LEARNING AND MOTIVATION **29,** 397–415 (1998)
ARTICLE NO. LM981012

The Elusiveness of Evaluative Conditioning

Paul Rozin, Amy Wrzesniewski, and Deidre Byrnes

University of Pennsylvania

Evaluative conditioning (EC) is an important variant of Pavlovian conditioning in which the outcome is a change in affective response to the conditioned stimulus (CS). It is the best extant account, with evidence, for affective change in humans. Good laboratory models are available. This paper reviews a set of findings which suggest that the actual occurrence of EC, and its magnitude, varies widely in both real world and laboratory situations. Attention to known parameters of Pavlovian conditioning may account for some, but not all, of the failures and successes. Six empirical studies on humans are described; two document frequent failures of EC to occur in real world situations where the Pavlovian conditions for development of animal phobias or taste aversions are present, two are real-world experiments in which no evidence for EC is obtained, and two are laboratory failures to produce EC, by pairing neutral odor CSs with a variety of unconditioned stimuli (USs). We suggest that there are important, not understood factors, that modulate the appearance of EC, and that for both theoretical and applied reasons, these factors should be identified.

Evaluative conditioning (EC) is usually conceived as a variety of Pavlovian conditioning, in which the unconditioned stimulus (US) is an elicitor of an affective unconditioned response (UR), and the ''conditioned response'' (CR) is a change in affective response to a conditioned stimulus presented contingently with the US. An alternative description focuses on valence and change in valence, rather than the more general term ''affect.'' The term, evaluative conditioning, was introduced by Martin & Levey (1978), and the principal basic paradigm was designed and studied previously by these same authors (Levey & Martin, 1975). Since the ''responses'' in evaluative conditioning are affective states, the literature has focused on humans, the only organism that can report on affective states, such as liking of disliking for an object. The phenomenon is presumed to occur in nonhumans, but is hard to measure without self report. Conditioned taste aversions in rats and humans are, according to most criteria, an example of evaluative conditioning (see discussion below).

Evaluative conditioning (EC) may well be pervasive in humans, and is

Send correspondence and reprint requests to Paul Rozin, Department of Psychology, University of Pennsylvania, 3815 Walnut St., Philadelphia, PA 19104-6196.

0023-9690/98 $25.00

currently the only articulated, experimental model of how likes and dislikes are created (Rozin & Zellner, 1985). The laboratory investigation of EC has been carried out, systematically, by a group of experimental psychologists at the University of Leuven in Belgium, under the leadership of Frank Baeyens (e.g., Baeyens, Eelen, van den Bergh, & Crombez, 1988; 1990; 1991). In their research, the Leuven group has used two laboratory exemplifications of EC: one is a contingent pairing of rather neutral pictures, usually faces, with attractive or unnattractive faces. This paradigm derives from the original work by Levey and Martin (1975). A second paradigm (deriving from work by Zellner, Rozin, Aron, & Kulish, 1983) pairs neutral flavors with unpleasant tastes (Baeyens *et al.,* 1990).

Using these two laboratory models, the Leuven group has shown that: (1) robust EC can be produced after five or so contingent pairings, (2) EC responses are highly resistant to extinction (Baeyens *et al*., 1988), (3) EC works well with simultaneous CS-US presentation (Martin & Levey, 1978; Zellner *et al.,* 1983; Baeyens *et al.,* 1991), and (4) awareness by the subject of the contingencies is unrelated to the success of conditioning (Baeyens et al., 1991). The latter finding argues against a demand characteristic account of EC.

This set of findings has led to a conception of EC as a somewhat different type of Pavlovian conditioning, in which the CS functions less as a signal for the US, and more as an acquired affective attribute of the US (Levey & Martin, 1983; Baeyens *et al.,* 1991; 1995). Garcia (1989) provides an alternative framing for EC, distinguishing it from traditional Pavlovian paradigms as a form of Darwinian conditioning, in which hedonic changes are produced via interoceptive feedback. Examples of EC, such as phobias and conditioned taste aversions, have also figured centrally in the literature on preparedness (Seligman, 1970), belongingness (Garcia & Koelling, 1966) or adaptive specializations of learning (Rozin & Kalat, 1971).

