set-up to a self-control design with progressive starvation and tapping into behavior at the set hours of deprivation or, best, rotate
the same, small group through the several experimental values in different orders thereby maximizing behavioral feedback and replication while minimizing the number of cases.

By the same token, these matters apply when we are interested in the discrimination matters known as "concept formation" in children and wish by our experimental procedures to build into the child a broad repertoire and a full reservoir of behavior both verbal and non-verbal. Here the point is to employ training techniques that will maximize generalization so that the child will respond in essentially the same way over a broad spectrum of stimulus conditions. The focus might be on the spatial dimension. First the child needs to be taught by straightforward discrimination operations the notion of spatial separation of objects. He may, for example, be trained to respond to a particular object in a variety of positions and then rewarded for responding to particular positions without regard to the specific object. Coupled with this or following it, conditioning of verbal labels for positions is conducted. During the process, generalization is maximized by presenting stimuli in various sensory modalities and by requiring different modes of responding. This is the heart of the generalization matter that will be treated later in greater detail. Of not so passing interest is the aside that such training programs could be instituted, implemented and conducted in the child's home environment by the mother, grandparents or other available mentors.

Statistics cannot be ignored in any consideration of methodology. Practically all textbooks and most experimental designers consider statistical manipulation the heart of the design matter. This is
wrong. Statistics need to be relegated to their rightful, secondary role as post facto tools for aid in clarifying the effects of experimental treatments. Elegant tables of sums of squares and degrees of freedom are dandy, but they tell us about the behavior of the analyst, not of the subjects. After all, the primary focus is on the behavior of individual organisms, but it is easy to bog down in the morass of statistical manipulation. In brief, the nature of the design determines the class of statistical analysis to be employed; it is a matter of expediency, ease and the nature of the behavioral data that dictates which member(s) of the class are actually applied. Statistics are fun and they're also safe, but the role of the statistical psychologist in aiding the experimentalist or "field" worker is to develop short-cut, "quick-and-dirty" techniques so the investigator can quickly determine what behavioral changes have occurred so he can then go on to the next experiment on his priority list. The rule of thumb, as always, is: the simpler, the better. Anyway, machines can do most of the elaborate computing.

A sequela of this point is to aim for whopper effects. The pathology of the psychologist includes the syndrome that might be labelled "the worship of large numbers". Large numbers somehow accomplish things small numbers don't. This also is wrong and the reverse is true. Compare a probability level of .05 attained in the one case with six behavioral events and in the other with 600. In the first instance, the value means that three events beat three others with non-overlapping distributions emerging. In the latter case, the entire picture is one of overlap in two groups of 300 and little comment can be made regarding group
far less individual behavior. Which is "better"?

In this connection there is another point. Large samples are wanted, but they are large samples of behavior from a small number of subjects. In addition, emphasis should be placed on supporting principles with relatively large numbers of experiments (say three or more) to establish not only the consistency of the effect, but also to obtain a plausible estimate of its magnitude. This leads into the next matter.

An obvious point, sometimes overlooked, is that chance is real. Coins do stand on edge although rarely. No amount of statistical twisting and turning will get around this. The only antidote is replication. Only replication enhances the probability that we are dealing with a basic phenomenon. There is no other way. An investigation once presented me with 100 Chi Squares reflecting a relationship between certain "personality" and "perceptual" measures. He had drawn partly modest, partly sweeping conclusions from the most significant ones. The complete set were buried in an appendix table. On inspection, they yielded a truly rare event: perfect agreement with chance with half positive and half negative. The obvious implications of this event struck me as fascinating, but seemed to disappoint the investigator.

Another statistical point follows from this last instance. Consistency should be equally weighted with magnitude. An effect may be small in magnitude, but have such high consistency that it yields a firm foundation for prediction. A negative case in point may be cited. A few years ago some investigators were concerned with the impact of "values" on "perception" in the coin size judgments of children (Carter & Schooler, 1949). They concluded "essentially no difference" on the basis of five
t-values all of which indicated the same direction of difference, namely, larger judgments for poorer children. Percentage-wise, the differences in means were minute, but five out of five events in the same direction (disregarding lack of independence) yields a P-value of .03. In other words, consistency was a more sensitive index of behavioral effects than magnitude. The two aspects of data go hand in glove, but the point is that both must be considered.

