Abstract — Classical conditioning experiments are reviewed in which the dependent variables are subjective responses, typically involving the evaluation of stimulus materials. These experiments suggest that classical conditioning produces a positive or negative change in the evaluation of previously neutral stimuli. We re-interpret these studies as well as presenting data of our own to support the view that apparently disparate conditioning techniques have an underlying common mechanism — the elicitation and transfer of an evaluative response.

We postulate that a subjective evaluative response, rather than implicit or explicit behaviour acts, carries the mechanism of conditioning. This leaves free the actual repertoire of responses (adaptive or maladaptive) in which the subject engages, and also suggests a way through the complex problem of how behaviour can be generalized outside clinical and laboratory situations. Subjective evaluations are differentiated from states of pleasure (and hence from hedonism), attitude, emotions and approach–avoidance behaviour.

The experiments reviewed are selected from the verbal, attitude and evaluative response conditioning literature, and they are used to shift attention from a model of classical conditioning in which the emphasis is on a specific motor or autonomic response as “the” UCR to one which substitutes a process of unmediated appraisal which we have called the evaluative response. The relevance of this formulation to behaviour theory and to behaviour therapy is discussed in relation to cognitively oriented models.

OVERVIEW AND ARGUMENT

The re-evaluation of fundamental aspects of behaviour theory, viewed in the context of its application to behaviour therapy, is engendering a lively debate. In particular, the question of how a general conditioning model can accommodate subjective experience is the focus of many contemporary discussions.

In this monograph we will suggest that some of the power and economy of objective behaviour theory can be retained by attending to a narrow but significant aspect of responding: the subjective evaluation of stimulus situations by the individual. We will suggest that these subjective evaluative responses can be conditioned in a “pure” form under appropriate circumstances, and will further suggest that they may carry the primary mechanism of conditioning. They also offer a partial solution to one of the problems which underlies current discontent with older models, namely the problem of defining meaningful units of behaviour.

Traditional conditioning theory, based on discrete, externally observable responses, opened the route to what was believed to be a genuinely “scientific” psychology, i.e. one based on comparative studies and with roots in biology and in the laboratory. Interest became firmly fixed on explicit glandular and motor responses; in Pavlov’s work on the glandular response of salivation, in Skinner’s on the skeletal–muscle response of bar-pressing. Other behaviour theorists were willing to admit implicit or covert stimulus—response sequences. Watson’s interest in the small movements of the vocal muscles accompanying speech and of the musculature of the arms and face during thinking began this trend, and Guthrie subsequently emphasized the sequence of proprioceptive responses and patterns of muscular activity underlying movement.
The acceptance of this type of responding as the cardinal criterion for scientific accessibility formed a basis for the development of Behaviourism, and, as is well known, carried with it the outright rejection of introspection and virtually all forms of subjective report. The effects of selecting a particular class of behaviour as a response unit and of rejecting other classes are manifold. In particular, the selection incorporates biases and assumptions of a historical nature and forces a commitment to specific methodologies.

Today the expansion of behaviour therapy as a body of techniques obliges us to call into question these early formulations. The purism of early Behaviourism is seen by many workers to fall short of the real-life demands of treatment. In particular the role of subjective responding assumes increasing importance as the techniques of therapy move away from the problems of isolated tics and phobias and into the areas of more complex and subtle human response systems.

In dealing with what we call a subjective evaluative response it becomes necessary to examine some of the biases mentioned above. The term "subjective" is not necessarily confined to the introspective analysis of mental elements nor of sensations or affects. Although subjective evaluations sometimes run a parallel course with pleasantness/unpleasantness and into the world (and problems) of hedonism, this is not intended, and discriminations between the two concepts will be made. Similarly, subjective responses tend to be equated with the cognitive, though this is not necessarily the case and it will be argued that unmediated subjective responses do occur. Another argument which will be dismissed is that the subjective is necessarily unreliable.

**What is a response?**

It is now widely acknowledged that behaviour is not made up of punctate responses but is a continuous stream. Nevertheless, some segment of it has to be abstracted for purpose of analysis, experimental manipulation and measurement. Records of significant aspects of human molar behaviour in ecological contexts have been well illustrated by Barker (1963), and many examples of animal behaviour in natural surroundings have been contributed by ethologists. Barker makes a useful distinction between behaviour units and behaviour tesserae. Behaviour units consist of inherent segments which are "natural" in the sense that they occur without intervention by the investigator; they are self-generated parts of the stream of behaviour like alpha waves and children's spontaneous games. The other aspect of the behaviour stream he calls behaviour tesserae; these are fragments of behaviour which are created or selected by the investigator in accordance with his scientific aims.

A fundamental place must be given to these conceptual building blocks of behaviour as constructs within a theory. While conditioning theory as we know it has been based largely on the explicit motor act as the major dependent variable, it has often incorporated other elements of implicit responding in the causal chain. It has viewed a given stimulus as arousing a whole pattern of autonomic-muscular events, organized as an S–R chain to determine the final response. The problem is whether those principles which apply to the implicit and visible UCRs can be extended to the subjective. The work on self-reinforcement is an example. The individual is reinforcing desired behaviours in himself, frequently with the aid of imagery, and these behaviours have no known or observable physiological basis; indeed, there is no way, apart from subjective report, for determining that a response has occurred.

In this paper we are concerned with the subjective responses of evaluating or appraising active stimuli in the environment. We make the assumption, following Arnold (1960) and Lazarus (1966) among others, that a characteristic reaction to any stimulus or stimulus complex is to evaluate it as liked/disliked, good/bad, dangerous/safe, etc. at a level which is typically well removed from intense autonomic arousal and well before actual approach–avoidance behaviour. If a reflective (cognitive) judgment does occur it is likely to be secondary. Subjective report, as distinct from immediate subjective responses to environmental stimuli, traditionally referred to conscious mental experiences or states. The introspectionist pioneers of experimental psychology were pre-occupied with psychology as the study of consciousness of mental content and mental elements. This is a very different conception of the subjective from
the use we will make of the term in this monograph. We will use the term *subjective evaluation* to refer to a mechanism for evaluating hostile or hospitable events in the environment. Such a usage recognizes the biological requirement that even the simplest organism must possess a mechanism for responding to and coping with environmental change, in terms of the potential significance of such changes for survival. Such a mechanism is subjective in the sense that it refers only to the individual experience of the organism, whether or not that experience can be reported. The most primitive organisms possess, for example, mechanisms for avoiding or coping with extremes of temperature. In a sense completely distinct from the anthropomorphic it can be said that the organism makes a negative evaluation of such extremes, which is essential for its survival.

Experiments requiring human subjects to indicate which object, real or imaginary, they "prefer" or "like" more, to state how pleasant or good they find something, or to make other evaluative judgements, have been carried out with an immense assortment of stimulus material and with a rich array of data-collecting procedures. One long-lasting branch of the scaling theory stream is experimental aesthetics, exemplified in Berlyne's work (Berlyne, 1973). There is also a long tradition of social–psychological research into attitudes, and workers in this area have been more active than anybody else in devising sophisticated data-gathering and data-processing techniques that can be used for probing evaluative reactions to all kinds of stimuli.

It is in this way, we suggest, that conditioning theory can accommodate subjective responses *per se*, and not by simply defining them in terms of externally observable acts or physiological measurements, or by assuming that they are detectable in terms of sub-threshold responding. In this sense subjective responses can be differentiated from "implicit" responses, a term which implies that while direct observation of the response is not possible it is potentially measurable by sensitive instruments. Technically these responses are unobservable except from "inside the skin", and it may even be the case that they are not always accessible to awareness or consciousness.

So, to review, it can be argued that a class of "truly" subjective responses can be differentiated from other categories, such as the overt, covert, and implicit. They can be independently measured in human subjects and not merely inferred from other types of responses like explicit behaviour or psychophysiological events. Techniques for their measurement are highly developed, including various forms of attitude scaling: Semantic Differential, Kelly Grid, Higher Ordered Metric Scaling, etc. These techniques have shown subjective evaluative behaviour to be systematic and reproducible, i.e. reliable and valid, and in these senses subjective responses can fulfil the desiderata of a dependent variable.

**Subjective responding: associations or strategies**

The question "What is a response?" can be asked and answered, then, in two general ways. The first is in terms of response definition: how to differentiate it, how to specify its beginning and end; how to measure its components. The second concentrates on what sort of activity the response is, on its function and meaning for the organism. One type of answer is that conditioned and unconditioned responses form a coordinated muscle/gland activity which serves the welfare of the organism. An alternative answer, and the theme of this paper, is to postulate that there is a response of evaluating stimulus input in relation to its beneficial or noxious effects on the organism, and that muscle and gland activity subserve this evaluative response.

Subsequent sections will consider contributions from the current experimental theoretical literature on human conditioning in relation to this point of view. Put briefly it is that all conditioning involves the subjective evaluation of an active stimulus (the UCS) and that the response changes produced in classical conditioning are attributable to the transfer of this evaluation from the UCS to the previously neutral stimulus such that the neutral stimulus acquires a positive or negative preference. We intend this formulation to include animal behaviour and the problems which it raises will be discussed briefly in a later section. Our focus of attention, however, is on classical conditioning in human subjects, viewed against a background of biological survival.
A fundamental assumption is that an organism cannot survive unless it has an innate mechanism that tells it when it is favourably or unfavourably correlated with its environment. This carries the implication that the only way it can "get the message" is by some form of subjective awareness of its own welfare and comfort, including the machinery for not "liking" to be uncomfortable. In the natural environment many stimuli will be neutral with respect to unlearned evaluations, while others, such as extremes of temperature and pressure, will produce an unlearned evaluative response on every exposure. It is probable that in simple organisms a fixed and limited repertoire of evaluative responses ensures primarily the survival of the species and only secondarily the survival of the individual. An enormous adaptive advantage is gained when the individual is able to acquire new evaluations of the stimuli in its particular environment as a consequence of its experience of the contingencies in that environment. According to this view it is through conditioned evaluations that the organism or person is able to adjust its behaviour to the situations which arise. An advantage of this view is that it determines how the specificity of learning can be translated into the generality of behaviour.

Clearly the "response" which is observed and measured in any conditioning experiment is not the only response which is "conditioned"; there may be dozens of autonomic responses, subjective responses, and/or adventitious motor responses, that is to say (in behaviourist terminology), many separate S–R connections. However, if one takes the view that what is conditioned is an evaluative response then, when this evaluative response is reinstated as a CR it can carry with it the components, sub-responses or some set of these, depending on the extent to which the response repertoire is "pre-wired" and on the conditions operating at the time learning took place. Stated more formally, we are arguing that conditioning occurs only if an evaluative response is first elicited, and that what is conditioned is the evaluative response itself. The behaviours consequent on occurrence of the conditioned evaluative response will be determined by parameters of the situation in which the response is evoked.

Such a viewpoint implies that conceptual weight is being accorded to the subjective evaluative state, that the total response pattern with its motor, autonomic, subjective elements, etc. is organized such that they follow from the unmediated evaluative response. In the case of certain "pre-wired" reflexes the processes of evaluation and action may occur virtually simultaneously, and possibly at certain phylogenetic levels the two may be indistinguishable. But in more complex instances it could be that it is the evaluation which has primacy, with or without attendant action.

Discussions in the behaviour therapy literature seem to be divided between those willing to retain the S–R conditioning model (e.g. Nawas, 1970) and those who insist upon "strategies", "plans", "programs" and other complex cognitive mediators (e.g. Beck, 1970). It seems to the latter group inappropriate to work within a classical conditioning model which avoids reference to sensations, thoughts and feelings. Are we to consider abandoning a behaviourally oriented conditioning theory for another, variously identified with "cognitive theory" or "humanistic psychology"? Are the shortcomings of behaviour theory severe enough to call for its replacement?

Part of the current unacceptability of the S–R approach to conditioning lies in the model of an associative bond between a specific CS and a specific CR, this latter being traditionally an overt act. Such a conventional associative conditioning model poses formidable problems in explaining how complex CR behaviour is achieved (it suggests manifold independent connections between the CS and independent response elements) and of how behaviour is generalized outside the clinical or laboratory situation.

The "direct binding" of acts to situations stands in opposition to the view that it is shifts in attitudes or perceptual organizations which can affect subsequent behaviour. This latter view was expressed many years ago by Zener (1937) in discussing the differences between observed behaviour to the conditioned and unconditioned stimuli. The conditioned reaction "is anthropomorphically describable as a looking for, expecting, the fall of food with a readiness to perform the eating behaviour which will occur when the food falls... taken all together, the behaviour elicited under the various changes in conditioning suggest that the entire conception implied by the term conditioned response is at fault. It suggests strongly that a response, even a complex one, is not conditioned at all; that the primary structural change is not a connection between the conditioned stimulus and a definite conditioned response, but rather
reorganization of a perceptual system related to an urge system whose combined action releases (not a single definitely defined response) but behaviour characterised primarily by its specific direction relative to the organization of the goal, means, and barriers of the given situation..."

From time to time these arguments recur, most commonly with an approach offered in the form of "cognitive" theories. Breger and McGaugh (1965, p. 355) offered an approach to "what is learned" as "what needs to be done in order to achieve some final event". "The facts of both stimulus transfer and stimulus equivalence seem much more adequately handled if we assume that what is learned is a strategy for obtaining environmental achievements." A few years later Beck (1970) reopened the issue. He argued that the study and analysis of introspective data suggest that the cognitive organization, far from being a mere link in the S–R chain, is a quasi-autonomous system in its own right. Introspective data indicate the existence of complex organizations of cognitive structures involved in the processes of screening external stimuli, interpreting experiences, storing and selectively recalling memories, etc.

Similarly Harlow and Harlow (1970) have expressed their dissatisfaction with traditional conditioning models: "The unlearned or unconditioned responses which form the ground structure for learned primate social responses are neither the specific, isolated Pavlovian-type unconditioned responses nor the random emitted behaviours advanced by Skinner. The unlearned responses... are complex, multivariate, highly organised responses, some of which are relatively late in maturing."

For a while it seemed as if private events could be incorporated in psychological theory by according them the same status as overt responses. Homme (1965) expressed the opinion that private events ("coverants") can be assumed to obey the same laws as public ones, and Cautela (1970) has sought to demonstrate that the manipulation of covert processes could influence overt processes in a predictable manner. Such thinking is in the Skinnerian tradition which, while admitting that a small part of the stimulus environment may be private, would deny that events which take place within the organism have any special properties for that reason. "A private event may be distinguished by its limited accessibility but not, as far as we know, by any special structure or nature" (Skinner, 1953).

It now seems doubtful that such a restricted conception of conditioning theory will be accepted, at least by behaviour modifiers (Mahoney, 1974). Attempts to specify the internal events mediating desensitization outcome have been only marginally successful, but the search for internal responses ("events", "modulations") continues at an accelerating pace. Meichenbaum (1975), for example, has suggested that clients' internal dialogues can be conceptualized as behaviours per se, as reflections of cognitive styles and faulty belief systems, as inadequate problem solving and coping skills, as defence mechanisms, etc., with each conceptualization yielding different forms of treatment interventions. On this basis the task of the behaviour therapist is to match the most useful conceptualization and treatment regime with each client's specific problem and goals of treatment. He has expressed the view that if the research task is framed in this fashion it will merely foster innumerable studies addressed to "engineering" problems. "Not that engineering problems are unimportant. But perhaps we can short-circuit some tedious and expensive comparisons by engaging in theory construction..."

He refers to a delightful quote from Paivio (1975): "Skinner warned us against the diversionary effects of fascination with inner life. I agree that the possibility is omnipresent. Mentalist ideas are so seductive that one is in danger of being led by them down the garden path of introspection and mysticism forever. For that reason, perhaps only a tough-minded behaviourist can afford to entertain the seductress."

If he were to do this, what might become of him? According to Mahoney (1974) who discusses the role and purpose of inferred variables in this controversy, he might become "a thinking behaviourist". Mahoney examined the "non-mediational' behaviourist perspective and considers that in restricting its focus to a relatively small fraction of human behaviour, i.e. that which is publicly observable, the non-mediational model is shown to be inadequate. A cautious optimism is expressed for two mediational models — information processing and cognitive learning.

Mahoney (1974), Berlyne (1975) and Paivio (1975) offer lively and readable views of these issues. As indicated, Mahoney is inclined to consider various cognitive mediational models.
Berlyne (1975) carefully considers why it is that Hullian neo-behaviourism has so fallen out of favor but reminds us of three valuable features which are absent alike from Skinnerian radical behaviourism and from the writings of most present day cognitive theorists and humanistic psychologists. These are (1) integration (compared with today’s fragmentation of psychology), (2) motivation, which must identify the conditions that generate pleasure and discomfort, which give something reward value or punishment value and (3) symbolic processes. The contention that behaviour depends on cognition encourages neglect of the need to state explicitly the principles that lead to the one from the other: “We need to distinguish the occasions when Occam’s razor leaves us no grounds for inferring the intervention of cognitions or thought processes and the occasions when we are compelled to assume that they are at work. Finally, we need to relate the higher, characteristically human forms of behaviour to the lower and more primitive forms from which our psychobiological point of view compels us to assume that they have developed.”

A fuller discussion of the cognitive conceptual position will occur later. The distinction which must be emphasized here is between a “direct binding” of stimulus and response, whether that response is an innate reflex, a pattern of autonomic activity or a complex repertoire, and the use of mediators like cognitions. Thus in processing the CS the organism may be saying “I know that something is coming that I fear”, or it may say simply “I fear the stimulus”. The first is essentially a cognitive position and the second is behavioural. The organism may acquire an evaluative response towards a previously neutral CS. A cognitive model would imply that it attaches significance to the CS, and is likely to use the concept of a template or plan of the external environment in such a way that the template or plan implies the necessary action. This is close to the situation—act—outcome analysis as used by Seligman and Johnston (1973) in their cognitive model. The behavioural (“unmediated”) model says, alternatively, that the evaluation of the CS is itself a feeling state of some sort that does not necessarily carry implications of understanding, though understanding may follow from it.

In contemporary psychology there appears to be a no-man’s land of subjective behaviour which is the playground for ad hoc speculations about cognitive mediation in conditioning. Yet a case might still be made that postulates of cognitions are the imaginary humours and miasmas of our world view while the more economical conditioning principles are the laboratory defined principles with which we can hopefully replace them.