The most robust and most investigated instance of EC is conditioned taste aversions (Garcia & Hankins, 1974; Rozin & Kalat, 1971). In this case, the US is usually nausea, and the CS is an associated taste or flavor. Conditioned taste aversions (CTA) are of particular interest because they provide a laboratory animal model of EC, since there is an affective (facial) readout available for rats. Two studies demonstrated that, in rats, nausea paired with a taste leads to an acquired distaste, as measured by a shift from a positive to an aversion pattern of facial and bodily gestures, whereas electric shock contingent on the same taste leads to avoidance, but no affective (facial) change (Parker, 1983; Pelchat, Grill, Rozin, & Jacobs, 1983). CTAs have also clearly been identified in humans, both in the laboratory (Bernstein, 1978) and in retrospective questionnaires (Garb & Stunkard, 1974; Pelchat & Rozin, 1982; Logue, 1985).

The extensive animal work on CTAs is important, because it is the most thoroughly studied example of an instance of EC. However, CTAs may be

an anomalous form of EC, because of the long CS-US intervals, rapid extinction, and probable importance of a CS preceding the US. Analogous to the CTA phenomenon are reports from humans of aversion/disgust towards foods associated with a disgusting experience (e.g., opening a box of cereal and finding ants in it; Rozin, 1986).

Another potential example of EC, not normally integrated into the EC literature, is phobias. According to the Pavlovian account of phobias, a CS acquires negative affective value by association with a powerful negative US (Ohman, 1979). Both phobias and taste aversions involve the change in valence of an entity (CS), and hence, on this criterion, are clear instances of EC. However, they may differ from other instances in important ways; since in both cases, simultaneous conditioning may be very weak, there is the possibility that some referential/signal conditioning is involved as well as an evaluative change.

The range of EC in humans has barely been explored. The experimental literature has focused almost entirely on physical events as USs. However, research on acquisition of fear of snakes by monkeys based on social referencing (experienced distress in another monkey as US) points to the potency of social USs (Mineka & Cook, 1988). Two important recent studies from the Leuven group demonstrate EC in humans with social USs (Baeyens, Kaes, Eelen & Silverans, 1996; Baeyens, Vansteenwegen, DeHouwer, & Crombez, 1996). Apparent enjoyment or displeasure at consumption of a specific beverage was indicated facially (on a video) by a ''demonstrator,'' and caused subjects watching the video to show a corresponding increase or decrease in liking for associated cues (glass shape in one study, flavor of the beverage simultaneously sipped by the observing subject, in the other study). We presume that such socially mediated EC is a major component of the acquisition of preferences, and the acquisition of culture. For example, it is very likely part of the way that foods consumed primarily by adults, such as coffee or hot pepper come to be liked by children (see, e.g., Rozin, 1990). Similar stories can be created for activities such as exercise, and for strong likings for anything, including fetishes (Rachman & Hodgson, 1968).

We are convinced that EC is of both great theoretical and practical importance. EC stands as the most articulated model of affective change with respect to objects/events in humans (Rozin & Zellner, 1985). In this paper we raise questions about the robustness, stability, and reliability of this phenomenon, in both real life and in the laboratory. Our interest is not in discouraging interest in EC; on the contrary, our aim is to direct attention to important and not understood modulating conditions that seem to determine whether or not EC occurs. These conditions, when explicated, may be of both theoretical and practical importance.

Looking over our own experience with EC, involving over 20 studies, and many by others, the issue of the unpredictability of the phenomenon emerges. EC is often robust, but sometimes it fails to appear.