Instrumentation seems worthy of comment in this context. Where the behavior is simple, simple equipment including the eye and ear suffices. The more analytic and delicate the phenomenon, e.g., action of single nerve fibers, the greater the need for elaborate and complex instrumentation. Instrumentation serves the three-fold function of communication of source of messages (stimulus presentation), transmission system (linkeage between stimulus input and response outcome) and receiving station (recording of behavior). If the system does not expedite the message, it may actually retard or distort it. Sometimes the sensory and motor apparatus of the investigator is enough. The criteria are objectivity and facilitation. If the machine system meets these, it will do the job. (Practical matters of cost, transportation and the like are not to be overlooked.) As a case in point, take tracking behavior as one index of sensori-motor coordination in the child. One could rig up an elaborate electronic device with variation in built-in pathways and moving targets for stimulus presentation. Various aspects of behavior such as time on target and summated spatial deviation from target could be recorded automatically and, in the limiting case, the information fed directly into a computer for analysis or storage. As a first approxima-
tion to this fancy model, one might present a path bounded by parallel lines between which the child is instructed to draw a line with a crayon. Number of times the boundary lines were contacted or crossed would be counted. Investigators must choose their weapons.

The Heart of the Matter: Generalization

Cue constancy and cue change are central to all learning situations be they "purely" theoretically oriented or deal with practical teaching. The cardinal rule is that rats, children and other organisms learn faster, the more the stimulus situation is held constant. One can worry about the units of the "similarity" scale, but in the long pull similarity is always defined behaviorally, initially by the experimenter's choice of stimulus values and ultimately by the behavior of the subject. There is no problem here. If one is enchanted with linearity, and the typical curvilinear generalization function emerges from the behavior, a transformation to a semi-log plot will take care of the matter. The main point is to show response decrement (or lack of it) over a wide range of stimulus variation. The limits of the scale of similarity hardly warrant comment. Zero and one hundred percent stimulus similarity (or dissimilarity) are impossible to achieve either logically or psychologically. This is not crucial; a wide range of stimulus and response values are left over. It is a worthwhile intellectual and methodological exercise to attempt to design experimental settings for relatively "simple" organisms that produce maximal and minimal response decrements.

The case needs to be pinpointed with an example. We had children say the numbers 1 through 9 in a "random" fashion with the instruc-
tion that one number was correct (Jenkins, 1955). "Right" was said to the number "five" during the conditioning trials. The response was acquired quickly and the reward pattern was shifted to approximately 50% (varying somewhat with the individual child) to maintain interest or avoid boredom and satiation effects and particularly to prolong extinction. Extinction consisted of the usual operation of turning off reward. The number responses throughout conditioning and extinction were recorded. One hundred extinction trials were given. Except for a couple of kids who perseverated with the number "five" throughout the 100 trials, extinction behavior proceeded according to plan. Since arithmetical averages are usually meaningless and frequently wash out the individual effects investigators are seeking, the data for a typical child - the one taking the median number of trials to condition - are presented in the accompanying figure.

There is nothing new or startling about this function. Even the individual orderliness is common with both human and infra-human organisms. The only special feature characterizing the set-up is that systematic stimulus variation (other than omission of reward, a standard procedure in generalization tests), was not introduced by the experimenter, but rather that the stage was simply set for its occurrence. It might be noted, in passing, that the stimulus and response materials are incidental to the lawfulness of the effect. Similar functions emerge with size and shape and a variety of other items.