To summarize, we suggest that in introducing subjective responses into classical conditioning theory a re-conceptualization is required. If behaviour therapists carry over a conventional associative conditioning model, formerly applicable to overt responses, really complex problems have to be faced, e.g. how is behaviour generalized outside the clinic or laboratory? The use of various kinds of implicit responses as mediators within a complex S–R chain leads into tortured accounts of stimulus generalization and probably of response generalization which are difficult to justify. It also carries with it a theoretical model applicable only to more mechanistic types of peripheral reflex responding.

On the other hand, the uninhibited introduction of numerous loosely defined mediational concepts does not contribute to rigorous models or productive research. Terms such as “cognitive conditioning”, “emotional conditioning” or the array of specific mediating factors such as “belief system”, “strategy”, “set”, “self-perception”, etc. must be critically examined against the history of classical conditioning and the requirements of a scientific system.

The classical conditioning of evaluative responses

The purpose of this monograph is to suggest an alternative formulation which retains the parsimony and generality of the classical conditioning model, while accommodating subjective responses in a central role. We have labelled this formulation evaluative conditioning, with the reservation that we regard it as more than just another form of conditioning, for reasons which will become evident after the basic paradigm has been described.

The concept of evaluative conditioning, as a “pure” form emerged in an experiment which we developed (Levey and Martin, 1975) and have since explored in order to throw some light on the issues raised in the preceding discussion. The experiment will be described briefly and its relevant features underlined. It may be worth noting first that the initial experiment was designed not to study either evaluations or attitudes per se, but to test inferences drawn from
our model of the genesis of classically conditioned responses (Martin and Levey, 1969). In the exercise of designing an experiment to test these inferences, we became interested in the implications of evaluative responding. While these inferences are not directly relevant to the issues involved in evaluative conditioning a useful perspective may be established if they are briefly outlined.

Working chiefly with eyelid and skin resistance response (SRR) conditioning we proposed the following model to account parsimoniously for the basic phenomena of classical conditioning. Whenever an organism encounters a salient stimulus or stimulus situation in the environment it stores the contents of immediate memory in a comparator store analogous to that proposed by Sokolov in his well-known model of habituation. The temporal contiguity of the salient event (e.g. corneal air-puff UCS) with the immediately preceding event (e.g. tone CS) ensures that the contents of the comparator store will contain both stimuli. We proposed that these are stored as a stimulus complex (CS/UCS complex) in which the stimulus events are undifferentiated. Thus when either of the stimuli occurs again it functions as if it contained the shared characteristics of both CS and UCS. If the CS occurs, it acts therefore as a weak UCS, and we used this supposition to account mechanistically for the appearance of the first overt CR. We further postulated that the CR and UCR themselves form a response complex which is modelled on the CS/UCS complex in such a way that CR/UCR integration becomes an important “end-point” of the classical conditioning process. We also suggested that in this way the CR is able efficiently to perform an increasing part of the “work” of the UCR, with a resulting economy of effort for the organism. One other implication of this model is important, namely that it allows for, and indeed predicts, the occurrence of “one-trial” conditioning.

Further aspects of this model, together with the evidence supporting it, are considered in the reference just cited, and have been discussed elsewhere (Levey and Martin, 1968, 1974). For present purposes the interest of the model centres on two problems: the definition of “salience” and the definition of “shared characteristics”. The difficulties of describing the subjective characteristics of a corneal air-puff and a neutral tone render it virtually impossible to test the notion of shared stimulus characteristics. We turned to complex visual stimuli which would have meaning for human subjects which they could describe by introspective report. We argued that visual materials which are strongly disliked or liked would serve as “salient” stimuli in the sense intended in the model of classical conditioning. We therefore argued that the stimulus characteristics which lead to the subjects’ liking or dislike of a specific visual material (e.g. pictures) would transfer to similar materials toward which the subjects’ attitude was neutral.

The initial experiment (Levey and Martin, 1975) investigated the effect of pairing neutral pictures with pictures rated as liked or disliked, in a compact Latin-square design which allowed us to test a small number of subjects in each of five different pairings. In brief, the procedure was as follows: Ss were presented with a set of 50 picture postcards of unfamiliar paintings and asked to sort them into three piles labelled “liked”, “neutral” and “disliked”. They were instructed to use their spontaneous reaction to the pictures and not to judge them. After re-sorting the neutral pile to eliminate borderline likes and dislikes, the pictures were paired in the sequences: neutral—liked, neutral—disliked, neutral—neutral, liked—neutral and disliked—neutral. They were then presented tachistoscopically in runs of each pair and, after exposure of all five pairs, all ten pictures were rated on a numerical scale. Sessions were run individually and each S was interviewed at the conclusion of the experiment to determine first whether he was aware of changes in evaluation and then his explanation for any changes noted. Results rather strikingly favoured a conditioning interpretation, as shown in Fig. 1, in that previously neutral stimuli became either positively or negatively evaluated according to the way they were paired.

The design of the study was also intended to compare the effects of contrast with those of conditioning in order to test the validity of a conditioning effect. It should be noted that if the neutral pictures changed as a result of contrast effects the changes would be in a direction opposite to that expected in a conditioning paradigm. An interesting result emerged: in the original study no subject was aware of the conditioning contingencies, all were aware that evaluative changes had occurred, though they could not specify their direction; and most attributed them to contrast effects. In other words the Ss offered an explanation which was opposite in direction to the changes shown by their ratings.
The experiment demonstrated several features which led directly to the concepts involved in the notion of evaluative conditioning. The actively liked or disliked stimuli, which functioned as UCSs, evoked no overt response, though the evaluation itself was clearly transferred to the neutrals. Thus, if conditioning occurred, it was of some unobservable internal response. The evidence for a conditioning effect was the set of ratings of the previously neutral stimuli. These ratings have no resemblance to the original sorting task. Thus the behaviour measured was a reflection again of some unobservable internal event, structured by the situation in which it was evoked. In short, the observable "UCR" and "CR" were quite different responses, mediated by a subjective evaluative response which was not itself observed. The question of demand characteristics and awareness of contingencies did not arise, probably as a result of the within-subjects design which changed the contingencies with each run of presentations. Responses to the post-test questionnaire showed that Ss were not only unaware of the direction of changes in evaluation, but attributed them to an irrelevant cause. Each subject was conditioned in terms of his own perception of neutral and liked or disliked items. When tested 18 months after the initial experiment, the conditioning effect had persisted, though subjects were unable to identify which pictures they had previously seen.

The experiment leads directly to the conclusion that a response can be conditioned in the absence of an overt UCR, and suggests that what is conditioned is the non-cognitive evaluative response of liking or dislike. The logic of our interpretation of this finding is that if positive and negative evaluations can be conditioned in the absence of a physiologically active UCS, they must be regarded as the sufficient components of the classical conditioning paradigm. It must follow from this that the motor and glandular responses which we count and measure in the usual conditioning experiment are in some sense peripheral or secondary to the basic mechanism of conditioning. Whether evaluative responses are also the necessary components of classical conditioning is open to argument, but not to proof. Since any UCS, noxious or pleasurable, which elicits conditionable responses must a priori be positively or negatively evaluated, it is not possible to condition an overt response without conditioning an evaluative response. Conversely, it is possible to condition the isolated evaluative response, and this suggests that evaluative responses are the necessary and sufficient components of classical conditioning.

A subsequent experiment (Levey and Martin, 1975a) tested the effect of having subjects rate the materials both before and after pairing. Interestingly this procedure produced relative changes in the neutral stimuli, shown in Fig. 2, which were in the predicted direction and at a satisfactory level of significance, but lowered the entire scale. In other words, the positively conditioned stimulus was significantly preferred to the negatively conditioned stimulus, but was itself disliked. We have since repeated the experiment a number of times, varying the parameters systematically in order to establish the factors underlying it. We have also encountered difficulties which need not be discussed in detail at this point but which hinge largely on the problem of ensuring that Ss make evaluative and not cognitive judgements. Indeed we have ample evidence that cognitive judgements disrupt rather than facilitate the effect. The results of these further studies will be discussed in the literature review to follow.
In this section we review a selection of studies which form the background of the evaluative conditioning concept, drawing on studies that emphasize the significance of subjective factors in classical conditioning. Rather than infer subjective states from overt behaviour or from autonomic responding we will consider an inversion of the argument: it is as a result of subjective process of evaluation that overt acts occur. The literature chosen is entirely based on human classical conditioning and entirely on those studies which have used verbal report for the assessment of subjective factors (such as ratings of liking, etc.). This therefore excludes all classical conditioning studies in which overt acts and autonomic responding have been measured exclusively as dependent variables. A considerable amount of the research to be discussed has been carried out within a general context of attitude conditioning, but with "what is learned" sometimes referred to as affectivity, sometimes as semantics and meaning and sometimes as cognitions. What these studies have in common is the reference to subjective processes of evaluation, carrying the implication that these subjective processes are primary determinants of behaviour.

A growing body of literature bears on this basic formulation. The literature can be categorized into three phases: (1) a series of precursors, studies which were concerned with the issues of interest, but not devoted explicitly to them; (2) a series of prototype studies which advance toward the concept of evaluative conditioning and (3) a series of more recent studies, growing from the earlier work, which we will label, for the sake of alliteration, proliferations of the basic concept. In the interest of clarity we will deal first with the prototype studies in order to establish a frame of reference. Within this framework the significance of the precursors will become clear. Finally we will review a selection of proliferative studies, selecting those which bear most directly on the subject matter of this review. In selecting the last named groups we do not pretend to have exhausted a field which is becoming surprisingly active. Rather, we have chosen studies which exemplify points of interest to the basic argument.

The prototype studies fall into three main categories, each associated with a particular author or group. These will be examined in historical order, since it is in that order that they move toward a concept of evaluative conditioning. The first of these prototypes was a group of studies conducted by Gregory Razran beginning in the 1930s, which explored the development of positive and negative attitudes reinforced by food. Second, the work of Arthur W. Staats on attitude development in the late 1950s explored the effect of verbal reinforcers in transferring positive and negative evaluations to neutral stimuli. Studies begun in the mid-1960s by Jum C. Nunnally and his co-workers comprise the third prototype, and were aimed at establishing the conditionability of reward values, reward expectancies and selective attention in a developmental setting. These three prototypes will now be described in some detail.

Prototypes of evaluative conditioning

Razran's basic experimental paradigm was labelled by him the "luncheon technique" (Razran, 1938). Typically, stimulus materials of a wide variety (musical selections, literary quotations, photographs, paintings, socio-political slogans, etc.) were presented and Ss asked to
rate them in terms of personal approval, liking, social effectiveness, and so on. In one representative experiment, the material (slogans) was then divided into two sets. One set was always presented while Ss were enjoying a free lunch and the other while they inhaled a number of "unpleasant odours of a putrid nature". To confuse the subjects' memories non-experimental slogans were added at each session. After 5–8 sessions of conditioning, the original procedure was repeated, with subjects again rating each original slogan. Slogans associated with the lunch clearly showed increases in the various ratings (i.e. more positive evaluations) whereas those combined with unpleasant odours showed decreases in ratings. Changes were not a matter of conscious memories since a check indicated that subjects' knowledge of which slogans were combined with pleasant and which with unpleasant stimuli was little above chance.

In a series of further studies (Razran, 1940, 1954) it was shown that the presentation of positively or negatively toned stimuli, in our terms the elicitation of positive or negative evaluative responses, resulted in transfer of hedonic tone to previously neutral materials. This effect was attributed by the investigator to the processes of classical conditioning. Apparently what was conditioned in the series of experiments, besides salivation, was in his view more a general affectivity than a specific attitude.

Razran (1971) subsequently developed a comprehensive phylogenetic description of learning and adaptive behaviour involving several levels of responding. At the time of the early studies, however, he was concerned with the problem of what is conditioned, and offered the possibility of two classes of conditioning, perceptual and non-perceptual. He was concerned with what, in contemporary theory, we term the problem of awareness, in his words the perception of relations between stimuli and reactions. He argued for two levels of learning, the higher dominating the lower. Whenever learning takes place, he suggested, it involves the operation not of one kind or the other kind but of either (a) the lower kind or (b) the higher and the lower kind.

He argued that whereas the phylogenetically lower order of classical conditioning could occur in human subjects as an isolated event, the higher form of conditioning, perceptual learning, required the lower as its substrate (Razran, 1955). This view was based in part on observations of human subjects becoming aware of stimulus contingencies during conditioning, and modifying their behaviour. The conditional behaviour was not "nullified" by awareness but modified in such a way that the conditional responding persisted. A noxious stimulus, for example, ceased to be an "annoyer" and became a signal of "what not to do". In this formulation Razran seems to have offered an account of complex behaviour which was well in advance of his era. It is difficult to disentangle the skeleton of logically necessary constructs from the rather florid literary style of this ingenious investigator ("The View that... our vast non-perceptual equipment is either unalterably fixed or unalterably subservient to perceptual learning is forsooth (sic!) more depictive of satans and saints than of true humans" (Razran, 1955, p. 93).) Nevertheless, his thinking anticipated, in underlying rigour, some of the problems raised today by the proliferation of ad hoc mediational concepts in behaviour theory.

To summarize, Razran seems to have been implying that a specific reaction (salivation) can be conditioned, but in addition a specific attitude is also conditioned. In some of his experiments this was a specific food-related attitude, as measured by the frequency with which food-related words were given by Ss, or the speed with which Ss unscrambled letter-scrambled anagrams of food-related words. But more general than this was what he termed "general affectivity" by which subjects conditioned to a positively evaluated unconditioned stimulus-situation (lunch) came to give a more positive affective rating. What therefore seems to be conditioned was a central evaluative state. When this evaluation was reinstated as a CR it carried with it the components or sub-responses of the original behaviour (or some subset of these) depending on the extent to which the response was predetermined. In this way Razran implied, among other things, a mechanism by which the specificity of learning can be translated to the generality of behaviour. It is remarkable that this formulation was offered at a time when most S–R theorists were concerned with the unitary S–R bonds of particulate behaviour.

The second experimental prototype to be considered is due to Staats. This author has contributed an extensive body of theory and experimental work to the field of behaviour theory. Apart from the studies to be considered here, which are concerned with the conditioning of what we would call "evaluative responses" he has developed a dynamic, as opposed to typological, theory of personality (Staats, 1970) based on principles of behaviour.
theory, which goes a considerable way towards meeting, in meaningful terms, some of the issues (e.g. uniqueness and continuity of experience) long held to be the property of personalistic theorists, and which may well form an important contribution to the body of theory underlying Behaviour Therapy. He has contributed, in addition, a novel reformulation of the body of theory in his concepts of Language Behaviour Therapy (LBT) (Staats, 1972) which avoids the vagaries of some of the ad hoc cognitive formulations to be found in the recent literature. The limitations of our subject matter do not permit a further exploration of this important body of theory. A cogent exposition of his view (Staats, 1975) has recently been published, and provides the reader with a systematic formulation of his theoretical position. For present purposes we concern ourselves with an early series of experiments, designed to explore the classical conditioning of evaluative meanings and attitudes which represent a step toward the concept of evaluative conditioning.

In one of their earliest experiments, Staats and Staats (1957) paired nonsense syllables with words all having in common a positive evaluative meaning (e.g. pretty, sweet, healthy). Their hypothesis was that the positive meaning of the words would be conditioned to the nonsense syllable, and to assess this shift in evaluation they used Osgood's Semantic Differential (SD) technique to measure the factors of evaluation, potency and activity. Nonsense syllables were found to have significantly shifted along the evaluative dimension. Other experiments in which "active" and "passive" meanings, or "strong" and "weak" meanings, were conditioned were also found to have shifted along the appropriate dimensions, though the shifts were smaller in magnitude.

Included in the rationale underlying the paradigm was the additive assumption that total word meaning is composed of response components which can be separately conditioned. In this early work the focus of interest was on meaning as defined by the attitude dimensions of the Osgood SD, that is, on connotative meaning. For the present purpose, however, the interest centres on the fact that these "meanings" are evaluative in character. In subsequent studies Staats and his co-workers came to regard these as evaluative attitudes, and referred to their technique as "attitude conditioning".

A subsequent study (Staats and Staats, 1958) extended the original experiments by studying the formation of attitudes (evaluative meaning) to socially significant verbal stimuli through classical conditioning. The socially significant verbal stimuli were national names and familiar masculine names. These stimuli, unlike nonsense syllables, were expected to evoke attitudinal responses on the basis of the pre-experimental experience of the Ss.

National names (German, Swedish, Italian, French, Dutch and Greek) or masculine names were used as visual CSs (projected) while the UCSs were words of evaluative meaning, either positive (gift, sacred, happy) or negative (bitter, ugly, failure) spoken aloud by E and repeated by S. Again, the Osgood SD evaluative scale was employed. Results of the analyses indicated that conditioning occurred in both cases. The conditioning paradigm is schematized in Fig. 3.

---

![Diagram](Fig. 3. Schematic diagram of stimulus and response relationships in attitude conditioning. (After Staats and Staats, 1958.))
The national name Dutch, in this example, is presented prior to the word Pretty. The authors' description of the conditioning process is as follows: “Pretty elicits a meaning response. This is schematized in the figure as two component responses; an evaluative response $r_{pv}$ (in this example the words have a positive value) and the other distinctive responses that characterize the meaning of the word, $R_p$. The pairing of Dutch and pretty results in associations between Dutch and $r_{pv}$, and Dutch and $R_p$. In the following presentations of Dutch and the words sweet and healthy, the association between Dutch and $r_{pv}$ is further strengthened. This is not the case with associations $R_F$, $R_S$, and $R_H$, since they occur only once and are followed by other associations which are inhibitory. The direct associations indicated in the figure between the name and the individual words would also in this way be inhibited.”

It was not thought that a mere rating response was conditioned in this procedure but rather an implicit attitudinal response which mediated the behaviour of scoring the SD scales. The authors did not suggest that the results of the study showed directly that Ss' behaviour to the object (e.g. a person of Dutch nationality) had been changed; rather, that the attitudinal response to the signs, the national names themselves, had altered. Nevertheless, the results of the study have special relevance for the understanding of attitude formation and change by means of verbal communication. The authors interpreted their findings as supporting the view that attitude formation or change through communication takes place according to principles of conditioning: “As an example, the sentence 'Dutch people are honest' would condition the positive attitude elicited by 'honest' to 'Dutch'—presumably to any person called 'Dutch'. In an individual's history, many words eliciting a positive attitude are paired with 'Dutch', then a very positive attitude toward this nationality would arise.”