The ''classic'' picture or face pairing paradigm, originated by Levey and Martin (1975) and refined by the Leuven group into the definitive laboratory example of EC has produced positive results on many occasions, as evidenced by many published studies from the Leuven group and by others (e.g., Stuart, Shimp, & Engle, 1987). However, Todrank (unpublished), in our laboratory, failed to obtain a significant EC effect using the Leuven picture-pairing paradigm, including the exact stimuli used by the Leuven group in their series of successful studies. Furthermore, in recent years, the Leuven group has experienced a substantial reduction in the size of the EC effect using the standard paradigm. The cause of this is unknown, but the effect reduction was sufficient to cause these investigators to shift to the more reliable flavor-flavor paradigm. The flavor-flavor paradigm was originated in our laboratory (Zellner, Rozin, Aron, & Kulish, 1983), and employed a neutral flavor as the CS, and a sweet taste as the US. The original set of studies involved three separate studies showing the enhancement of liking for a neutral flavor contingently paired with sweetness, in comparison to non-sweet paired neutral flavors. In further unpublished research in our laboratory, in which we attempted to use this paradigm to explore features of EC, the effect in the standard conditioning group (used for comparison with other treatments) was usually not statistically significant. As a result, we abandoned this line of investigation with the flavor-flavor paradigm. Meanwhile, the Leuven group employed a variation of the flavor-flavor paradigm (Baeyens *et al.*, 1990), and reported a substantial EC effect using a negative US, and a non-significant effect in the predicted direction with the sweet US used by Zellner *et al.* (1983). The flavor-flavor paradigm, with the negative US (Tween20 [polysorbate]), has been a steady source of significant EC results in Leuven.

Meanwhile, in our search for another paradigm, a number of studies were done using odor as a CS, and pleasant or unpleasant pictures as USs. These studies, carried out by Todrank, did not yield positive results. However, a procedure in which a Leuven type of procedure was used, using the Leuven faces as CSs, and pleasant or unpleasant odors as USs did lead to significant effects in four studies (Todrank, Byrnes, Wrzesniewski, & Rozin, 1995). In these studies, which constitute a basic demonstration of the effect with three replications, effect sizes varied from very large to modest in size. No obvious difference in the procedures could account for these differences.

This set of results has led us to believe that there are some important and unknown factors that influence the occurrence of EC. Systematic investigation by the Leuven group (e.g., Baeyens *et al.*, 1991), confirmed by some findings from our own laboratory, suggest that neither awareness of the contingency nor strength of US (mild or strong) has a major effect on the appearance of EC.

Convinced of both the pervasiveness and importance of EC, we bring together here six studies we have done that we hope will form part of a body

of research that will clarify the essential conditions for the appearance of EC. Some of the elusiveness of EC may not be unique to EC; rather, it may be a feature of Pavlovian conditioning in some types of situations. For instance, we have the impression that the classic Pavlovian salivary conditioning, which is rarely used now, is rather difficult to obtain reliably.

In our first two studies, we examine by questionnaire the two most robust examples of evaluative conditioning: phobias and taste aversions, both in humans. We show that even these robust, one-trial learning phenomena often fail to occur. In the next two studies, we describe two failures to achieve significant EC in real world situations in which we engineered multiple pairings of relatively neutral odors with negative or positive events. In the final two studies, we report negative results in two laboratory studies using relatively strong visual negative and positive USs.

Because most of the results we report are negative, we have made our presentation of procedures and results briefer than the norm, while keeping the essential information that might be useful in helping others to integrate these findings into a fuller picture of EC.

STUDY 1: CONDITIONED TASTE AVERSIONS IN HUMANS

Following on the demonstration of robust, single-trial flavor aversion learning in rats, originally by Garcia and his colleagues (reviewed in Garcia, Hankins & Rusiniak, 1974; Rozin & Kalat, 1971), the very reliable conditioned taste aversion phenomenon became one of the basic paradigms in the field of animal learning. Laboratory demonstrations of the phenomenon in humans were accomplished by Bernstein (1978), and four survey studies gave clear evidence for the effect in retrospective reports from humans (Garb & Stunkard, 1974; Pelchat & Rozin, 1982; Logue, 1985; Rozin, 1986). The questionnaire studies as a group are quite convincing, and reveal the same basic properties of CTAs as are shown in the animal literature. Furthermore, the study by Pelchat & Rozin (1982) identifies nausea as the critical event in producing the effect. (USs that do not include nausea typically do not lead to an acquired dislike of the associated food.) All of the questionnaire studies took the form of asking subjects about cases in which they got sick after eating something, and queried subjects about the type of illness, time between ingestion and illness, and any changes in liking for the associated food. Subjects were prompted to produce examples in which a change in liking occurred. In the present study, we change the form of the question. We begin not with the pairing of food and a negative event, but the last occasion in which the subject became sick to his or her stomach (the critical US for CTAs).