The overall effect from a large number of investigations is obvious: response decrement is a function of stimulus change. Conversely, the less the stimulus alteration, the more constant the behavior. If one
The number "5" was rewarded in conditioning and the response frequency recorded in extinction.
wishes to condition a response quickly or maintain it at a high level, the stimulus situation should be held as constant as possible. The rule for breaking habits is to change the stimulus setting for the behavior over a wide range. The ramifications of these simple findings are enormous and extend into all areas of behavioral change including education and psychotherapy. Furthermore, an extensive, basic experimental program is immediately suggested involving the parameters of generalization. For example, it has already been demonstrated that certain kinds of deprivation increase generalization (Jenkins, et al 1958). Exploration is needed of the extent of spread of this finding to determine its applicability to deprivation of safety (increased "anxiety") and economic security.

Returning to the main stream of the pervasiveness of generalization and its particular applicability to training experiments, the statement that cue change weakens behavior does not draw a complete picture. So far the emphasis has been on cue change as it relates to response decrement and cue constancy as it bears on response strength. We need, here, to turn the coin over and look at the other side, and make a careful examination of how stimulus change can enhance response strength. It's really a matter of emphasis. Cue change increases response strength when the members of a class of behavior are conditioned to a wide variety of stimulus situations. Note that the defining operations for a generalization test are cue change and omission of reinforcement. The other side of the coin deals with the more long-drawn out application of the reinforcement operation to a particular kind of behavior in the face of a wide variety of stimulus classes. In other words, to strengthen a response in these terms, associate it with as many stimuli as possible coupled with reinforcement so that generalization is maxi-
mized. All other things equal, the more the cues associated with a response, the likely it is to occur on presentation of a different situation some portion of which has been previously enlisted to the behavior.

Thus, if we wished to teach a child "to talk", we should elicit talking behavior and reinforce it in the presence of all stimuli where language is appropriate e.g., at least all classes of people. This point raises the interesting experimental question of over-generalization - the limits of generalization which is seen clinically, but has not been thoroughly examined in an experimental sense. What are the conditions under which a child, say, learns to talk to birds and bees, dogs and cats, to himself, to dolls, in his sleep, etc. etc.

Simple examples of broadening and flattening the generalization come easily to hand. In teaching a child the directional dimension of near-far, it is obvious that a variety of stimulus objects should be employed in a variety of settings. Again, if we wish to build into the child a skill at knife-throwing, we should use different targets and different environmental conditions.

It is to be noted that training time is greatly increased by this varied conditioning procedure. The behavior has to be attached simultaneously or successively to many different aspects of a number of situations. The purpose of the training need not concern us here, but a decision has to be made whether quick acquisition to a stripped-down stimulus situation is desirable or whether training should be greatly prolonged by applying reinforcement procedures in a wide variety of instances. From the viewpoint of environmental enrichment and cutting back on stimulus deprivation, it would seem wise to use the long way around.
The implications of the wide scope approach for methodology are so obvious that only one-large scale example is needed. If the training program is concerned with teaching the child spatial and distance concepts or rather responses to these stimulus dimensions, then training on them should involve not only many different kinds of objects and settings, but also maximize the sensory and motor avenues involved. For instance, the stimuli should be spaced, say, near or far apart in the visual, auditory, tactual and kinesthetic categories. Objects should be presented visually separated and sounds or words spaced out temporally and in loudness and spread over the tactual continuum. Kinesthetically, the child's limbs can have movements induced in them over a lesser or greater extent. Similarly on the motor side, the child should make the response with his body (walk it), with his limbs in finer type movement (reach near and far), track a moving target with his hands and eyes, and say the words involved while performing the actions. It is well established that the greater the motor involvement in a learning task the more the learning is stamped in (Jenkins, 1943). One should keep in mind that a response can serve as a stimulus for the same person. To this point should be added the more numerous the stimuli enlisted to a response, the stronger it is. There may well be an upper limit on both the sensory and motor sides beyond which further involvement impedes learning, but this is an experimental question.

As a closing note, the action of stimulus constancy and change is so ubiquitous and pervasive, that it behooves investigators not only to consider the effects, but to use them to their advantage.
Bibliography


Tolman, E. C. There is more than one kind of learning. *Psychol. Rev.*, 1949, 56, 144-155.