In a subsequent experiment, using a similar “language conditioning” technique, Staats and his colleagues tested for semantic generalization. (Staats et al., 1959). Thus the word CARPET for example, was paired with a word having positive evaluation, and the word ROCK with a word of negative evaluation. Later the evaluative meaning of synonyms RUG and STONE was measured on the SD scale. It was found that synonyms had significantly acquired evaluative meaning, thus the conditioned meaning had generalized to the synonym of the originally conditioned word.

Further experimental work established a reinforcement gradient (Staats et al., 1960) using the same general conditioning technique, i.e. visually presented nonsense syllables (CS) paired with auditory presentation of affectively toned words (UCS) and SD ratings. Three levels of reinforcement (100, 5 and 0 %) were used, and revealed significant differences in response intensity in the expected direction. Thus, in a series of experiments this group of investigators were able to apply one of the basic criteria of classical conditioning, viz. that it obeys well-known principles.

In terms of our present focus of interest these studies support the view that an evaluative response can be conditioned using the classical paradigm. The attention of Staats and his group to the response components of verbal meaning, and to the concept of attitude, while interesting in themselves, represent a level of theorizing which may be felt to be superimposed on the simpler finding that positive and negative evaluations were conditioned. At this level there is a parallel with the suggestion of Razran, discussed earlier, that "higher" forms of perceptual learning use as their substrate the lower forms of classical conditioning.

Unlike the work of Razran which was largely ignored, though its implications were ahead of their time, the work of Staats has been rather widely criticized on methodological and theoretical grounds. It may be an interesting comment on the Zeitgeist that Razran's work, which was methodologically hazardous, to say the least, escaped such criticism. It is perhaps gratifying to infer that both methodological standards and theoretical awareness have improved with time! In the interests of continuity, the criticisms of this work and an assessment of its significance will be discussed briefly before turning to a consideration of the next general prototype.

The two main foci of criticism which have emerged are: (1) whether or not awareness of the reinforcement contingency is necessary for conditioning to occur and (2) whether or not the conditioning observed is an artifact of the demand characteristics in the experimental situation. Both of these issues were addressed by Insko and Oakes (1968) in an experiment which attempted to produce conditioning of attitudes towards a nonsense syllable both with and
without the presence of an interfering intertrial activity, colour naming. The assumption underlying this manipulation was that colour naming interferes directly with the hypothesizing behaviour of the Ss (awareness) but not directly with conditioning. A similar procedure to that of Staats and Staats (1958) was used, with the addition of measures of S's awareness. Results indicated that conditioning and awareness are correlated. The authors comment "it is not clear from the discussions of previous researchers whether the conditioning of attitudes was intended to be a conditioning of symbolic evaluative meaning, or of affect, or of both" (p. 495).

Page (1969) provided further evidence to support his claim that the results of the higher-order conditioning paradigm "are entirely artefacts of demand characteristics . . . ." An internal analysis of his data revealed that only those Ss who reported awareness of the CS–UCS contingencies and, before the post-test, awareness of how the experimenter expected them to evaluate the CSs, showed a conditioning effect. In contrast, Ss either unaware of the contingencies or aware of the contingencies but unaware of the demands, did not show any evidence of conditioning. Staats (1969) has challenged this interpretation on the grounds that post-experimental questionnaires may produce varying levels of awareness as well as measure it.

Page (1974) has recently continued his experimental attack on the problem in a series of experiments to counter Staats' objections that the post-experimental questionnaire itself contains demand characteristics that suggest awareness to conditioned but unaware Ss. Again when Ss who were "aware" were removed from the analysis, the unaware Ss showed no significant effect.

It has also been argued that experiments which attempt to show the classical conditioning of attitude cannot in their very nature escape the subjects' awareness of the demand characteristics of the experimental situation. Zanna et al. (1970) took up this problem in an experiment which appears to meet the criticism. They argued that because the experimental hypotheses are often direct and uncomplicated (especially to introductory psychology students) subjects may easily be able to detect what they are expected to do in the experiment and then simply comply with the experimenter's presumed wish. Using a first-order conditioning paradigm (adjectives paired with onset or offset of shock) they hypothesized that the adjectives would come to be evaluated, respectively, more negatively and positively. In order to make the demand characteristics of the experimental situation irrelevant to the hypothesis, an elaborate cover story was created, which motivated subjects to cooperate with the experimenter in developing "new and more sensitive" indices of physiological responding. Skin resistance response (SRR) measures were used as an objective index of conditioning to the electric shock.

That the cover story was effective in deflecting attention from the demand characteristics of the situation was shown by a post-test questionnaire which revealed that all Ss reported that they believed the contrived purpose of the experiment. Thus the experimental design appears to have ruled out a demand-characteristic explanation of the results. Conditioning results were not strong and were complicated by the fact that the adjective CSs used differed in their level of affect. However, the conditioning effect was stronger in those Ss showing independent physiological evidence of conditioning, whereas in those showing no physiological evidence the conditioning effect was very weak.

The issues surrounding the identification of awareness of stimulus contingencies on the one hand, and of the influence of experimenter demand, on the other, are thorny indeed, and will not be considered in detail here. The problem arises, in part, from purely methodological difficulties in assessing awareness, a seemingly simple problem until one attempts to solve it. It is not enough to establish, for example, that S knows that shock follows the presentation of a particular trigram. He must know that it always and only followed the relevant stimulus. In our own experience in this and other fields of conditioning, it is remarkable the extent to which Ss are unsure of their answers to these questions, when they are probed. Yet the mere act of probing introduces an experimenter demand which may influence Ss level of certainty. Conversely, failure to probe (either verbally or by multiple choice distracters) leaves the essential question unanswered.

An ingeniously bizarre technique devised by Page (1974) in the study just mentioned requires Ss to revise their ratings to "completely opposite" to the way they think they are "supposed" to be in the eyes of the experimenter. The bizarre aspect of the procedure emerges in the following instruction: "Now for those of you who don't know what I am talking about... just go ahead and rate the syllables as best you can. Do not spend time trying to
figure out what I am talking about, just rate the syllables according to your natural feelings without this interruption" (p. 471). Whether such an instruction might dissuade the already fragile effects of attitude conditioning and destroy a genuine conditioning response is a matter for speculation. In fact, data from a control group suggested that it did not, but other difficulties were raised which are discussed in the original paper.

One further difficulty with the experiments using this paradigm merits attention, though it is not crucial. In most, but not all, of these studies S has been required to repeat the UCS word aloud on each trial. This Talmud Torah procedure could have the effect of producing a Pavlovian state of induction, which must be differentiated from “true” conditioning. In other words, the production of a momentary set could mimic the effects of conditioning since it would lead Ss to repeat the associated words. If the attitude conditioning were in fact due to induction, and not to conditioning it would have interesting consequences. The semantic differential rating of nonsense trigrams is a difficult and implausible task for many subjects. If the process of induction were at work it would be the aware subjects who would rate, not the trigram, but the UCS itself, implicitly vocalized, and this would be equally true in the reversal technique just described. In summary, while the concept of verbal reinforcers has considerable theoretical power, investigators or clinicians who use it in an explanatory role must be aware that it is fraught with peril.

Although discussed within the context of attitudes, and although positive and negative evaluations of CSs are obtained, it is language and the semantic aspects of his theoretical system which seem to be Staats’ main preoccupation. He has become concerned with the roles of language as a reinforcer, as a mediator, as the modus operandi of a complex process of integration of classical and instrumental conditioning mechanisms in language and in some other aspects of behaviour. This interest can be differentiated from our own in that we are not primarily concerned with semantics, but rather to stay “inside” the unmediated part of the system by attempting to measure the subjective evaluation itself. Staats’ interests have become engaged in the semantic system itself, and with, for example, the “detachable part” of meaning, but with a growing emphasis on the powerful emotional effects which words can have. This has led him to contend that words can act as emotional reinforcers, and there is considerable evidence to support his view including a study in which emotional meanings, conditioned to verbal stimuli, could serve as reinforcers of instrumental behaviour (Finley and Staats, 1967).

It is difficult to assess the interesting body of work just reviewed. It is unfortunate that to date the problem of the role of awareness remains unsolved. The contention of Staats that it is inherently insoluble may have considerable validity. At the same time, the argument that the demonstration of awareness precludes a conditioning explanation needs critical examination, as the sophisticated early analyses of Razran showed. It can be argued, post hoc, propter hoc, that the establishment of a conditioned response is the necessary condition for awareness, and that subjects who are not aware have simply failed to condition. This possibility, in the writers’ view, makes a nonsense of Page’s reversal technique. (“If you don’t know what I’m talking about just go ahead as best you can”!)

The attitude conditioning paradigm has passed some of the parametric tests (e.g. generalization, reinforcement gradient) accepted as valid by Pavlovian theorists. It has been replicated and extended by workers who are not associated with the original investigator and must be regarded as a robust phenomenon. Costello (1972), for example, has used an exact replication of the original study to demonstrate superior “conditioning” of introverts. For the purposes of our present argument we would like to believe that the conditioning explanation is true; however, we are fully aware that the data base for making this evaluative response has not yet been established, and the prototype must be regarded as illustrative rather than definitive.

The problem of how neutral stimuli acquire reward value has exercised the interest of Nunnally and his co-workers and their work provides the third general prototype of evaluative conditioning. Working with children, Nunnally et al. (1965) reported an experiment in which positive and negative value was conditioned to neutral stimuli by pairing that stimulus with a positive or negative consequence. Nonsense syllables were employed as neutral stimuli in a game designed to appeal to children. Copies of three trigrams were placed around the perimeter of a spinwheel so that each stimulus appeared in random order, and depending at which trigram the pointer stopped, S experienced the positive, negative or neutral event, viz. gain of 2 pennies, loss of 1 penny and no outcome. An interesting variety of methods was used to measure the
effects of this pairing of neutral and positive/negative stimuli. Measures of verbal association assessed the pleasantness/unpleasantness of the nonsense syllables following the spinwheel game. This consisted of ten triplets of words each composed of one word having positive, one neutral and one negative connotation, e.g. sweet, bland, bitter. Each trigram was paired with each of the word triplets and the thirty pairings were randomly ordered on a response sheet. S was instructed to spell aloud each trigram and quickly mark the word in the triplet that reminded him most of that syllable.

Several measures were used to indicate the “reward expectancy” of the nonsense syllables. These were based on the hypothesis that after a neutral stimulus has achieved cue value for rewards in one situation it will serve as a cue for rewards in a different situation. For this purpose a “guess-who” game was used. Fifteen questions were given, five of a negative character (“Who steals pennies from other children?”), five of a positive character (“Who has the most friends?”) and five neutral (“Who walked down the street?”). S was asked to say to whom each question applied by indicating one of three stick figures placed before him. The stick figures were identical, but beneath each figure one of the nonsense syllables appeared and was referred to as the stick figure’s name.

The authors discuss their results in terms of a conditioning effect. Overall, the results of the studies indicated that pairing a neutral stimulus with a positive or negative consequence results in that stimulus acquiring positive or negative value. Association of neutral stimuli with rewards and punishments affects not only their ability to serve as secondary reinforcers but also their connotation as pleasant/unpleasant stimuli. Other studies by this group of workers have demonstrated that the effects observed are not transient but are maintained over at least a 24 hour period. Previously neutral stimuli were used as tokens for the receipt of future rewards even in tasks that differed markedly from the original task of association and in which the expected reward was different from that with which the stimulus has been associated. Thus, conditioned reward value effects observed were strong, lasting and generalized. The method of measuring semantic appraisal of neutral objects associated with pleasant and unpleasant events has been refined and further standardized (Rielleh and Nunnally, 1970) and provides a promising technique for investigations of conditioning effects.

In a subsequent study by Faw and Parker (1972) the spinwheel game was adapted to young adults, using as positive reinforcers 15 second viewing of an attractive nude, and as a negative reinforcer immersion of the arm in ice water. Ss made their own spins. The results confirmed for young adults the findings reported above for children.

A further observation of interest concerned the effectiveness of merely anticipating a positive or negative association on the acquisition of conditioned reward value. Ss who only anticipated the associations were as much influenced in their behaviour as those who actually experienced those associations. This finding raises important problems concerning the psychological processes which underlie the acquisition of conditioned reward value. In the authors’ words: “If a simple association model is to explain the present results, it seems necessary that such a model expand its concept of a positive or negative event to include covert anticipatory responses of those events as well as the events themselves. If this expansion is necessary, then one might ask to what extent the earlier results, as well as present ones, are determined by simple stimulus-stimulus associations and to what extent they are dependent upon Ss’ cognitive activities. The present study does not answer these important questions but does provide data which motivate their formulation.”

A subsequent study (Parker and Rugel, 1973) examined the effect of counter-conditioning on the positive or negative value acquired by a neutral stimulus in a conditioning situation. Children were employed as subjects, and the conditioning apparatus was the spinwheel mentioned above, using the same reward relations. Dependent measures as follows were taken within a week following the third conditioning session: (i) verbal evaluation, i.e. the “guess-who” game mentioned earlier and (ii) reward expectancy. Each nonsense syllable appeared on six boxes (eighteen boxes in all) and Ss were given twelve chances to guess which box contained 25 cents. Within a week following administration of the dependent measures, the counterconditioning sessions were begun for appropriate groups. It was explained to Ss in these groups that the syllable that had been associated with reward would now be associated with loss of reward, and vice versa. Five sessions spaced 1–2 days apart were run under these conditions. Following the last counterconditioning session the dependent measures were again administered...
to all groups. For the verbal evaluation measure the counterconditioning groups changed their evaluations of the initial reward syllable from positive to negative, and changed their evaluation of the initial loss-of-reward syllable from negative to positive. For the reward-expectancy measure, however, the data supported only a positive effect, in that association of a neutral object with reward resulted in acquired positive value. No reward expectancy accrued when a neutral stimulus was paired with a negative consequence. The pairing of the initial reward syllable with loss of reward during counterconditioning also had no effect. Since the reward-expectancy measure asks S to anticipate only positive effects, it may be sensitive only to conditioning using positive consequences. However, again, the results suggest that this paradigm contains elements of some cognitive process, and indeed the instruction that reward values would be reversed is consistent with this.

This series of studies is apparently on-going and further results will be awaited with interest. Their relevance to evaluative conditioning lies in the concept of transfer of reward, though the results for anticipation of reward and for reward reversal are probably not consistent with a simple conditioning model. These experiments differed from the previous series in that "real" rewards were used, but more important these rewards were contingent on the occurrence of an event of interest (target behaviour of the spinwheel). They will not be further analysed in the present context.

To summarize: the early experiments of Razran showed that temporal contiguity of a pleasant luncheon experience or the presence of unpleasant odours, with neutral stimuli could alter the affective evaluation of these stimuli. The experiments of Staats et al. have shown that positively and negatively toned words transfer their value to neutral stimuli presented in contiguity. Finally, the experiments just described show that direct reward contingencies also impart an evaluative attitude to the neutral stimuli associated with them. In the terms of our formulation, we would suggest that each of these techniques can be regarded as instances of the elicitation of evaluative responses: production of a pleasant "lunchy" feeling; association of the pleasant and unpleasant connotations of emotive words; and the use of direct reward contingent on a fortuitous pairing. They differ from conventional working paradigms in that they each include measures of the change in subjective evaluation. It remains to examine further the effects of eliciting a "pure" evaluative response.

*Studies in evaluative conditioning*

In describing the original evaluative conditioning experiment (Levey and Martin, 1975) it was mentioned that a further series of parametric studies was undertaken and these will now be summarized.* The object of these studies was to examine stimulus and subject parameters which might be expected to have some relevance for evaluative conditioning. The strategy was to isolate two levels of a particular parameter by assigning them to independent squares of the original design. Thus for each partial replication it was possible to hold known factors constant and to assess the effect of varying one aspect of the original procedure. It will be remembered that the design incorporated both forward and backward paradigms, and positive and negative reinforcement, together with a control condition. These stimulus combinations are diagrammed in Table 1. Each of the studies used small numbers of subjects (ten or twenty) and were less concerned to establish the significance of the evaluative conditioning effect, than to dissect out

<table>
<thead>
<tr>
<th>Table 1. Summary of experimental conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Positive Neutral Liked (NL) Liked Neutral (LN) Neutral–Neutral (NN)</td>
</tr>
<tr>
<td>Negative Neutral Disliked (ND) Disliked Neutral (DN) –</td>
</tr>
</tbody>
</table>

*We acknowledge our debt of gratitude to our research assistants: Johanna Beyts, Ruth Hodges, Annette Jedwobski, John Spinks and Maja Turcan, each of whom cheerfully bore the burden of data collection and analysis of one or more of these evaluative conditioning studies.*
Evaluative Conditioning

73

its sources of variation. Nevertheless, the basic finding that neutral stimuli acquire positive or
negative evaluations as a consequence of pairing in the classical conditioning paradigm was
generally upheld.

Before turning to a description of this series of experiments reference should be made to two
aspects of the stimulus combinations which raise some problems of interpretation that are not
yet resolved. The control combination of two neutral CSs (NN) which was expected to produce
near zero ratings has in fact produced rather inconsistent results, tending in general toward
positive evaluation. A good deal of recent work on the effects of mere exposure of unfamiliar
materials (e.g. Berlyne, 1970; Stang, 1974) has shown that repeated exposure tends to produce
liking. This phenomenon was well known to early workers (e.g. Meyer, 1903) in the field of
experimental aesthetics. However, it does not entirely account for the present results which, as
noted, have tended to be inconsistent. There is no reason to expect, for example, that the two
neutral stimuli of the control pair would differ from one another as a consequence of pairing.
Yet this has tended to be the case though the effect was seldom significant. References in the
early literature to the observation that psychophysical judgements of very similar stimuli tend
to be forced apart (Helson, 1951) are consistent with this finding, but again do not fully
explain it. Further investigation of these effects on the control pair (NN) would require a
change of paradigm which would attenuate the strategy we have employed in our parametric
studies. Until such further studies can be carried out it should be noted that the control
differences do nothing to undermine the conditioning result, which depends upon
differentiation of positive and negative evaluations of the appropriately paired neutral stimuli.
Consistent with this, we have found that re-analysis of the data using difference scores based on
the control pair does not abolish but sometimes weakens the significance of the conditioning
effect. Clearly, the results of pairing similar stimuli are complex and include contrast effects,
context effects and the effects of “mere exposure”. To date, however, our results suggest that
while these components require fuller investigation in the examination of the evaluative
response, they are nevertheless subservient to a more consistent conditioning effect.