Method

Students in an introductory psychology class (n = 232 returned forms) anonymously filled out a short questionnaire in class. They were asked to:

''remember the last occasion at which you became sick to your stomach. Did you throw up: (yes, no). Was there nausea: (a lot [3], some [2], none [1]); how sick did you get (extremely [3], moderately [2], slightly [1]); are you confident about the cause of the illness (yes, no); time in hours between eating and the illness event (hours/minutes); was any novel food eaten at that time? (yes/no); What food (write in); Did any food come to be disliked as a result of the event (yes/no); If yes, rate that food for liking before and after the event (9 point standard hedonic scale with 1 = dislike extremely and 9 = like extremely).

Results and Discussion

Altogether, 214 of 232 subjects (92%) reported at least some nausea (36% reported ''a lot''). Only 72 of the eligible 214 (34%) subjects reported a subsequent dislike of a food from the preceding meal (including declines of at least 1 scale point). The average drop on the hedonic scale for the 70 of the 72 subjects who provided pre- and posthedonic ratings was −4.5, a large effect.

In an attempt to explain this rather poor yield, we introduce further subject selections which incorporate variables that we believe (on the basis of the CTA literature) might increase the proportion showing CTAs. If we limit the sample to subjects who got ill in less than 6 hours after eating, and who rate the severity of illness as greater than slight we obtain an eligible sample of 109 subjects, of whom 46 (42%) show an acquired food dislike. This improves the yield modestly from 34% to 42%. The most critical variable, based on the animal literature on CTAs, and basic principles of Pavlovian conditioning, should be the consumption of a novel CS (food) before the illness. If we add this condition to the time interval and illness intensity constraints, there remain 17 eligible cases, of which 13 report an aversion (76%). Hence, on the basis of these calculations, a clear majority (76%) of cases in which the basic conditions for CTAs seem to hold actual result in a CTA.

The results we report are based on retrospective data. However, the same sort of data has reliably replicated the basic CTA phenomenon and its properties, as derived from animal research. There are abundant biases in this data, producing, in particular cases, increases in the likelihood of reporting type I or type II errors. We presume that it is not hard for most people to remember a relatively recent incident in which they became sick to their stomach. We presume people are much less likely to remember what they had eaten previously, unless an aversion occurred. This memory bias should not, in our opinion, affect our results. However, we may have overestimated the number of cases in which an aversion occurs because a novel food followed by illness is more likely to be remembered if it became the target of an aversion. Since our major point in this study is to show that CTAs often fail to occur when the ''conditions'' for them are met, the principal bias does not disturb our

conclusion, since it operates to reduce memory for ''failures.'' Laboratory experience with CTAs in animals with an ideal US, a novel US, and a CS-US interval of one half hour or so produce a very high conditioning rate. We cannot explain the apparently lower rate in humans, but suggest that it might be related to both how humans ''frame'' the situation and the degree of novelty of the CS.

STUDY 2: ANIMAL PHOBIAS

Animal phobias typically have a well defined target, and constitute one of the best documented examples of rapid, real world Pavlovian conditioning in humans. Dating from the classic laboratory work of Watson and Rayner (1920), there have been a series of laboratory studies which suggest a conditioning model for these phobias (Ohman, 1979). Furthermore, therapies based on extinction and counter-conditioning (assuming Pavlovian conditioning as a cause) have been highly successful. A retrospective questionnaire used with college students employing a methodology similar to that used here examined the cause of dental phobias (anxieties) (Davey, 1989). A very strong, but not perfect, relation was established between painful experiences at the dentist and development of dental anxiety; 94% of subjects who developed dental anxieties had at least one painful experience at the dentist.