The second problem is raised by the results for backward conditioning. In the studies
discussed previously (Figs 1 and 2) it may be noticed that there was a fairly pronounced
backward conditioning effect (LN and DN) particularly under negative reinforcement. Again
this effect was not uniform and in one study to be discussed later (see Fig. 5) the negatively
reinforced backward conditioning effect (DN) was stronger than the forward effect. We are not
accustomed in Western psychological circles to thinking seriously about backward conditioning,
and this effect was not predicted. However, the pioneering work of Dostalek (1964) in
examining classical backward conditioning in human and animal subjects over a number of
years has established that it demonstrates certain lawful regularities. One of these is the
consistent finding that backward conditioning is enhanced if both CS and UCS are of moderate
or low intensity, a situation which is characteristic of the visual stimuli used in our experiments.
We have not as yet examined the relative strength of the forward and backward paradigms in
evaluative conditioning and this would again require a change in the basic design of the
parametric studies. However, recent interest, in the West, in the “bidirectional” hypothesis
(Gormezano and Tait, 1976) may stimulate research on backward conditioning and throw some
light on its role in evaluative conditioning. For the present we are unable to specify the
conditions under which the backward paradigm is effective and our interest is centred on
delineating the factors which contribute to the general effects of evaluative conditioning.

The series of parametric studies will now be reviewed briefly without further reference to
the problems just discussed. The first study, already described, involved presentation of either a
300 or 1000 ms UCS. This study showed no significant effect for UCS duration within the
range examined. The next study, also mentioned earlier, was designed to assess the effect of
number of conditioning trials (ten vs twenty) on the establishment of the conditioned
evaluative response. In this experiment, slides of contemporary works of art were used as
stimulus material, and displayed by projector. A significant conditioning effect was obtained
and the treatment X squares effect was also significant (p < .05) indicating that positive
conditioning was less effective and negative conditioning more effective under the ten-trial
condition (Levey and Martin, 1975a). As previously mentioned, an unexpected feature of these
results was that the overall post-conditioning ratings shifted in the negative direction, a finding
which was attributed in part to the effect of Ss’ rating the materials both before and after
pairing. The shift of ratings due to a change in the anchor effect of the first stimulus is well known (Helson, 1964) and may partly explain this result.

The next study was carried out using art photographs of landscapes, with sophisticated and unsophisticated subjects assigned to the two Latin squares (sophistication referring to the familiarity of subjects with the visual arts). Results were, in general, not significant but followed a trend which had already been noted, namely that unsophisticated Ss tended to show the conditioning effect more strongly. The non-significance could be attributed in part to the fact that reproductions of the photographs were used for the stimulus selection procedure (it is easier for Ss to sort through cards than to shift to and fro using the projector) whereas the experiment itself was carried out on slides derived from the original pictures. Two factors emerged: one, that the projected slides were perceived as less interesting than the original photographs; and two, that there was too much homogeneity in the material, all pictures being black and white and of rather similar landscapes.

The next series of three experiments returned to the stimulus materials used in the first prototype study. They were postcard reproductions in colour representing a wide range of "modern" art of a by-now-familiar kind; for example, many were from the Impressionist school. The major factor investigated was the effect of similarity of content and form between CS and UCS. Previous studies had used as CSs pictures selected from the neutral pile which were, in general, of similar form and colour as the UCS. In line with the findings of the prototype experiment where selection of CSs was based on their general similarity to the UCS, a significant effect was obtained. Where the selection of CSs was based on dissimilarity to the UCS the conditioning effect was abolished, while selection of CSs at random attenuated, but did not eliminate, it.

This result was replicated in the next set of three experiments, where again similarity of CS to UCS produced a significant conditioning effect. This set of experiments was designed to test the observation that Ss who respond more on the basis of an immediate "feeling" response condition better than those who respond on the basis of cognitive judgement. Although Ss were routinely told that the experiment was about simple spontaneous likes and dislikes, many consciously responded to the form or structure of the picture rather than with a direct evaluation. A comparison of these "feeling" and "cognitive" groups, identified by a brief questionnaire, showed the conditioning effect more clearly in the former group, for the forward but not the backward direction. Figure 4 shows that, following pairing with a liked or disliked picture in forward conditioning, the neutral card was rated more liked or more disliked by the "feeling" group, but not by the cognitive group.

A final experiment considered the effect of personality, extreme groups being selected along the dimension of introversion—extraversion, using the Eysenck Personality Questionnaire. Extroverts showed better conditioning, as shown in Fig. 5. To some extent this result interacts with the fact that extraverts are more likely to give a direct "feeling" response than introverts, who, conversely, are more likely to adopt an analytical, cognitive approach to the materials. These results on individual differences suggest a complex set of interactions between level of

![FIG. 4. Mean ratings of neutral stimuli following paired presentation, comparing feeling (FG) and cognitive (CG) groups of subjects.](image-url)
Evaluative Conditioning

sophistication, feeling or cognitive style, and introversion–extraversion. They also suggest that the conditioned evaluation in its pure form is susceptible to interference from subjects' attitudes.

![Image: FIG. 5. Mean rating of neutral stimuli following paired presentation, comparing introverts (I) with extraverts (E).]

To summarize the results of these experiments (see Table 2), we have shown that the conditioning effect relates strongly to the type of CS (similar or dissimilar), with similarity of content and form between CS and UCS enhancing the effect and dissimilarity eliminating it. The nature of the stimulus materials is an important factor; heterogeneity is required, but the inclusion of unfamiliar or bizarre stimulus materials, e.g. avant-garde paintings, disrupts the conditioning effect. Both the factors of dissimilarity of the CS/UCS pair and unusual materials have the effect of a general lowering on the final overall ratings, indicating a tendency to rate more negatively following repeated presentations.

Table 2. Summary of parametric studies of evaluative conditioning

<table>
<thead>
<tr>
<th>Experiment</th>
<th>N</th>
<th>Presentation</th>
<th>CS selection</th>
<th>Materials</th>
<th>Latin square</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>Tachistoscope</td>
<td>Similar to UCS</td>
<td>Art reproductions</td>
<td>1000 CS, 1000 UCS/1000 CS, 300 UCS</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>Projection</td>
<td>Similar</td>
<td>Modern paintings</td>
<td>10/20 trials</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>Projection</td>
<td>Similar</td>
<td>Landscape photos</td>
<td>Sophisticated/unsophisticated Ss</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>Projection</td>
<td>Similar or random</td>
<td>Modern paintings</td>
<td>Similar/random CS</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>Tachistoscope</td>
<td>Similar</td>
<td>Art reproductions</td>
<td>1000 CS, 1000 UCS/1000 CS, 300 UCS</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>Tachistoscope</td>
<td>Dissimilar</td>
<td>Art reproductions</td>
<td>1000 CS, 1000 UCS/1000 CS, 300 UCS</td>
</tr>
<tr>
<td>7</td>
<td>20</td>
<td>Tachistoscope</td>
<td>Random</td>
<td>Art reproductions</td>
<td>1000 CS, 1000 UCS/1000 CS, 300 UCS</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>Tachistoscope</td>
<td>Similar or random</td>
<td>Art reproductions</td>
<td>Similar/random CS</td>
</tr>
<tr>
<td>9</td>
<td>20</td>
<td>Tachistoscope</td>
<td>Random</td>
<td>Art reproductions</td>
<td>300 CS, 300 UCS/1000 CS, 1000 UCS</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>Tachistoscope</td>
<td>Random</td>
<td>Art reproductions</td>
<td>Personality</td>
</tr>
</tbody>
</table>

Questioning of subjects following all experiments has repeatedly confirmed the finding that subjects are unaware of the conditioning effect, many subjects being unable to remember which cards had been in the neutral pile prior to testing. Many Ss tended to attribute changes, in the wrong direction, to contrast effects. Often Ss observations about how they rated the cards bore no relationship to their actual rating behaviour. We have been particularly interested in the relative absence of the contrast effects which subjects intuitively expect. Picasso once remarked, however, that "in a room full of good pictures, even a bad picture looks good", suggesting on good authority that contrast is not the powerful perceptual mechanism it is sometimes held to be.

Subjects who say that they responded on the basis of feeling are more likely to condition than those who respond on the basis of aesthetic judgement. Similarly, subjects who are unsophisticated with respect to the visual arts are more likely to condition than sophisticated
subjects. Extraverts show the conditioning effect significantly more than introverts, both with respect to the liked and the disliked conditioning pairs.

To conclude this section, it can be said that the evaluative conditioning effect is persistent across a range of conditions. It is not, however, a robust effect, in its pure form, using the materials and techniques which have been described. Not all subjects are able to give straightforward unmediated reactions without excessive cognitive involvement. Varieties of UCS stimulus materials have yet to be explored, and the relationship between CS and UCS may be more important than has been believed. It should be remembered that these experiments attempted to isolate the evaluative response from active physiological responses usually evoked by the UCS of a conventional conditioning paradigm. The postulate of a subjective evaluative response while important as a causal mechanism may in fact be difficult to isolate. In this respect it may be no different from the postulates of the physical sciences which often refer to conditions which are extremely difficult to isolate except under exceptional circumstances and with a high degree of singularity.

Precursors of evaluative conditioning

The concept of evaluative responding is by no means novel, though it has been widely neglected until recently. It seems probable that the role assigned to motivational variables in the Hullian tradition, and the neglect of the subjective response in the Skinnerian school tended to reduce subjective evaluation to a status of irrelevancy in general behaviour theory. Everyday experience suggests, however, that the balance should be redressed, and that subjective evaluation may be a major determinant of behaviour. Some of the forerunners of this type of work are to be found in the pre-Watsonian literature. In particular, early studies on the perception of feeling or evaluation, and in the definition of attitude are relevant.

McDougall, in developing his comprehensive system of Psychology, paid attention to the role of feelings in the context of psycho-physical judgments. He concluded, from the vantage of his Gothic window, that the simultaneous presentation of stimuli of different feeling tone results in an algebraic summation of the feeling components. This principle, reminiscent of Bentham's Calculus of Pleasure and of utilitarian principles generally, seems not to have had real empirical support in the modern sense. He summarized his views as follows: "When several sensations of pleasurable tendency are present together, their tendencies re-enforce one another; and when sensations of opposed tendency are present together, the opposed tendencies partially or completely neutralize one another. Or, if the pleasurable feeling tendencies be regarded as of positive sign, and the disagreeable tendencies as of negative sign, we may express the facts by saying that the feeling-tendencies of the various sensations simultaneously present to consciousness are algebraically summed, and according as the resultant is of positive or negative sign, the individual feels pleasure or displeasure." (McDougall, quoted by Kellog, 1915.) Interestingly, the notion of algebraic summation as applied to neutral and active stimuli seems to foreshadow our notion that stimulus characteristics are shared in an undifferentiated stimulus complex.

Attention to the minutiae of intrapersonal feelings which the quotation exemplifies, and against which Watson was to rebel, dominated the field in the early decades of this century. One focus of attention, relevant to the method used in our own studies, was the problem of whether feelings are "pure" or occur in combination; whether they are or are not related to sensory impressions.

Kellog (1915) reported some experiments on relatively small numbers of subjects, with the aim of "producing feelings as normal as possible under laboratory conditions and in such a manner as to lead to interference". Picture postcards were used as stimuli, and interference was secured by exposing two pictures, generally one pleasant the other unpleasant, alternately, in a modified form of the Dodge tachistoscope. No attempt was made to adhere to a rigid schedule. The number of trials and the materials used depended upon the tendencies and "success in observation" of the various subjects. Rates of alternation ranged from ten exposures per minute to about one-hundred-and-fifteen. Subjects reported a variety of feelings and were trained to describe them in analytic detail. One of the phenomena identified was the "transference" of feeling. In one instance the two stimuli which were alternately presented were pictures of a
kitten sitting on a pedestal alternating with an unpleasant surgical picture, and the subject reported as follows:

“The human figure was sickening, repulsive. The cat picture came as a relief; it looked snug, cosy, contented. The repulsive figure, however, was stronger and began to lessen the relief derived from the cat . . . .”

In the following example the two pictures were (i) “Sunset at Woods Island” and (ii) “God Help the Poor Sailors”. The subject’s introspective report was somewhat more complex:

“Agreeable picture dominated from outset. Persistence of pleasant feeling during exposure of ugly picture. Involuntarily disregarded what was most unpleasant in picture. Fixed on the neutral features. This tendency stronger with each new exposure, while sunset grew more and more agreeable. The ugly picture eventually grew meaningless, except as a rather annoying halting-place between exposures of the other”.

In the following example the two pictures were (i) “Lovers’ Retreat” and (ii) “God Help the Poor Sailors”. The subject’s report seems to mingle description with inference.

“At first two distinct feelings, (a) pleasant, (b) disgust. The pleasurable tone increased on its own account and encroached on the other. This continued until the pleasurable tone carried over from exposure to exposure, thereby becoming continuous, the disgusting picture coming in as a check or damper on the pleasurable tone. The pleasurable tone grew in two directions, backward and forward, the former as a pleasant lingering tone, and the latter as anticipatory and, of course, pleasant.”

Kellogg rhapsodically summarized his results in what must have been one of the swan songs of the Introspectionist era: “Our results show considerable variety, ranging all the way from inhibition through partial and complete parallelism, with occasional fusion, to increase in cases of contrast. Feelings may increase gradually, change intensity suddenly, with a sort of staircase effect, or stay fairly constant for a considerable length of time . . . . When the fields of appeal become nearly identical, the feelings either become one, or ‘increase by contrast’ in alternation, according as they are of like or opposite value.”

His net conclusion was that simple algebraic summation, following McDougall’s idealized recipe does not occur; rather the picture presented is a kaleidoscope of shifting evaluation. The overall relevance of these early studies may be to remind us that we are still treading deep waters when we again venture back across the skin barrier into the subjective worlds of our patients and subjects. Of more immediate interest is the frequent reporting of the “contrast” effects by the subjects of these studies. It will be recalled that in our experiments, described above, subjects tended to perceive contrasts, and to attribute changes in evaluation of the pictures to contrast effects, although their ratings showed that the objective changes were in the opposite direction.

The interest of European psychologists in introspective phenomena and in the classification of feelings continued long after the Behaviourist revolution in America. At the same time they were aware of the work of Pavlov and interested in the possibilities of conditioning. The result was a body of work which seems quaint by today’s standards but contains fascinating implications for the study of evaluative conditioning. Among the curiosae of the era was Phelan’s exhaustive treatise on “Feeling Experience and its Modalities” (1925). In the present context the work is almost prophetic.

Phelan used an experimental procedure which resembles the aims and methods of an evaluative conditioning experiment. In all, 789 individual “experiments” were undertaken using 7 subjects over a period of 8 months, all recorded in impressive detail. Random sampling was not a feature of the design and the investigator assembled what must have been one of the most distinguished experimental groups in the history of Psychology: three full professors and a lecturer, all “experienced introspectionists” with a control group of graduate students and tutors. The sampling frame covered the university faculties of three European countries. The design was as follows: a variety of sensory stimuli (auditory, olfactory, gustatory, visual, etc.) were first presented singly and Ss were required to dictate a description of their experiences on receiving the stimuli. In subsequent sessions these stimuli were given in pairs, employing a wide range of sensory combinations, for example, odour—taste, taste—odour, odour—sound, taste—sound. The stimuli were paired in such a way that one which had been neutrally evaluated was followed by one rated positively, neutral by negative, positive—positive, negative—negative, and so on. Thus a neutral odour (e.g. guiacol) might be followed by a
negatively rated solution of sulphuric acid. Obviously, the temporal interval between stimuli is not easy to control in the case of olfactory/gustatory stimuli; it seems that sometimes there was a temporal interval, sometimes temporal overlap and, where auditory and visual stimuli were used, the pairs were presented simultaneously, each pair being presented twice.

Detailed introspective reports were given by the Ss on their reaction to these pairs of stimuli, and the investigator distinguished four types of reaction.

1. Isolation  "Where the two experiences remain distinct and separate."
2. Expulsion  "Where the second response interrupts the development of the first, (shifting attention seems to be a factor in this)."
3. Domination  "Where the feeling is so strong to the first stimulus that it suppresses or modifies the second."
4. Coalescence  "In which a total unitary response seems to occur. This was most frequent in the odour—taste groups, especially when taste occurred first."

In these experiments, the aim was to investigate whether the normal feeling response to a stimulus undergoes modification through the preceding action of another stimulus which might arouse a feeling response of the same or different kind. In a second series, an attempt was made to use a conditioning paradigm to establish whether the habitual feeling response to a stimulus could be "permanently" changed. The plan adopted was to present simultaneously with a CS (the one to be modified affectively) another stimulus (UCS) which ordinarily produced a different feeling response, and "to repeat the experiments a great many times in order to create a powerful association between the two sensorial impressions in question".

Interestingly, Phelan described at a common-sense level a phenomenon close to that which we are referring to as "conditioned evaluations": "... This evidently touches upon the question of the transfer of feelings, and it would seem that the experience of everyday life offers enough examples of this phenomenon to render a special research to prove its existence, quite superfluous. Everybody knows that certain pleasant odours, for example, become unpleasant when associated with sorrowful events in life, sickness, or death, for instance; tastes which once were pleasant, become decidedly disagreeable or even revolting when they have been connected at some time or another, with discomfort or nausea. A host of similar examples will occur to the mind of each individual from his personal experience. It seems beyond question that the feeling character of certain impressions may be modified. Yet, even if we suppose that it is firmly established by occasional observations such as these, there still remains the problem of controlling the fact in experimental conditions, and, above all, of establishing its laws experimentally." He speculated, in turn, on our area of interest: "'May it not be possible," he wrote, "to produce conditioned feeling reactions, i.e. to succeed in arousing feeling-reactions with stimuli which were previously indifferent, or nearly so, by submitting the subject to an appropriate preparatory treatment? Here again, daily experience and pathological observation will immediately reply in the affirmative. We have only to recall the sentimental value of certain objects, once quite indifferent, when they have been connected in our lives with pleasant or unpleasant circumstances, separation, childhood days and the like, or in the pathological domain, the case of sexual fetishism."

Phelan specified a list of experimental details which need to be investigated in this connection, and foresaw problems which we have also encountered. What is the role, for example, of sensory modalities, i.e. the differentiable reactions to smell, tastes and sounds? Some sense impressions (e.g. scents) give rise more immediately than others to feeling-reactions which can readily be labelled as pleasant/unpleasant. Reactions to taste and odour are more likely to be accompanied by direct approach/avoidance behaviour than those of sight and sound. Experiments comparing different kinds of sensory stimuli suggest that people react more frequently (in the positive or negative direction) to odours and tastes than to visual stimuli. Early on, however, marked individual differences were observed, both in the frequency of feelings experienced and in the sense that some Ss tend to rate stimuli more positively or negatively than others. Finally it was noted that certain stimuli "go together" naturally, while others remain isolated, and it seems entirely feasible that some principle of "preparedness" underlies the ease or difficulty with which evaluative conditioning can occur.
Underlying these investigations was the dimension of pleasant/unpleasant which Phelan characterized as feeling. "From the introspective reports of our subjects, it is evident that, when they distinguish between certain experiences which they call 'pleasant' and others which they call 'unpleasant', the experiences, as such, have a special aspect. These distinguishing marks of a given experience differ from other characteristics of the same experience, such as colour or pitch, brightness or strength . . . . In analysing the protocols, we have carefully distinguished between judgements of value concerning the responses and introspective description of them. Judgements of value are frequent in the reports of our subjects. They bear on the hedonic value of an experience, as well as upon its aesthetic value. But there is a decided difference between these judgements and actual affective experiences. We restrict to these latter facts, the term 'Feelings', which throughout this work, signifies those responses, exclusively, which are reported as immediately experienced as pleasant or unpleasant."

This difficulty in separating aesthetic and cognitive judgements from feelings was paralleled by the difficulty we have encountered in trying to assess what our subjects experience. "There is no a priori reason," Phelan wrote, "why we should conclude that an individual feels a pleasant or an unpleasant response when he simply states that the stimulus-object is pleasant or unpleasant. Furthermore, there is ample a posteriori evidence to show that judgements of value, whether of hedonic or aesthetic value, are clearly differentiated, by the subjects themselves, from the actual experience of affective phenomena. As a matter of fact, we have reports of experiments in which it is explicitly asserted that no feeling was experienced, yet the subjects formulate judgements of value in terms of pleasantness and unpleasantness. On the other hand, there are numerous cases in which the affective character of the experience is actually felt, and, in addition, a judgement of value is made, which, on the subjects' own statements, quite distinct from the feeling itself, and, sometimes, quite contrary to it." There is perhaps a law of diminishing returns in devoting a great deal of attention to those early studies. Their fascination lies in the similarity of issues which arise in the context of evaluative conditioning. Nevertheless it is tempting to suggest that if behaviour theory is now poised to re-invade the inner world, it would behove us to read the military manuals of our predecessors before, rather than after, the campaign is begun.

It is evident that what we call "evaluation" is largely co-extensive with what others (e.g. Staats) call "attitude", and the remaining group of precursors fall within the domain of attitude theory. Probably the earliest systematic work on attitudes was undertaken by Thurstone (1931) who defined attitude as "the affect for or against a psychological object". Affect in its primitive form was categorized as appetition, or positive affect which in more sophisticated situations appears as "liking the psychological object, defending it, favouring it in various ways"; and aversion, the negative form of affect, described as "disliking the object, destroying it, or otherwise reacting against it". Attitude was used to describe potential action towards the object with regard only to the question whether the potential action will be favourable or unfavourable toward the object.

Thurstone believed that the positive and negative affect constitutes a linear continuum with a neutral point of zero and two opposite directions, one positive and the other negative. Measurement along this affective continuum is of a discriminatory character with the discriminial error as a unit of measurement. This view, which has dominated attitude scaling in most areas of research, encounters difficulties when the situation involves contiguous exposure of stimuli in close temporal approximation. We have found that our subjects sometimes report simultaneous feelings of like and dislike which are reminiscent of the feelings of the introspectionist. At some point in the explanation of attitude conditioning and evaluative conditioning this problem will require solution.

Mention should also be made of the work of Doob (1947, 1969) whose $S-R$ formulation of attitude theory has been influential in the field. Early $S-R$ interpretations of attitude were based on the associationistic principles of repetition and reward. Each attitude was seen as an overt affective response which becomes associated with its own specific stimulus conditions. This view, lacking the conceptual flexibility provided by the later-developed conception of an implicit mediating response failed to account parsimoniously for the diverse phenomena of attitude. In Doob's analysis, an attitude was viewed as an implicit response in its own right -- a hypothetical construct positioned between an objective stimulus and an overt response. The attitude response, although unobservable by an outside observer, is both a response to the
observable stimulus and a stimulus to the observable response in a sort of “chaining” mechanism. Both of these S–R bonds (observable stimulus–attitude and attitude–objective response) were assumed to obey the laws of behaviour theory. He defined an attitude as “an implicit drive-producing response considered socially significant in the individual's society”. By “implicit” Doob meant that the response is not directly observable; it may affect overt behaviour, but it is not overt behaviour. The response “may be conscious or unconscious, distinctly verbal or vaguely proprioceptive”. The influence of these formulations is to be seen in the work of both Osgood and Staats. The importance of regarding an attitude as a response is the mainspring of this system. Whether it may be desirable to distinguish between a phasic evaluative response and a tonic attitude is a question which is relevant to our own formulation. The possibility is open that an evaluative “response” triggers an evaluative “state” which on successive recurrences develops into what might be called an attitude in the usual non-technical meaning of the term.

**Proliferations of the basic concept**

We have now examined a group of prototype studies, in which some form of evaluative conditioning can be identified, or assumed to operate, and we have traced some of the antecedents of these studies in the early literature on the analysis of feelings and attitudes. It remains to select for further examination a group of studies which exemplify, or develop from, the basic prototypes and their precursors. We have employed two criteria in selecting a small range of studies: (1) that explicit reference is made to the subjects' evaluation of the CS; or (2) that the UCS is clearly able to evoke a response of liking or dislike. It should be noted, of the latter criterion, that we are looking for studies which specifically isolate the evaluative component. It is one assumption of our formulation that every UCS, be it shock, corneal air-puff, food pellet or attractive fantasy will elicit an evaluative response which is then conditioned to the non-evaluative neutral CS. The studies to be considered fall into various categories, and carry a variety of labels — “verbal conditioning”, “attitude conditioning”, “emotional learning” and so on. What they have in common is adherence to one or both of the criteria stated above. Interestingly, some authors take the principle we have called “evaluative conditioning” for granted on the basis of plausibility rather than empirical evidence. We are reminded of Phelan's comment that “everyday life offers enough example of this phenomenon to render a special research to prove its existence quite superfluous”. Nevertheless, it must be proved.

Before looking at recent studies, passing mention should be made of the early work of Peters (1935, 1939a, 1939b) who anticipated our interest in the effects of the like–dislike dimension in a series of experimental studies and theoretical papers. This author showed, in a number of experiments, that the perception of stimuli as liked or disliked influences a variety of behaviours, including language behaviour and the recall of foreign words. While he did not specifically employ a conditioning paradigm, being concerned only with a judgemental theory of feeling, his findings can be interpreted as supporting the notion of an evaluative response. Of equal interest is the presentation, through the series of papers, of exhaustive bibliographies of the literature on feeling states up to the time he was writing. These bibliographies should be consulted by any researcher interested in the contemporary analysis of feeling states.

More recent studies fall into two convenient groups determined by the interests of the authors: (1) those which are chiefly experimental and theoretical; and (2) those whose orientation is primarily clinical–social. The experimental studies will be considered first.

Sachs and Byrne (1970) have reported a study in which the conditioning of evaluative responses was assumed, and the experimental manipulation was designed to assess the effect of attitude discrepancies as reinforcers. College students were given a 24-item attitude questionnaire which assessed political and social attitudes and their own attitudes were individually scored. Subsequently they were presented with geometric figures paired with compatible and incompatible attitudes in terms of their own responses to the questionnaire. Incompatible attitude statements produced negative evaluation of the figures, while compatible attitudes produced positive evaluations. The assumption of the authors was that incompatible attitudes arouse negative affect, and compatible attitudes positive affect. The choice of
meaningless geometric figures was designed to ensure that subjects did not merely attribute the incompatible attitude to the paired stimulus, a finding demonstrated earlier using pictures of same-sex strangers as neutral stimuli. This study differs from the evaluative conditioning paradigm only in that the mechanism of reinforcement is assumed to be an implicit affective response rather than an evaluative response in the sense in which we have defined it.

A study which has already been mentioned in the context of demand characteristics and awareness was that of Zanna et al. (1970). In this study meaningful words were paired with onset or offset of shock. Semantic Differential ratings showed that pairing with onset produced negative evaluation while pairing with offset produced positive evaluation. Concomitant physiological records showed that the likelihood of conditioning was partly dependent upon the degree of physiological response to the shock UCS. The conditioned evaluation also generalized to words of similar meaning. Studies of this type are of particular interest since the noxious UCS is an adequate stimulus for motor or glandular responses, for which subsequent evaluation of the CS would normally be ignored or regarded as a by-product of conditioning.

The obverse of this paradigm would require that autonomic conditioning be demonstrated to a UCS which does not involve actual noxious stimulation. Such a study was undertaken by Geer (1968) who showed that random pairing of tones with photographs of violent death are effective in producing SRR conditioning in the absence of other noxious stimulation. Needless to say, we would regard the role of the UCS in both of these studies as that of eliciting an evaluative response.

Of the family of studies which has extended the work of Staats, those which extend the parameterization of the original paradigm are of interest. DiVesta and Stover (1962) demonstrated both higher-order conditioning and semantic generalization to be features of attitude conditioning. It is of interest that this prototype experiment appears to conform to the requirements of classical conditioning criteria in the hands of a variety of investigators, although the problem of subject awareness remains unsolved. In the experiment being considered the first and second order CSs were nonsense syllables and nonsense figures respectively, while the UCSs were evaluatively toned words and thus the effect cannot be attributed only to verbal-cognitive effects.

Finally, in reviewing selected experimental studies, we cross the border between experimental and clinical-social interests with a study by Masters and Santrock (1976) on the effect of verbal self reinforcement on the persistence of task oriented behaviour in children. This was a series of five experiments in which children were presented with a repetitive game involving the control of lights by means of a handle-turning task. Ss were instructed to imagine pleasant and unpleasant events, in the sequence of interest, and it was found that the evaluative tone of these images affected maintenance of the behaviour in the predicted direction. Task relevant and irrelevant images were equally effective in determining this effect. Task relevant self instructions such as "this is fun" had the predicted effect. The interest of this finding, though it is unrelated to classical conditioning as such, is that the task irrelevant images can be assumed to have evoked an evaluative state which operated to influence behaviour as effectively as specific self regulating monitoring of the task. This buttresses the notion that some form of evaluative mechanism is the fundamental underlying mechanism in reinforcement.

Studies oriented directly towards the clinical use of evaluative mechanisms are relatively few in number, though the use of techniques based on evaluative responding is probably widespread. A well controlled study by Hekmat (1972) derives directly from the theoretical formulation of Staats, discussed earlier. The study concerned counterconditioning of anxiety and will be discussed in some detail as a prototype of the clinical application of attitude conditioning.

Hekmat used a procedure of "semantic desensitization" in which phobic subjects employed a repertoire of words depicting positive evaluative experiences (vacation, gift, joy, etc.) which have a meaning antagonistic to anxiety. Two basic techniques were examined: semantic visualization and verbal emission. In the visualization technique individuals were asked to conjure up a series of controlled visual images in response to pleasant connotating words, e.g. "vacation", following the stimulus word, e.g. "spider". In the verbal emission technique Ss were requested to say aloud the positive word rather than to form images. Pre–Post difference scores were obtained on semantic differential scales, on a behavioural avoidance test (in the presence of a huntsman spider) and on a Fear Survey Schedule (FSS). Having conditioned the evaluative
response, the behavioural consequence which followed included rating the spider more positively, less fearfully (FSS) and approaching the spider, a behaviour not trained but following as an open-ended consequence of altered evaluation.

It is important to note that this technique is not merely a second-order (or second signalling system) effect mediated by language, if we accept that evaluative conditioning has occurred. An evaluative conditioning interpretation of the semantic desensitization procedure is that it is a semantic technique for arousing a subjective evaluative response, i.e. an alternative to eating sandwiches (Razran) or relaxing (Wolpe). Whereas Hekmat concentrates on the reinforcing property of words however, we concentrate on the nature of the subjective experience or feeling of evaluation which is, we believe, the reinforcer. In this context it would be interesting to explore the use of a loud positive verbalization: “Spider — I like it”, to see whether this has the same effect. If so, this might imply a closer relation to elicitation of the evaluative response. Conversely the elicitation of negative evaluation might have the same effect, i.e. produce a subject who disliked the spider but no longer feared it.

Whatever the outcome of these speculations, it is apparent that the role of verbal mediations cannot easily be separated from the evaluative state which they induce. Where Staats and his colleagues stress the potency of verbal reinforcers we would prefer to focus attention on the evaluative response itself, regarding the verbal reinforcer as a trigger. The fact that we obtain similar results for “pure” positive or negative evaluations, independent of the use of words, supports this view.

The same argument applies to the work of Cautela (1970) who uses covert reinforcement as the agent in desensitization and other procedures. Patients are encouraged to produce a list of situations which have strong positive value for them individually. These are then associated in imagination with the feared object or event. (A refinement of the technique uses a standardized survey of reinforcing events and situations in the form of a questionnaire, Cautela and Kastenbaum, 1967.) This technique, well known to, for example, hypnotherapists and sergeants-major, has been formalized and given theoretical status by referring it to the principles of reinforcement and S-R learning theory. In our view, such theoretical elaboration is unnecessary if it is assumed that any stimulus event, whether external or imaginary which evokes a positive or negative evaluative response can serve to classically condition that evaluation to previously neutral or negatively evaluated stimuli.

The relevance of evaluative attitudes to social psychology has long been recognized and a number of workers have explored the implications of attitude theory in this area. Kaplan (1975) has extended this thinking to person perceptions in a way which can be readily assimilated to a conditioning model. In a well-controlled study he showed that trait adjectives, as words applied to people, are effective in their evaluative, but not in their denotative aspect. While the controversy to which this study was directed does not concern us here, the results again suggest that a simple mechanism of evaluative responding has a pervasive influence on learned behaviour.

Similarly, one of the well known phenomena of recent behaviour studies, that of vicarious conditioning, seems to be more simply explained by an evaluative conditioning model than by the rather complex theoretical speculations sometimes advanced to account for this puzzling phenomenon. Venn and Short (1973) exposed children to films in which a model manifested negative or positive emotional responses to toy plastic figures. An ingenious test situation was devised in which operant conditioning was employed to assess S’s response to the stimuli before and after vicarious conditioning. The result demonstrated the expected vicarious conditioning effect. One advantage of this technique was to eliminate modelling as an explanation of the effect, since the response required in the operant test situation bore no resemblance to the behaviour demonstrated in the film. The authors do not attempt a theoretical explanation of the vicarious conditioning and the experiment was designed mainly in order to demonstrate that this type of conditioning occurs. In terms of the evaluative conditioning model, however, this type of experiment exemplifies what we mean by open-ended responding. The child confronted by an adult responding with fear to a neutral object experiences a negative evaluative response, and this becomes the basis for its subsequent behaviour toward that stimulus. While this interpretation is speculative, it could readily be tested.

Finally, attention is drawn to a promising therapeutic technique described by Beech et al. (1971) in a preliminary report of a single case. These authors were led by the problems
Evaluative Conditioning

associated with aversion therapy in the treatment of sexual disorders to devise a classical conditioning procedure which would eliminate them. The method they chose was virtually identical to the method used in the evaluative conditioning procedure. The patient was a young paedophiliac and the treatment consisted of exposing pre-rated photographs of sexually mature women for 5 sec (CS) followed by exposure of comparable photographs of immature girls (UCS) for 30 sec. The subject was instructed to create a fantasy in response to the UCS exposure. Penile plethysmography served as the dependent measure and the authors note that “penile responses were successfully conditioned to all of the pictures”, and that this was accompanied by a diminution in sexual deviance. It is noteworthy that the result was considered due to conditioning of the penile response, and this was the purpose of the procedure. It makes equally good or better sense to suggest that what was conditioned was a positive evaluative response, of which the penile response was a concomitant.

In summary, a selected series of studies drawn from a variety of substantive fields has been reviewed, for which we feel that an evaluative conditioning model offers a parsimonious account of the underlying mechanism. While the studies cited are merely illustrative, it would be possible to re-design experiments in which the evaluative component of responding is explicitly controlled. The merit of such an approach would be that it can be referred to a paradigm whose economy and effectiveness are well established, and whose general principles are well known — that of classical conditioning. It is to be regretted that simple explanatory models are currently in danger of being abandoned in favour of elaborate or vaguely defined theoretical formulations, which sacrifice the overall generality of the basic principles of behaviour theory. As novel treatment procedures are developed the tendency to invoke ad hoc theories, or merely descriptive models increases. We applaud, then, the plea of Nawas (1970) in favour of “the canons of science rather than personal conviction and predilection” in the interpretation of behaviour.

THE NATURE OF THE EVALUATIVE RESPONSE

Having proposed the evaluative response as a primary determinant of behaviour it becomes necessary to attempt to clarify its nature. This involves an attempt to describe its attributes and functions; it also involves considering its near neighbours and making some comparisons and discriminations from other positions. In this section some of the implications of the concept of evaluative responding are considered in relation to other dimensions of behaviour, and other ways of construing evaluation. The evaluative response can be differentiated from the cognitive appraisal system of Lazarus (1966) from the hedonic system of Young (1955) and from the sensory-reinforcement system of Wyrwicka (1975) although it has relationships with all of these. It can also be differentiated from motivational systems such as that of McClelland (1955) and from physiological treatments of evaluation exemplified in the work of Arnold (1960). Each of these approaches will be considered in relation to the concept of the evaluative response developed here.