We believe that, just as with CTAs, animal phobias fit the definition of evaluative conditioning. In this retrospective questionnaire study, we explore cases in which dogs have been associated (usually on a single occasion) with great fright and/or bodily damage, as a means of uncovering the reliability of dog phobia development.

Method

Brief questionnaires about dog phobias were distributed in two introductory psychology classes, with total returns from 381 subjects. Subjects were asked the following questions: ''Have you ever been bitten, attacked or seriously frightened (chased, witnessed someone else being harmed) by a dog? (yes/no); (if more than one encounter, questions refer to most serious encounter); Do you remember it or were you told about it? (response); Describe the physical damage to yourself: (no physical injuries, minor, moderate, serious [requiring immediate emergency treatment and hospitalization]); Rate the damage to others on the same scale; Rate your degree of fright (5 point scale, not at all to extremely); How many prior encounters did you have with (1) all dogs, (2) the particular dog in this incident, (3) dog breed or dogs that look like the dog that attacked in this instance? (all answered on the scale: none at all, minimal, occasional, moderate, frequent, don't know); Did you grow up with a dog in the house? (yes/no); Did you develop a dog phobia (defined as a fear of dogs out of proportion to the danger of dogs) at the time? (yes/no); To what dogs? (all dogs, dogs of certain breed, size,

TABLE 1
Study 2: Relation Between Presence of Dog Phobia and Presence of Dog Trauma

		Phobia absent	Phobia present
Dog attack or fright	No	171	27
	Yes	111	72

color, dogs you don't know, the specific dog that harmed or frightened you); Explain your phobia as best you can, even if you didn't have a negative encounter with a dog.''

Results and Discussion

Our principal report is a two-by-two table relating presence of dog attack/fright and presence of phobia (Table 1).

The table and analysis include only subjects whose dog phobia post-dated the trauma the subject reported. There is clearly a substantial and significant relation between dog attacks and dog phobias. While only 14% (27/198) of people who did not remember experiencing dog attacks had a phobia, 39% (72/183) of people who remembered such an attack reported the experience of a dog phobia. However, there are 27 cases of phobia without remembered incident. These ''failures'' could result in part from memory failures, but are likely to be, for the most part, phobias generated by something other than traumatic Pavlovian conditioning. The issue of direct relevance to the concerns of this paper are the 111 subjects who report a dog trauma without a phobia; indeed, only 39% of dog attacked subjects report a phobia.

Of the 111 subjects in the critical ''wrong'' cell, 34 rated themselves very or extremely frightened during the dog incident. Overall, comparing phobics (72) and non-phobics (111) who experienced dog attacks/frights (using independent t-tests, $p < .05$ two-tailed), the phobics rated the dog event as significantly more frightening, and they had significantly less exposure to dogs prior to the event (as predicted by informativeness principles of Pavlovian conditioning). Considering all subjects reporting a negative encounter; degree of fright ($r = .30$) and amount of prior contact with dogs ($r = .31$) showed the largest correlations with dog phobias.

As with the case of CTAs, we find a substantial relation between dog attack/fright experiences and dog phobias, but a relation that is far from perfect. A substantial number of people report an experience of serious fright and/or serious physical damage from dog encounters, while being moder-

ately unfamiliar with dogs and completely unfamiliar with the offending dog, and still did not develop a phobia.

STUDY 3: EVALUATIVE CONDITIONING IN THE REAL WORLD—AN EXPERIMENT ON GETTING TO LIKE SHAMPOO AROMAS

This study followed on our observation that hair-washing was a very positive event for some people, while quite neutral for others. This US was paired in as natural a way as possible, with the odor of an unfamiliar, neutral shampoo (CS). This study was part of an attempt by our laboratory, and in different paradigms, the Leuven laboratory, to establish real world contexts for EC. As it happens, one of us (AW) was a principal experimenter in both of the (unsuccessful) studies reported here and in the one quite successful and one marginally successful study carried out at Leuven (Baeyens, Wrzesniewski, de Houwer, & Eelen, 1996). The more successful study at Leuven involved conditioning of distinctive and relatively neutral odors by using the odors as fragrances in a public lavatory regularly frequented by the subjects. Hedonic changes in the odors was related, in the predicted manner, to attitudes to going to the lavatory. In the less successful study, there was some indication of a greater positive hedonic change in odors paired with pleasure-giving massage as opposed to relatively painful therapeutic massage.

Method

Students in an introductory psychology class, as part of a questionnaire on another topic, answered a question on their reaction to washing their hair, from neutral to extremely pleasant. They were also asked to indicate the brand/type of shampoo that they usually used. A subset of 21 hair-washing likers (rating it as extremely pleasant), and 19 hair-washing ''neutral'' subjects were selected at random from those eligible.

Each subject came into the laboratory under the cover story of participating in a study on the effect of odor on mood. Subjects first rated the pleasantness of the fragrance of seven shampoos presented in unmarked bottles. The shampoos were selected to be relatively different from one another (and did not include the subject's own shampoo). The ratings were carried out on a 200 mm analog scale, with ''neutral'' marked at the midpoint. Subjects were given an unmarked bottle of the shampoo whose odor they rated closest to neutral. If they asked about the basis for assignment, subjects were told that this shampoo was randomly assigned.

Subjects were instructed to use this new shampoo for a period of approximately two weeks. Before and after each shower, subjects rated their moods on a -9 (''extremely negative'') to $+9$ (''extremely positive'') point scale on a waterproof plasticized rating sheet that was provided.

After two weeks, the subjects returned to the laboratory, and rated the

seven shampoo odors. After answering a few written questions about their idea about the purpose of the study, the subjects were debriefed and paid.

Results and Discussion

There was no difference between the mean change in rating of the target shampoo odor between the hair-washing-liking and hair-washing-neutral subjects. Both groups showed a slight improvement in liking for the target shampoo odor (7.0 mm on the 200 mm scale for shampooing likers, 7.9 mm for shampooing neutrals), an effect somewhat larger than the positive effect on the average of the seven non-target shampoos (2.0 and 3.2 mm, respectively).

Subjects were debriefed; only one realized the true purpose of the study; all claimed to believe the cover-story.

We were very hopeful of positive results in this study, because of the totally natural, and substantial number of odor/positive-event pairings. Furthermore, in the real world, we expect that there actually is some conditioning of this very sort that goes on.

STUDY 4: EVALUATIVE CONDITIONING IN THE REAL WORLD—AN EXPERIMENT ON GETTING TO LIKE ODORS ASSOCIATED WITH POSITIVE OR NEGATIVE EVENTS IN DAILY LIFE

This experimental study is modeled on the prior study, but uses positive, neutral or negative events individually selected from the personal reports of subjects to serve as the USs, to be accompanied by odor CSs.

Method

Three groups of 15 students were chosen from an introductory psychology class on the basis of a brief screening questionnaire. Students were asked to list the most positive, negative and neutral activities that they engaged in at least twice a week. The experimenters selected from among these activities those: (a) during which the subject could sniff an odor in a bottle, and (b) were predictable, appropriate and in the subjects' control. After this activity evaluation, each subject had between one and three eligible activities. Subjects were assigned randomly to the positive, negative or neutral activity conditions, with resampling if the selected subject did not have an eligible activity of the type designated.

Subjects were brought into the laboratory and asked to rate the pleasantness of 8 relatively neutral odors on a 200 mm analog scale with a midpoint, ''neutral'' mark. The odors were: andrane, benzyl acetate, linaloal acetate, D-limonene, α-pinene, cedrenyl acetate, geraniol, and hedione. The experimenter selected the odor rated closest to the neutral point. Subjects were given a small one ounce plastic squeeze bottle containing this odor. They were asked to smell the odor in their bottle each time they were engaged in their designated activity. It was suggested that they smell the odor for one

TABLE 2
Study 4: Mean Shift in Ratings of Odor in Target and Other (Seven) Odorants by Sniffing Conditions

Group	Change in nontarget odors[a]	Change in target odors
Positive ($n = 31$)	10.3 (13.2)	1.0 (39.8)
Neutral ($n = 30$)	6.2 (16.9)	1.9 (29.6)
Negative ($n = 33$)	3.2 (16.3)	8.0 (33.8)

Note. Mean (*SD*). Measures in mm on the 200mm scale. Post–preodor ratings.