First, however, an attempt will be made to define the evaluative response more explicitly, and to examine its origins in and significance for adaptive behaviour. It may be useful to summarize what has been said in the previous sections. The evaluative response is seen as a characteristic reaction to environmental stimuli in terms of the evaluation of good/bad, liked/disliked, pleasant/unpleasant. The response is defined as subjective in the sense that it is unique to the individual organism. It is distinguished from physiological and emotional changes, from sensory experience, from explicit motor acts of approach and avoidance, and from cognitive—judgmental analysis. At the same time it will be evident that on most occasions the evaluative response will be associated with elements of all of these. It is described as “truly” subjective to indicate that it cannot be defined in terms of sensations, autonomic arousal or motor behaviour, although it can be inferred from these. Finally, it is important to emphasize the definition of evaluation as a response since it is on this basis that it can be assigned a primary role in conditioning.
Evaluation as a response

There can be no doubt that all organisms evaluate their surroundings in ways which substantially influence their behaviour. In common parlance we say that some plants "like" the sun while others "prefer" the shade. Kittens "love" to be stroked but they "hate" water. These expressions reflect our recognition that the like/dislike dimension of experience is universal. However, it is the response-like character of evaluation which needs to be examined. To qualify as a response a segment of behaviour must be a reaction to changes in the external or internal environment; it must have an onset and offset; and it must have a dimension of amplitude. Each of these criteria can be seen to apply to the concept of evaluation.

We say that plants "like" the sun because they turn toward it, and this response to an external object occurs without sensation in the usual sense, and without emotion or cognition. Most animals withdraw from intense heat, and do so because it is unpleasant, disliked, negatively evaluated. Yet the fact that animals and men can be trained or motivated to endure unpleasant stimulation suggests that the response of withdrawal is not the same as the "response" of disliking. The same argument applies to the avoidance or postponement of pleasurable experience. The evaluation which is precipitated by an external event or a change in the environment has an onset, when the object or event is encountered, a dimension of amplitude reflected in the intensity of liking or disliking, and an offset when the object of evaluation is removed. There seems, therefore, to be a strong case for regarding the evaluation as a response.

Such a response precedes the behaviour to which it gives rise, and the emotion which follows from it. It also occurs without cognitive judgement. Everyone has experienced a sudden flush of pleasure on seeing a beautiful flower, a pretty face or an attractive child. We do not first make the cognitive judgement "this is good for me", rather we enjoy it in the moment of perceiving, and the reflection that the pleasure is good comes after. It is this unmediated character of "pure" evaluation which most clearly qualifies it as an independent response. It has been recognized by writers such as Arnold, who is primarily concerned with its emotional consequences, "An object or situation is immediately appraised as liked/disliked" (Arnold, 1960); and by those whose interest is in cognitive events "the process is often nearly instantaneous... an individual need not be fully aware of the evaluations he is making or of the factors that enter into them" (Lazarus, 1965).

The theories of Arnold and of Lazarus in particular, exemplified in the quotations given, are often classified under the heading "cognitive", and this presumably implies that the coping behaviour follows the perceptual analysis of stimuli. It could be argued, consonant with these quotations, that in many instances, before a stimulus or event has "meaning" for us, it is detected instantly in relationship to personal wellbeing. Hence the response can be regarded as a kind of "welfare reflex" which includes elements of "it's nice/not nice for me", "I want more/less", etc. Such an evaluative response is essentially of biological origin, a mechanism for recognizing hostility or hospitality in the environment to enable the organism to survive.

Thus the response of evaluation can be proposed as a meaningful unit of subjective experience and as a primary determinant of conditioning. According to this view actively liked or disliked stimuli can function as UCSs without evoking physiological or behavioural responses. However, this postulate of a subjective evaluative response while important as a causal mechanism may be difficult to isolate. In practice, as a mechanism for responding to and coping with environmental change it is likely to occur concomitantly with a pattern of physiological and behavioural activity.

Probably the most characteristic behavioural aspect of evaluation is that it gives rise to motor movements of approach or avoidance. As soon as we evaluate something positively, we feel an attraction toward it. As soon as we evaluate negatively we feel repelled from it, and urged to avoid it. This theme recurs in the literature of approach-avoidance behaviour, and has a long history. One of the earliest attempts to formulate a systematic account was that of Troland (1928). This theory reiterated the close connection between pleasantness and unpleasantness and biological advantage and disadvantage. His view was that pleasantness is accompanied by movements toward the object of attention, and unpleasantness by movements away from it. Lewin (1952) distinguished positive valences effecting approach and negative valences producing withdrawal or retreat. However it would be a mistake to equate approach
Evaluative Conditioning

and avoidance behaviour with evaluative responding in a one-to-one relation, as has already
been suggested.

Young (1961) studied the relation of hedonic tone to types of movement by having his
observers report motor experiences accompanying pleasant and unpleasant stimuli. He
concluded that unpleasantness was associated in a large number of cases though not always with
tendencies to “react away”, “to put the stimulus away from oneself”; pleasantness, however,
was seldom associated with seeking movements. In general, work in this area led to the
conclusion that the typical involuntary reaction to pleasantness is a relaxation and not a seeking
reaction. Thus there is further evidence that evaluation represents an independent response.

An interesting and difficult aspect of evaluative responding is the problem of sensory input.
It has been suggested earlier that organisms are likely to have a repertoire of innate evaluations.
How are these mediated? It is possible that “pre-wired” evaluations depend directly on qualities
of the sensory receptors. Rejection of bitter tastes is widespread throughout the animal
kingdom and may have originated in the Cambrian period with the first coelenterates (Garcia
and Hankins, 1975). That conditioned aversion to bitter flavour can be obtained in
sea-anemones has been demonstrated and may represent the earliest demonstrable form of
evaluative conditioning. In this context we recall Razran’s observation that simple conditioning
(i.e. classical, Pavlovian conditioning) “was already in full bloom on this planet some
500,000,000 years ago in organisms with no cortex, very little forebrain and, indeed
comparatively, not much brain altogether”.

Clearly, the majority of sensory processes in higher organisms are plastic with respect to
evaluative responding as are the behaviours which evaluation provokes. Nevertheless the
possibility that some sensory input is inherently evaluative cannot be dismissed, just as some
withdrawal reflexes are apparently innate. It is possible to define a hierarchy of evaluative
responses from the simplest innate evaluations to those in which both input and output are
“open-ended”. An interesting early attempt to speculate on the nature of evaluation in receptor
systems was that of Troland in the reference quoted earlier. This theory aimed to provide an
objective view of the processes of evaluation without reference to introspective concepts. He
discussed the problem in connection with biological systems which he classed as nociceptive,
beneceptive and neutroceptive.

“We may be tempted to define a nociceptive sense channel as one, the normal excitation of
which produces unpleasantness in consciousness; whereas a beneceptive channel could be
specified with reference to the pleasantness which normally accompanies its operation. But
such a method of classification would lay us open to the charge of employing a psychological
criterion . . .” Instead “ . . . the criterion which we must actually adopt in classifying the various
receptive systems as nociceptive, beneceptive, and otherwise, must apparently be of a general
biological nature, depending upon the relationship between the given receptive process and the
welfare of the species. By welfare, we merely mean the best chances of survival, referring not at
all to such concepts as progress or happiness. On this basis, we may define a nociceptive system
as one which responds specifically to stimuli that are injurious to the organism or the species,
wheras a beneceptive system may be characterized as one which is especially sensitive to
beneficial stimuli” (Troland, 1928, p. 204).

Troland continued his discussion in relation to the various sensory systems, which, as he
pointed out may be nociceptive, beneceptive or neutroceptive, depending upon circumstances.
He concluded that “the neuroceptors may be regarded generally as recording the environment,
and the relation of the organism thereto; whereas the nocic- and beneceptors transmit to the
cortex constant reports concerning the effect of these factors upon the welfare of the individual
or species”. While the neurological basis of these speculations may have been dubious his
formulation recognized the problem of the role of sensory input in generating evaluative
responses.

The argument to this point has concentrated on regarding evaluation as a response. However,
the existence of evaluative states must also be recognized. Everyone has experienced the sort of
mood in which the whole world seems pleasant, and the black mood in which nothing seems
enjoyable. It seems possible that in certain circumstances a strong evaluative response could
trigger an evaluative state, and this is consistent with ordinary experience. Interestingly,
however, what seem to be enduring endogenous states are seen most clearly in pathological
conditions and in altered states of consciousness, where they are not triggered by a response.
The notion of evaluation as a primary component of behaviour is consistent with euphoric/dysphoric experiences, for example, those found in manic-depressive patients. Negative evaluations of the self, the environment and the future, are some of the main features of depression. These individuals commonly overemphasize negative evaluations to the relative exclusion of the positive. By contrast, the manic phase brings pervasive positive evaluations which colour the patient's view of his entire universe, and significantly influence behaviour.

Similarly, the drug induced "highs" and "lows" of the hallucinogens may be construed as pervasive evaluative states or as states in which the frequency of evaluative responses is greatly increased. These states can be seen as supporting the view that some form of unmediated subjective evaluation is a major determinant of behaviour. Interestingly, therapists who have used LSD as an adjunct to therapy are strong in the belief that learning is enhanced during the positive phase of the experience. Members of the soft drug culture believe that the experience leads to enhanced social learning. These convictions are clearly not in the category of hard evidence, but again support the notion of a central role for evaluative processes in behaviour and in behaviour change.

To summarize, it has been suggested that evaluations of external and internal events can be regarded as independent responses, unmediated by cognitions as such, and having primacy over subsequent behaviour. Indeed the three systems model of emotionality as comprising partially independent systems of cognitive, behavioural and autonomic responding can be viewed as a relatively open-ended consequence of primary evaluations. Given that an evaluative response occurs, the past history, present circumstances and native temperament of the individual would determine how it is handled by each of the three systems. Such a model emphasizes the adaptive character of the evaluative response, and this will be examined next.

The adaptive nature of the response

It has been suggested that a primary determinant of behaviour is the organism's capacity to recognize harmful stimulation as "unpleasant" and beneficial stimulation as "pleasant". This fact is so obvious that it sometimes escapes notice. In order to survive in the natural environment an individual or species must be equipped with some sort of evaluative responses. An animal escapes from injurious stimuli not because they are injurious but because they are painful. It is not necessary to postulate consciousness or even to invoke a sensory system in order to establish a mechanism of evaluation. For lower organisms with a relatively fixed repertoire of behaviour the subjective evaluation and the external approach/avoidance behaviour are intrinsically linked. For man, and the higher animals, subjective evaluations can lead to a variety of behavioural patterns. Fundamental to the repertoire of behaviours, however, is the evaluative response which initiates them. The response can be regarded as a kind of welfare reflex which includes elements of "it's good/bad for me", "I want more/less" etc. Such an evaluative response is essentially of biological origin, a mechanism for recognizing hostility or hospitality in the environment to enable the organism to survive, and it need not operate at a "conscious" level.

If such a mechanism were omitted, what would the organism need to survive and operate successfully? Presumably a system of hedonic feeling would be adequate to respond to the environment as it is, but an inadequate mechanism for detecting when it is becoming hostile. Perhaps a number of different detectors could be installed; to detect when heat is too strong, air insufficient, light too bright, tastes too noxious. To a certain extent this has happened, especially in the internal environment at a biochemical level. Chemoreceptors in the carotid bodies detect deficient oxygen levels in the blood and adjust the cerebral blood flow accordingly. The liver can detect toxic substances and deal with them appropriately, over a wide range of biochemical processes. In this sense the whole evaluative machinery is working internally without sensations, hedonism or consciousness. There is a natural division between the internal environment which cannot be escaped, and the external environment which can be approached or avoided. In the latter case there has to be a mechanism for detecting external noxious factors, and nature has provided ways of coping with many of these.

But to multiply the sensory mechanisms required to cope with every external eventuality would imply an organism too complicated to function. It is a more parsimonious approach to
keep the specific mechanisms and avoidances to a minimum and to build in another mechanism which enables the organism entering a novel environment to identify the warning signals of hostile events and to learn to anticipate danger. Suppose, for example, animals could be provided with a mechanism to prevent burning, a mechanism, say, that would ensure it avoided red objects. This would inadvertently ensure that it avoided red-coloured food. A better mechanism is that which allows it to be burned once and then to evaluate red flickering as a warning of a stimulus to be avoided, i.e. to provide for the conditioning of an evaluative response. The essential strategem is to equip the organism with a minimum basic repertoire of instinctive and reflex reactions, and also with a mechanism for detecting signals in the environment which are relevant to its well being, while leaving open the response pattern (feeling, thinking, emotions, motor activities) required to accomplish adaptive coping behaviour.

The point need not be laboured further. Recognition of some form of hedonic evaluation has formed an implicit or explicit component of most theories of behaviour, and is a common theme in folklore. In the present context it is argued that the biological function of evaluation can usefully be regarded as a response in its own right, that the response has an adaptive basis in that it enhances the probability of survival, and that its adaptive significance is increased by the fact that it can be conditioned to relevant stimuli in the environment.

Motivational components of the response

Having reviewed the concept of evaluation as a response and commented on its adaptive significance it remains to examine some alternative formulations in the literature. Of these perhaps the most important are theories concerned with the role of motivation in behaviour. The behaviour theoretic literature of the past decade has demonstrated a return of interest in the role of hedonic processes in behaviour. It seems probable that the dominant position of the drive concept in learning theory focused attention away from hedonic processes in the explanation of motivation. However, a number of authors maintained an interest in evaluation as an explanatory mechanism, and the eclipse was not total.

The history of hedonism in relation to theories of learning and motivation has been reviewed by Young (1961) and Cofer and Appley (1967). Young's own impressive series of experiments were carried out on food preferences in rats (Young, 1952). Arguing that positive and negative processes constitute a form of primary motivation in that they arouse, sustain and direct patterns of behaviour, he believes that through affective experience the organism learns preferences, builds up attitudes of liking and disliking, motives and stable value systems. The importance of affective processes as contemporary events within the organism was emphasized in his system, and he consistently used terms like preference, palatability and acceptability throughout the period when S–R formulations dominated learning.

According to this essentially hedonic theory, summarized in the Nebraska Symposium of 1955 (Young, 1955) the organism behaves so as to maximize positive affective arousal (delight, enjoyment) and to minimize negative arousal. In the same symposium Olds described how a rat's behaviour could be steered by brain stimulation, and proposed that the reward measured as self-stimulation interacts with other neural functions to provide the affect or "pleasure" experienced by the animal. These sets of ideas have generated an active body of research into brain reward and aversion systems.

A large area of work has been carried out on the hypothalamus and feeding behaviour, and an important aspect of this research is its emphasis on affective arousal in controlling the pattern of approach and withdrawal. The results suggest that there are brain reward and aversion systems which together finely control approach/avoidance behaviour. The concept of a reward-aversion mechanism postulates that stimuli generate approach or withdrawal according to the combination of physiological signals within the animal, not just the strength of peripheral stimulation. This mechanism apparently senses the physiological state as signalled by energy and osmotic related factors and then links food-related stimuli to complex approach responses when the animal is in need, and to complex escape responses when surfeited. An interesting extension of the work to human subjects showed that the taste of a sucrose solution that Ss rated as "pleasant" shifted to "unpleasant" after a large meal of the sucrose. Thus the
reward of sugar gave way to aversion as a result of post-ingestional factors, just as self-stimulation gave way to stimulation-escape in rats (Hoebel, 1974). Hoebel uses the term "homeostatic hedonism" to apply to the realm of feeding in humans, with the interesting rider that ". . . it is more than hedonism because it may well apply to the behaviour of simpler animals to which few would attribute pleasure or displeasure". His reward-aversion theory of behaviour control is postulated to apply to a wide variety of phenomena ranging from the approach and withdrawal behaviour of the fly to the subjective pleasure and displeasure of humans. Such a view would seem to imply a common underlying mechanism of evaluation, with the interesting implication that the direction of evaluation of external stimulation is a function of the contemporary physiological state of the organism.

Wyrwicka (1975) has more recently emphasized the hedonic (experiential) aspects of motivation, and reward and punishment in a Pavlovian framework. Reward consists of obtaining desirable sensations while in avoidance behaviour it consists in the disappearance of undesirable sensations. As an expression to describe the communality of these events she suggests an "improvement in the sensory state" which acts as a reward in each case. An improvement in sensory state is understood as a subjective experience which can be described in such colloquial terms as "I feel better", or "better-being". The expressions "improvement in the sensory state" or achievement of "better-being" imply that it is not necessarily full satisfaction but only some degree of it which is obtained as a result of instrumental action. It seems, therefore, that these expressions are closer to reality than, for instance, the expression "well-being" which implies an obtaining of full satisfaction. Wyrwicka's views are related to ideas based on the principle of natural selection according to which the chance of survival is much higher in those populations in which pleasurable sensations are "correlated", in the Pavlovian sense, with actions beneficial to the organism and in which disagreeable feelings of pain are correlated with actions destructive to the organism.

Working in the context of learning theory, Silverstein (1973) has pointed to the extremely important distinction between "pleasure" and "incentive", noting that they are not necessarily related. His argument, important to the concepts of hedonic theory, cannot readily be summarized; briefly he suggests that "pleasantness" is defined by subjective experience, while "incentive" is defined by motor behaviour. This formulation, presented in detail in the reference quoted, offers an attractive alternative to earlier forms of hedonism. Similarly, newer conceptions of hedonism (e.g. Hoebel, 1974) contain sophisticated views on the role of hedonic components in the homeostatic regulation of behaviour.

These references to hedonic theory suggest that interest in the concept is reviving, as noted earlier. The purely hedonic position differs from our own in that its principle emphasis is on the motivating and steering functions of pleasant or unpleasant experience.

At least one formal theory of motivation placed a major emphasis on the role of subjective evaluations in coordinating behaviour with the environment. McClelland (1955) regarded affective processes, particularly those associated with pleasantness/unpleasantness, as having an objective existence which could be described in physiological terms. In describing the genesis of motives this thesis (1955) was that we are first of all presented with cues in affective situations, for instance, sugar is put in the mouth and this produces a pleasurable affect. This type of cue then becomes paired with an affective state in such a way that the cue will, as a result of association, come to "redintegrate" the affective state first associated with it. Behaviour gradually comes to be coordinated and directed towards situations which are associated with positive affect and away from situations associated with negative affect. McClelland used the concept of affective arousal as Hull used the concept of drive — to explain both the acquisition of motives and their activation once acquired. This view differs from the evaluative conditioning concept chiefly in terms of the level of analysis, being concerned with the development of complex motivational systems. Both views hold that what is learned is the affective relevance of cues in the environment.

It should be noted that both the hedonic theories and the motivational theories just discussed are logically distinct from reinforcement theory and this is also true of the evaluative conditioning concept. Thorndike's formulation of the Law of Effect which underlies modern reinforcement theory stated that specific acts are "stamped in" by the experience of annoyers and satisfiers. In contemporary terms, the occurrence of an evaluative response influences the frequency of occurrence of the behaviour associated with it. By contrast, the concept of
Evaluative Conditioning

Evaluative conditioning is that the evaluation is transferred to the stimuli in contiguity with the evaluative response, leaving the appropriate behaviour undetermined in whole or in part.

The notion, just stated, that what is learned is the "affective relevance of cues in the environment" inevitably invites comparison with cognitive theories of evaluation. It is argued here that the organism learns to make an evaluative response to a previously neutral stimulus, and that this occurs without cognitive mediation. The distinction, noted earlier, is between knowing that a given stimulus signals an impending event and responding toward the stimulus as if it were the event itself. The former is cognitive and the latter is not, though the distinction may be very fine in practice. Nevertheless, it is an important one, since the two positions imply quite different models of behaviour.

Processes of evaluation or appraisal are increasingly discussed in the context of cognitive factors and the concept of cognition may be assuming some of the roles previously assigned to the concept of motivation. Arnold, for example, refers on occasion to her system of appraisal as cognitive although she very frequently expresses the view that appraisal is immediate and intuitive. In a recent volume on feelings and emotions (Arnold, 1970) she discusses the ways in which psychological experience, affectivity and physiological changes might interact, referring for example to Kohler's Gestalt point of view that feelings and emotions are integral parts of perception. It is probably this intimate association between perception, evaluation and cognition in human experience that makes the problem difficult.

For extensive treatment of the subject of cognition and perception in the conditioning context one must refer again to Razran (1971). The underlying hypothesis of his book is that learning is an evolved and evolving ascending hierarchical system of levels of continuity and novelty, and he cites a vast amount of data to support his view. Razran discusses a type of conditioning which perhaps comes closest to our view of evaluative conditioning under the heading "reanalysis of sensory preconditioning" (1971, p. 252). The core of his analysis is that what is learned in sensory preconditioning is more perceptual than sensory, and he describes this as a perception of simple and obvious temporal sequences and spatial contiguities. It is learning not only without "insight" but also without its traditional rival, "trial and error" "...it is indubitably a perceptual learning, although low in the package.... Learning to perceive that a dark cloud is followed by rain or that food or a predator - or prey - is in a particular location is as modal a phylectic acquisition as learning that a stick may rake in food (when the organism is equipped to do it); that a disappearing food box may reappear; or that a correct pathway to food may be a combine of the learning of two prior pathways...." In brief, Razran is arguing for a non-cognitive interpretation of kinds of learning that are often considered to be cognitive.

One of the active writers of today in the areas of stress and coping behaviour, Lazarus (1966) makes use of concepts of appraisal and reappraisal, and assigns to them a cognitive significance. His research is concerned largely with the role of cognitive mediators in emotion, and frequently introduces experimental variations to alter the mediating cognitive processes. Every emotion is said to "flow from appraisal processes by which the person or infrahuman animal evaluates the adaptive significance of the stimulus". A further concept is that the person or organism must be "regarded as an evaluating organism, one who searches his environment for clues about what he needs and wants and evaluates each stimulus as to its personal relevance and significance" (Lazarus et al., 1970). We would suggest that these concepts do not, in fact, imply cognitive mediation in a necessary role. Nevertheless it cannot be denied that the notion of cognitive interpretation of evaluative responses merits serious attention. It is obviously difficult to disentangle what we call the evaluative response from cognitive elaborations of it, and from cognitive judgements per se. The suggestion is offered however that unmediated evaluative responses can occur, and that they are separate from cognitive appraisal. Indeed, it has been noted that one difficulty of the "pure" evaluative conditioning paradigm is to ensure that subjects avoid cognitive appraisal.

In summary, it is suggested that the concept of evaluative conditioning is logically distinct from those of hedonism, of motivational theory and of cognitive theory. It is further suggested that the distinction carries interesting implications for conditioning theory and for the analysis of behaviour.
Finally, in attempting to define the nature of the evaluative response, it is necessary to consider its subjective components. The evaluative response has been described earlier as subjective in the sense that it is unique to the particular individual. It is "truly" subjective in the sense that it cannot be identified in terms of sensations, autonomic arousal or motor behaviour. Nevertheless, if it is to serve as a dependent variable in conditioning studies it must be identifiable. In human subjects the evaluative response can be reported as the experience of liking or disliking, provided precautions are taken to exclude cognitive judgements. Techniques such as the Semantic Differential or the Kelly Grid may prove useful in by-passing merely cognitive appraisal. Interestingly, Osgood (1971) has pointed out that the three Semantic Differential dimensions of evaluation, potency and activity although primarily semantic could equally well apply to the most primitive organism when confronted with a novel stimulus. It has to decide how pleasant/unpleasant, how active and how powerful the stimulus is in order to adjust its motor output in terms of approach or avoidance. In the present context, however, we are concerned with the identification of the evaluative response in a pure form.

In this connection there is a large early literature concerning the introspective analysis of hedonic tone. Two fundamental views are in conflict; one (Wundt and Titchener) believed hedonic tone to be an attribute of special affective elements which were different from sensations. Another school (Ziehen and Stumpf) denied the existence of special affective elements, believing hedonic tone to be an attribute merely of sensations. In their view hedonic tone never occurs independently but may occur with images as well as with sensations. When it occurs in conjunction with an image it is very easily transferred to other conscious contexts. "When I have suffered an accident at a certain place, a hurt, for instance, at future times not only will the recollection of this accident be accompanied by unpleasantness, but frequently the recollection of the place itself will be spoiled, i.e. also accompanied by unpleasantness" (Ziehen, quoted by Beebe-Center, 1965). Ziehen seems to have come close to describing a kind of conditioning of hedonic tone through the juxtaposition of images, a procedure not unlike the use of "emotive imagery" in the treatment of children's phobias described by Lazarus and Abramowitz (1962) and by Cautela (1970) in his description of covert reinforcers. These techniques use as reinforcing stimuli clear images which are perceived as highly desirable, pleasurable or enjoyable, without cognitive intervention.

Whatever the means of inducing strong positive or negative evaluations we must still depend upon the subject to report his evaluation whenever we attempt to deal with pure evaluations, or with evaluative conditioning. No specific physiological pattern of responding has been associated consistently with positive or negative evaluation. Obviously strong evaluations occur in states such as fear and revulsion, and hence physiological reactions are likely to be frequent concomitants. A typical definition of an emotion, for example, is "... an experience in which the person appraises the object as affecting himself... as the felt tendency toward anything intuitively appraised as good (beneficial) or away from anything intuitively appraised as bad (harmful). This attraction or aversion is accompanied by a pattern of physiological changes organized toward approach or withdrawal" (Arnold, 1960). However, at a less intense level of evaluation, i.e. that which does not involve emotional arousal, there is no reason to suppose that peripheral autonomic activity or somatic–muscular activation is involved. Nevertheless there is little doubt that subjective evaluations are experienced within awareness, and the problem is how to monitor them.

The subjective experience of evaluation is an internal event which can be analysed and reported. But we have little experience of analysing these private events. It seems likely that if we were to do so, some training would be required to distinguish internal evaluative sensations from tensions and autonomic events. In this respect, introspective observation is no different from other observations. Any student who looks for the first time at protozoa under a microscope will hardly know what to look for, and may not "see" anything at all. And if untrained observers look at X-rays, little meaning can be attached to light and dark areas. However, the use of introspection carries difficulties which are well known, including the problem of how to use introspective data. A common suggestion is that we use them not as direct evidence about the occurrence of conscious contents but as a verbal report about them.
In this view it is "merely" language behaviour, and its relationship to processes within the individual is irrelevant. Others argue for a kind of isomorphism between subjective experience and physiological activity, such that private events can be made public by means of electrophysiological instruments which "detect the same events as do the organism's internal receptors" (Hefferline and Bruno, 1971).

Probably the most feasible approach is to use verbal self-report as a reflection of an internal state, without the associated corollary that it could be validated in terms of electrophysiological measures of that state. Nowlis (1970) and Thayer (1967), for example, have used mood check lists to represent general bodily activation and emotional reactions, and Thayer has suggested that self-report may be an integrative variable more representative of the general state of bodily activation than any single peripheral physiological measure. Davitz (1970) has accumulated a pool of phrases that people use to describe their subjective experience of emotions. Such authors accept that this form of self-report extends beyond the merely verbal and represents an attempt to describe inner feelings. As we have suggested earlier, the existence of well standardized scaling techniques opens the door to the use of subjective report as a reliable dependent variable in conditioning studies.

To summarize, the evaluative response is necessarily subjective, but its measurement may be important to the understanding of conditioning. It has been suggested several times that the evaluative response may carry the basic mechanism of conditioning. In any case, the observation of subjects' evaluations of conditioning stimuli add a new dimension to the study of conditioning. The usefulness of defining subjective evaluation as a response is partly strategic in the sense that it enables the development of an approach to conditioning that has interesting implications for behaviour theory. In the last analysis the usefulness of the strategy must be judged in terms of these implications.

**IMPLICATIONS OF EVALUATIVE RESPONDING FOR BEHAVIOUR THEORY**

The preceding sections have considered some of the issues and problems associated with the role of subjective evaluation in classical conditioning, and have reviewed some of the experimental literature pointing to the importance of such evaluations. It remains now to examine some of the implications of evaluative responding for behaviour theory. The studies quoted have the effect of bringing the concept of evaluation to the centre of the stage, where it can be seen to play a major role in the conditioning process. It is probably fair to say that in most of the classical conditioning literature, if evaluation is considered at all it is regarded as a by-product of the conditioning process. In our own experiments we have attempted to determine whether a "pure" evaluation can be conditioned and have demonstrated that this phenomenon exists.

In this final section the implications are explored of placing evaluation and the evaluative response at the core rather than at the periphery of conditioning. The suggestion is made that it may be possible and useful to consider the evaluative aspect of responding as the carrier of the central mechanism of conditioning, able to account not only for simple forms of conditioning but for the more complex and subtle forms of conditioning which are involved in covert behaviour and in subjective states. We are aware of the dictum of Razran (1936, p. 335): "To say that one's learning is no more than mere conditioning is more of an insult than a theory", and we do not propose to fall into this pitfall.

We are equally aware of the intricacy of human response systems and of the intimate relations pertaining among the components of behaviour that we call cognitive, affective, attitudinal, and so on. In these terms, we will expect that pure evaluative conditioning will be a rare event, frequently or usually overlaid and camouflaged by these systems. It is not our purpose to suggest that a simple or simplistic mechanism can account for the complexities of human behaviour with which the clinician must deal. Rather, we are suggesting that to examine a simple mechanism may throw some light on the operation of these more complex systems. Before turning to this discussion, however, it may be useful to summarize briefly the literature discussed earlier.
Resume of studies

The early studies of Razran form the first link in a chain of argument leading to the concept of evaluative conditioning. In these studies it was shown that the temporal association of food or agreeable music produced positive shifts in political and social attitudes, which Razran attributed to a conditioning effect. While, in this context, he hovered between concepts of cognitive conditioning, the role of attitudinal set as awareness, and the possibility of perceptual conditioning, it is interesting that he seems to have been the first to suggest that what is involved in classical conditioning is the underlying modification of general affectivity. The development by Staats of the attitude conditioning paradigm carried the argument a step further. This paradigm demonstrates the potency of language reinforcers in producing positive and negative shifts of evaluation by a process of classical conditioning. In terms of the model suggested here, we would regard these language reinforcers as evoking a positive or negative evaluative response, and would construe the attitude conditioning paradigm as the transfer of this evaluative response to the previously neutral material.

The demonstration by Nunnally and his colleagues that reward values can be directly transferred advances the argument yet further toward the concept of evaluative conditioning, and introduces the element of open-ended responding. The studies of Homme, Cautela, Sachs and Byrne, among others, introduced the use of subtle covert responses as unconditioned responses, and carry the phenomena of interest into the realm in which we are dealing with unobservable responses whose reinforcing value, in the Pavlovian sense, can be inferred. Finally, our own experiments have attempted to use merely the subject's evaluation as a UCR, and this represents the evanescent but nevertheless persistent phenomenon that we have called evaluative conditioning. While the term evaluative conditioning will be used throughout this section, we are arguing that the role of evaluation in conditioning should not be assigned merely to another category of conditioning but should be examined as a possible core of the conditioning process.

The historical thrust of these studies and of the argument is toward the recognition of an evaluative response as a basic response in human behaviour. We have been struck by the pervasiveness in human behaviour of evaluative responding - however defined - which becomes apparent when the concept of evaluation is used in an explanatory way. To quote Osgood and McGuigan (1973, p. 455): "What is important to us now, as it was in the age of the Neanderthal, is the good or bad for us of the things signified by signs . . . ." While it may be argued that individuals differ in their readiness to evaluate the objects and events about them, it is certainly the case that most of us make spontaneous evaluations and base a part of our behaviour on them. It is also evident that these evaluations summarize information, as Arnold (1970) clearly recognizes in her concept of "affective memory". It is probably a common experience for most people that events in the past, a good meal, a vacation, a theatre performance, and so on, are recalled in the first instance as pleasant or unpleasant events without necessarily remembering the contents or the details of the event. We remember that the Bistro was atmospheric, but not exactly what we ate or drank. An interesting analogy is the disconcerting situation when a person tries to tell a joke and is unable to remember the punch-line. He remembers that it was funny, the positive evaluation, but he does not remember how it was funny.

The central role of evaluation in classical conditioning

The argument, then, is that evaluation occurs in all conditioning situations, where it is usually considered a by-product. The fact that evaluation is assigned a subsidiary role, if it is assigned any role at all, may probably be traced to the conceptualization of motivation as a drive factor in traditional theories of learning and conditioning. An exception can be found in the work of Lewin (1952) whose summary discussion of learning defined two basic forms. His concept of cognitive learning dealt with the acquisition of information, while his concept of motivational learning seems to have been the only serious attempt in learning theory to recognize the importance of changes in motivation. These changes in motivation were construed, within the terms of his theory, as changes in valence and were attributed among
other factors to modification of goals. With this exception, the role of evaluation in learning theory has been that of a motivator of behaviour, and this role precludes the notion of evaluation as being the thing that is learned.

We would argue that the organism is equipped with a set of native evaluations which enable it to secure a favourable correlation with its environment, in order to ensure survival. While it is self-evident, it is worth noting that even the simplest organism must contain a built-in mechanism for "recognizing" extremes of stimulation as harmful. The simplest plants and animals are able in a variety of ways to avoid noxious stimulation and to maximize favourable conditions. At the lower phylogenetic levels, the evaluations are simple and the organism is limited by its repertoire of evaluations. The classic case of the moth and the candle-flame is an instance of a native evaluation gone wrong, and points to the inefficiency of a fixed repertoire of positive and negative evaluations. An organism which is able to form new evaluations of unanticipated stimuli in its environment has an enormous advantage in individual survival.

In other words, a mechanism for transferring native evaluations to previously neutral stimuli, that is, a mechanism of evaluative conditioning, would be a powerful adjunct to continuance of the species. Indeed, such a mechanism would be considerably more powerful than the model usually proposed of the acquisition of a motor response to a previously neutral stimulus. Whatever the representation of the external environment, a topic well beyond the scope of this paper, it is necessarily placed within the organism. What we are saying, in these terms, is that a mechanism of evaluative conditioning gives the internally represented external world its adaptive meaning and significance.

Chomsky (1959) in his well-known review of Skinner's theory of language ridiculed the notion that "a person can be reinforced though he emits no response at all" and suggested that the concept of reinforcement becomes meaningless. Since we are speaking here of reinforcement in the Pavlovian sense of classical conditioning, rather than in the sense of drive reduction, we can readily accept the notion. Instead of being a devastating criticism of the classical conditioning concept of reinforcement, it may be an excellent summary. While for lower organisms the evaluative response is confined to approach or withdrawal, it is self-evident that in human behaviour an internal evaluation can have reinforcing properties, as the experimental studies have shown, though no response is "emitted" in the sense implied by Chomsky. By labelling this internal evaluation as a response, and suggesting that the response can be conditioned, we preserve the power and economy of the classical conditioning paradigm. That the response may be relatively inaccessible raises problems of measurement but does not detract from the potential organizing force of the concept.

It is important to note that the selection of literature reviewed earlier was designed to stress the role of such an evaluative response. It should not be forgotten, however, that the evaluative response also occurs in those simple forms of conditioning in which the experimenter is interested only in a motor response or a glandular reaction. It may be likely, though it is admittedly difficult to test empirically, that the laboratory conditioning situation actively constrains the subject to perform only the behaviour of interest to the experimenter. This observation has often been made (e.g. Prokasy, 1962) but a different emphasis is placed on it here. It can be suggested that because of this constraint, the apparent conditioned response is an artefact of the situation rather than the behaviour which is conditioned.

Of the many apocryphal anecdotes to issue from Pavlov's laboratory, one is particularly apposite. A dog, in the course of daily salivary conditioning sessions, was tested on one day free of its constraining harness. On presentation of the CS, a metronome click, the dog approached the instrument, wagged its tail, and licked it enthusiastically. We are not told whether the animal also salivated, but this would appear to be a clear-cut instance of evaluative conditioning. Indeed, every pet owner knows that the production of a food dish is not primarily the cue for mastication nor is the production of a leash the cue for initiation of walking behaviour. Rather, both these CSs produce joyful leaping and other signs of positive evaluation.

This aspect of animal learning has been widely neglected until recently, the outstanding exception being the work of Ploog (1970) using objectively categorized and analysed vocalizations in monkeys, and showing the powerful expressive effect of positive and negative affective signals. The discovery by Brown and Jenkins (1968) of the phenomenon of "auto-shaping" has corrected this deficiency, and has led to a substantial literature on the
tendency of animals in the classical conditioning situation to approach or avoid a CS source which signals positive or negative reinforcement. The approach and contact movements, reminiscent of Pavlov's dog described above, occur whether or not the UCS requires a similar motor response. While the focus of this monograph is on human conditioning, the recent work on auto-shaping in animals is highly relevant to the concept of evaluative conditioning. The literature has been well reviewed by Hearst and Jenkins (1974) and by Schwarz and Gamzu (1977), and will not be considered here. Nevertheless the occurrence of expressive movements, and movements of contact and avoidance in animal conditioning permit the inference that there are reasonable grounds for regarding the transfer of evaluative responding to previously neutral stimuli, the redefinition of these stimuli in non-cognitive terms, and the modification of subsequent evaluations as basic processes in conditioning rather than as peripheral side effects.

It is of interest to consider one further implication of this view for the role of reinforcement in classical conditioning. In this context it is important to remember that the mechanisms of classical conditioning are still ill-understood and open to question. It is interesting that a veteran Pavlovian of the stature of Horsley Gantt was recently led to speculate (Gantt, 1973) that what is conditioned in aversive conditioning is the sensation of pain, rather than the response of withdrawal. While he dismissed this speculation in a series of arguments, the fact that he entertained it at all emphasizes the point just made, that the role of reinforcers in classical conditioning remains an area of intriguing mystery. The effect of suggesting that the evaluative response is conditioned is to reduce reinforcement to a single dimension. This meets a criticism offered by Breger and McGaugh (1965, pp. 346–347) which may have some validity, though we are unable to endorse the grounds on which it was proposed. “The hard job, but the only one that will lead to any meaningful use of the concept of reinforcement is specifying what the various events or reinforcers have in common.” The context of this observation, intended as a criticism, is their suggestion that the only solution is to compile an endless catalogue of reinforcers with the implication that this sort of inexhaustible repertoire cannot explain behaviour. What the variegated reinforcers have in common is positive or negative evaluation, and this is an idea worth considering if only in the interest of parsimony. We turn now to a discussion of some of the implications, and, hopefully, some of the advantages of this formulation for behaviour theory and for behaviour therapy.

**Conditioning models in behaviour therapy**

We regard it as a matter of regret that the behaviour therapy movement is drifting inexorably away from its original theoretical foundations. These foundations were laid out early by Eysenck, on the one hand, and by Skinner on the other; and it is not an unjust criticism to suggest that the building has subsequently gone askew. The power and economy of the classical and operant models of basic conditioning processes probably constitute the strength of the behaviour therapy approach and the basis for its repertoire of techniques. Recently, however, there has been a drift toward the introduction of rather vaguely defined cognitive concepts (ironically offered as an alternative reformulation of behaviour theory by Breger and McGaugh in their early criticism of behaviour therapy, and stoutly rejected by the pioneers of the movement), and toward a proliferating array of internal responses in the form of images, fantasies and self-instructions, uncritically given the status of responses and carrying with this status an illusory gain in objectivity. We do not propose to criticize the methods which emerge from these formulations, where they are effective. We wish to suggest, rather, that some of the methods might be profitably re-examined in terms of the concept of an evaluative response, and that this concept genuinely retains the role of a response, as the foundation of classical conditioning mechanisms, with undiminished potency.

We would suggest then that it is worth our time as behaviour theorists to explore the possibility of reducing the growing number of mediating reinforcements, postulated for heuristic purposes, to a single manageable concept, that of evaluative conditioning. It would be rash to suggest that this strategem will produce miracles of increased effectiveness, but we suggest that it could serve to clarify the field, in terms of its theoretical foundations. At the same time, it may be argued that some advance in technique would be gained if the result of such a strategem were to be that the core mechanisms with which we deal were correctly
Evaluative Conditioning identified. It would be interesting and practical to base a clinical behaviour analysis on identification of the patient's hierarchical structure of evaluations, and to do this explicitly rather than by implication. In other words, and in summary, it may be possible to reduce the number of subjective factors to a central ideational and methodological core, with the hope that if the core is correctly identified there will be an ensuing gain in technique. The history of clinical practice in many areas has often demonstrated this development.

It would be prematurely speculative to suggest that the concept of evaluative conditioning may have direct relevance to the treatment of neurotic behaviours. The use of an evaluative conditioning technique by Beech et al. (1971) in the successful treatment of a single patient has already been mentioned, and we would hope that procedures of this sort might be further explored. It can be suggested in passing, however, that some neurotic syndromes may represent failures or misdirections of the evaluative response. The indecision of the obsessional patient may fall in the former category, while failures of sexual identification may fall in the latter. The acquisition of conflicting evaluations may underlie the origin of some neurotic behaviours, for example phobias. Further work on the identification of evaluative mechanisms will be required before direct clinical applications are immediately feasible. It is one of the purposes of this monograph, however, to suggest that further work may prove rewarding.

The problem of the generality of behaviour

A purely theoretical advantage accrues from the formulations outlined in the preceding sections, in attacking a problem which is fundamental to the application of behaviour theory to clinical practice: how is the specificity of learning translated into the generality of behaviour? We will argue that the conceptualization of an evaluative response, as the core mechanism of conditioning, offers a satisfactory answer to this problem. Before examining the alternative answers and formulating the answer which the evaluative concept implies, it may be useful to summarize once again what is meant by evaluation and by the conditioning of evaluative responses.

In introducing the concept of evaluation we are not referring to peripheral receptor events nor to the single dimension of pleasantness/unpleasantness. Rather we are referring to a central response which occurs in the context of changes in the environment favourable or unfavourable to survival. It is defined as immediate, in the sense that to make an evaluation does not require cognition or affect, and this is inferred from the presence of evaluative responses in the lowest species. Nevertheless, as noted earlier, we are well aware that human behaviour involves many dimensions of responding and have noted that these may camouflage the evaluative response.

The suggestion that this response is neither affective nor cognitive is based in part on the argument from phylogeny just cited. In addition, however, our experiments have shown that evaluations can be conditioned which have no affective component as far as we can detect. Similarly our finding is that cognitive appraisal acts as an interfering mechanism rather than as a component of evaluation. The term cognitive is this sense refers to deliberate judgement, but it is also important to note that a part of the parsimony of the concept of evaluation is that it excludes the necessity of postulating awareness of means—end expectancies as well. Further, it is suggested that the evaluative response is available to subjective report and can be reliably identified, though it may be the case that subjects need to be trained to recognize it. Indeed, much of the current literature on behaviour therapy is curiously preoccupied with the induction of pleasant and unpleasant internal states which may be subsumed under the single dimension of an evaluation. The work of Barendregt (1973) and Ramsay (1974) in training emotional responses, their recognition and their facilitation, provides an analogy to the training of evaluative responses.

It is suggested, then, that the evaluative response can be conditioned, and further suggested that evaluation is a necessary component of what is transferred during the process of classical conditioning. While this claim rests on speculation, buttressed by our experimental explorations, it is not a difficult claim to accept and indeed is probably accepted in the folklore of conditioning. A number of authors write as though they take something in the nature of evaluative conditioning for granted, although it has not been finally and satisfactorily demonstrated beyond all doubt. This suggests that the concept has a heuristic and theoretical
range of convenience which makes it acceptable. We merely argue that the concept could be placed in the centre of the stage as the chief protagonist in the drama of conditioning. In summary, then, the concept of an evaluative response which is not a receptor event but a central response, immediate and non-cognitive as well as non-affective, and which can be identified by subjective report on a reliable basis, can be assigned the role of principal mediator in conditioning with some theoretical advantages. Chief among these is the answer it offers to the problem of the generality of behaviour and to this problem we now turn our attention.

How do specific learning events translate into the generality of behaviour? Misunderstandings of behaviour theory by its critics have been based on the historically outdated notion that conditioning deals only with a particular response to a particular stimulus. It is on this basis that Breger and McGaugh (1965) posed the question as a criticism of SR theory. Nevertheless, the question does pose problems for traditional behaviour theory, and unsuccessful attempts to answer it date back to the time of Watson. It has been an arena of conflict both for clinicians involved in behaviour theory methods and for those involved in more personalistic or dynamic schools, as well as between them. In the therapeutic setting, it refers to the transfer of learned behaviour outside the consultation room. While we are sympathetic to the formulations of SR theory in general, the move towards identification of increasingly global stimulus constellations and response frameworks, in the context of SR associationism, eventually becomes unconvincing.

Atkinson and Raynor (1974) revive a distinction, made earlier by Atkinson in his analysis of motivational theory, between a general group of theories characterized as Habit X Drive, which include the Hullian and neo-Hullian formulations, and a category characterized as Expectancy X Preference. It is of interest that this distinction has assumed a new relevance with the work of Seligman and Johnston (1973) in reviving a rigorous cognitive theory based on postulates of expectancy and preference. Indeed, cognitive theorizing has regained its popularity, in particular among behaviour therapists. The issues involved in choosing between cognitive and behavioural models of learning have recently been re-formulated by Dulaney (1974), who is disposed to favour the cognitive models. The antecedents of these models are the cognitive theories of the 1940s, eclipsed in the following decades by the SR theories. Interestingly, the early cognitive theories have also given rise to a group of information processing theories, having a similar general format, but taking their terms of reference from computer technology. These two theories, or groups of theories, the cognitive and the information processing, appear to offer answers to the problem under consideration: the transfer of specific learning to general behaviour.

Before considering these alternatives we should first examine a more sophisticated formulation of the problem. Mandler (1975), in a more philosophical treatment of the problem, imagines a situation in which a spectator sees a child threatened with danger and he asks what it is in the history of the spectator that makes him respond emotionally to this situation. The interest in Mandler's way of putting this question is that he clearly recognizes the deficiencies of a model which requires that every element of a response be associated with some previous stimulus in the history of the organism. Mandler writes: "If we were to ask how a particular individual comes to be in a state of autonomic arousal when presented with the sight of a child being nearly run over by an automobile, then it is difficult (though some learning theorists sometimes make the attempt) to describe the variety of prior conditionings that might have occurred in order that this particular set of events should produce the release of autonomic discharge. It is highly unlikely that the intricate concatenation of previous experiences have always been favourable in just such a way as to produce the 'conditioned emotional response' that occurs. It is in fact the very fine grained analysis of classical conditioning and the precise relationships among unconditioned stimuli, conditioned stimuli and unconditioned responses that these studies have shown to be necessary for the substitution paradigm to work, that argue for the implausibility that those conditions are met in the conditioning of autonomic responses in everyday life." (Underscoring added.)

Mandler is attacking a straw man, since no behaviouristically oriented psychologist would today wish to pursue and defend the challenge he offers, though there have been elaborate attempts to do so in the past. The essence of today's situation in behaviour theory is that it is recognized that something analogous to a central state can be learned and that such a state can then exert a distal influence on subsequent behaviours in a vicarious way. This suggestion, first
Evaluative Conditioning has been widely ignored until recently. The cognitive theorists of the 1940s, particularly Tolman and Krechevsky, adumbrated such a view but were unable convincingly to demonstrate or defend it in the climate of the times. It seems probable that the advent of the computer as a familiar model, with its central processing activity taking the place of the homunculus, and its emphasis on programs and strategies as mediational events, has been the main factor in making central theories currently acceptable. Whatever the reason, we are now presented with the two general theoretical approaches just mentioned, viz. the cognitive framework and the information processing framework, both of which have descended directly from cognitive (S-S) learning theory, based on the formula Expectancy X Preference.

The use of these two categories is not arbitrary but is based admittedly on convenience. No attempt will be made to define them, other than in descriptive terms, and it should be pointed out that the semantic quibble which asks whether a particular behaviour is "really" cognitive is simply that - a semantic quibble. It should also be made clear that we are not talking about theories as such, but about two approaches which subsume a large number of conflicting theories. It can also be noted that there are theories which span both categories, and it would be interesting to analyse these into their propositional frameworks, in order to see whether the categories are in fact exclusive. More detailed discussions of the issues have been offered recently by Mahoney (1974) and Razran (1971). The present concern is to consider the two approaches, and to compare them with the approach formulated here, in terms of the problem of the generality of behaviour.

The cognitive theorists, in particular Irwin (1971), suggest that behaviour is governed by its preferred outcomes, and the organism learns to anticipate the outcomes that it will prefer. This formulation lies somewhere close to a hedonistic view, though, of course, it is not simply that, and its main direction is to ask how the expectancies are erected given that preferences govern behaviour. A cognitive theory implies that the organism understands its world, and understanding is taken to mean prediction of outcomes. It may be useful however to examine the concept of cognition in slightly more detail. It is not always clear exactly how the term is being used. Sometimes it is restricted to the category of verbal reports of subjective awareness of CS-UCS contingencies. At other times it seems to be accepted that a cognitive view can readily encompass the non-verbalizable, even the unconscious. We see no difficulty in extending the concept of cognition to this realm, provided that the terms of the theory are rigorously defined, as has been done, for example, by Seligman and Johnston (1973) in their treatment of avoidance learning. Operationally, the definition of cognition need not imply conscious awareness; rather the term hinges on the possibility of defining expectancy in a non-circular fashion.

Central to the cognitive theories however, by our descriptive definition, is the notion of some form of modelling of the external environment. This modelling may be in the form of means-end expectancies, the older concept of cognitive maps, or more recently the situation-act-outcome model proposed by Irwin (1971). For such a view it must be argued that in the representation of the external world, multiple models are formulated which overlap, interact and integrate in many and complex ways. Some part of these can be consciously assessed in human experience and verbally reported. To take the simplest example, habituation, which can readily be interpreted in terms of expectancy, the contingencies can be verbalized and this conscious knowledge may affect, for example, autonomic activity in the conditioning situation. In such a case it would be argued that the model of the stimulus environment is important in influencing responsiveness.

Without going further, then, it is possible to identify a logically distinct group of theories which hinge on the notion of cognitive expectancy, of understanding the world, and of the construction of models which represent the external world in terms of expectancies. In the view of such a theory the basic element in conditioning is the formation of overall models and expectancies, which describe the current and future state of the organism's environment. For this type of theory, the problem of the generality of behaviour is easy: general expectancies can be learned and the adaptive behaviour which ensures favourable outcomes follows as a natural consequence. It may be trivial to say that this formulation simply begs all the important questions. A more serious objection is that few cognitive theories seem to be aware of a new set of difficulties raised by changes in preference. How such changes come about is one problem,
sometimes met by saying blandly that they are "acquired" by classical conditioning (e.g. Boneau, 1974). A second problem is that of specifying the modifications required of the model's system of expectancies in the face of altered preferences. Until these problems are met the cognitive solution to the problem of the generality of behaviour must be regarded as tentative.

The approach of the information-processing theories is logically distinct from that of the cognitive theories in the mechanism they suggest for behaviour, though parallels between the two groups are apparent. Such theorists as Grant and Broadbent, arguing from widely differing positions, suggest that what is learned is some form of program accessible to a central processor whose job it is to determine priorities in the filtering and processing of input. Pribram (1967) has adopted a similar view in the past, and a number of theorists have defended the notion in recent years. The salient feature of this type of theory is that it distinguishes between reflexes and strategies, on the one hand, and strategies and expectancies on the other. What the organism is learning is a set of strategies or programmes which equip it to deal with input experienced in the past and to process this in terms of output strategies which have been successful. Modifications of these strategies represent modifications of the program and the analogy with computer systems is clear. To emphasize the distinction between this and the preceding group of theories it might be said that the cognitive theorists deal with mapping of the world and its contingencies, while the information processing theorists deal with the elaboration of strategies for performance in the world. Without analysing the propositional frameworks of these kinds of theories in detail it would seem that the cognitive theorists are in a better position to explain novel and vicarious behaviour, while the information processing theorists are equipped to deal with those complex behaviours which form the organism's stable repertoire. The distinction may be artificial however and as is often the case, both general approaches may be correct, insofar as they are concerned with the texture of everyday experience. It might be suggested that both approaches fall short of a fine-grained analysis of that texture.

The problem of the generality of behaviour does not seriously arise for the information processing theories, since it is answered by default. It is easy to imagine program modifications, branching sub-routines and so on, and a drawing board circuitry which will handle these problems. It is much more difficult to demonstrate that these sub-routines exist as some meaningful form of modifiable software, that the neural hardware is well designed for this kind of processing, or indeed that a central processing unit exists at all. Needless to say, these objections are not critical, since theories of this type are not intended to represent physiological reality. But one may question the usefulness of a theory or model which is merely descriptive and self-fulfilling in that it describes a kind of learned introspection conditioned to the computer jargon of our era.

The evaluative conditioning principle is offered as a valid alternative. This approach suggests that what is primarily learned is some form of central evaluative state triggered by an evaluative response to a salient stimulus, in the presence of a previously neutral stimulus which is specific to the stimuli exposed, but which leaves the response pattern in some degree open-ended. Such a view suggests that the organism possesses a number of innate evaluative responses to specific stimuli, which have survival value in avoiding harmful stimuli and approaching stimuli which are beneficial. It suggests that further evaluations are learned, and that these in turn may be transferred to neutral stimuli. Such a model would postulate an n-level learning process in which the acquisition of an evaluative response would be the primary mechanism. It can be postulated that evaluative responses vary in the extent to which they are pre-wired to specific behaviours. Thus the dislike evaluation of intense heat, for example the legendary stove which the burned child learns to fear, would contain its own predetermined motor component. At another level evaluations may involve a hierarchy of behaviours, so that the organism selects the appropriate behaviour in the light of other elements of the situation.

The evaluative conditioning model makes room for novel behaviours and for vicarious responding, on the basis that a given positive or negative evaluation will result in different decisions or action patterns on different occasions. That this analysis freely makes use of the operation of cognitive processes, as the term applies to everyday experience, is readily admitted. The emergence of a stragedium, based on previous experiences, is consistent with the model proposed. The formulation has the advantage, however, that it retains the power and
generality of the classical conditioning paradigm in an explanatory role which the other two approaches lack.

It is evident that none of these three general models is mutually exclusive, in terms of a functioning organism, although it is convenient to keep them logically distinct. Thus an evaluative reaction might call an appropriate program strategy in line with the information processing model, or may contribute to modelling of the environment in terms of the cognitive theoretical position. It remains to ask which of these three general models offers the most appropriate strategy for research into the genesis of complex behaviour. The cognitive theories offer obvious problems in that their definition of cognition tends to be difficult to make specific, and it is hard to imagine where the locus of a cognition might lie. The information processing models are equally vague in their specification of the location of programs and strategies. It is easy to imagine behaviour as a series of block diagrams and arrows but difficult to coordinate these with the functional units of behaviour.

The viewpoint put forward here is that the evaluative conditioning mechanism offers a more simple preliminary model of learned behaviour than do either the cognitive or the information processing frameworks. This view retains the parsimony of the conditioning paradigm on which behaviour therapy was originally founded. We see a danger in uncritically admitting to the theoretical framework of behaviour therapy an increasing number of loosely defined mediational concepts, and suggest that the role of evaluative mechanisms, and particularly of learned evaluations, can serve as an organizing core concept in the analysis of complex behaviours. In these terms we would champion the heuristic and theoretical gain to be realised by placing evaluation in a central role, rather than regarding it as a by-product of adaptive behaviour.

Acknowledgements - We gratefully acknowledge the support of the Bethlem and Maudsley Research Fund (to I.M.) and the Medical Research Council (to A.B.L.).

REFERENCES