[a] Entry for each subject is the mean change in the seven nontarget odors.

to two minutes during the activity. The cover story was that the study was about the effects of odor on mood. A label was placed on the bottle on which subjects rated their mood (on a −9 to +9 scale) both before and after engaging in their activity. Included on the label, and in the subject instructions, were ''control'' trials, in which mood was to be rated before and after the activity, but the odor was not to be sniffed.

After subjects had completed 20 trials (3 of which, in designated positions, were designated as ''controls''), or a trial period of one month had passed, they returned to the laboratory to sniff and rate the odors again. In addition, subjects filled out answers to several questions about what they presumed the purpose of the study to be, and then were debriefed and paid.

In order to increase the *n*, a second set of subjects was run, with identical procedures, a few months after the first, with between 15 and 18 subjects in each group.

Results and Discussion

The critical measure was the change in rating of the target odor. This was compared to the mean shift in the rating of the remaining seven rated odors. For reasons that we do not understand, there was a suggestive effect on the liking for the ''positive-paired'' target odor in the initial study ($n = 15$), with a movement of 21.8 mm in the positive direction. This effect was counteracted by a movement in the negative direction for the additional 16 subjects run in the second phase. For the negative and neutral activities, results were highly similar across the first and second phase samples,

In the full sample, there is no significant effect of treatment on either the target odors (Table 2), or the nontarget odors (Table 2). No difference between scores for any pair of the three groups meets the generous $p < .05$ (2-tailed *t*-test) evaluation. There is weak evidence for a mere-exposure effect from scores on the non-target odors (Table 2).

The mood ratings of the subjects indicate movement in the predicted direc-

tion after the positive or negative activities. Only two subjects were aware of the real purpose of the study.

STUDY 5: DISGUST EVALUATIVE CONDITIONING IN THE LABORATORY—ODOR CSS AND AFFECTIVE PICTURE USS

This and the subsequent study represent attempts to produce EC in the laboratory, using odor as a CS. The basic design is similar to that used in the original Levey—Martin and the later Leuven picture-pairing studies.

Method

Subjects were 26 undergraduate students who participated in a single .5 to 1 h experimental session. Subjects were randomly assigned to one of two groups that differed only in the particular pairings of odors and pictures.

The presence of adequate odor acuity in subjects was tested with a brief odor identification test. Five odorants (garlic, cinnamon, coffee, bacon bits, nothing) were presented to subjects in random order. Each odorant was delivered from an opaque squeeze bottle. Subjects were given a list of the five odorants listed above, and asked to select the relevant odorant. There were six trials (one of the odorants was presented twice). We required a score of at least 5/6 correct for a subject to qualify for the experiment. All subjects qualified.

Subjects were told that the experiment was about memory for odor-object pairings, as a function of liking for the odor and the object. Subjects would be pretested to get ratings of characteristics of the odors and objects, then presented with a memory task, followed by a post-test of the same attribute ratings of the odors and objects. The post-test was justified to subjects as a way of getting more reliable data on liking; this could conceivably have reduced an effect because it may have created a demand characteristic to maintain constant ratings. In debriefings, all subjects seemed to have accepted the rationale for this study.

In the pre-test and post-test, subjects rated the odors on three 200 mm analog scales, anchored by weak-strong, masculine-feminine, and dislike-like. The odors employed were: lavender, sweet birch, jasmine, cajuput, sassafrass, sandalwood, coconut, and walnut. Similar ratings, on the same scale, were obtained for the US pictures, except that the weak-strong scale used with the odors was changed to one that was anchored by simple-complex for the pictures.

The eight US pictures, classified here by our a priori assignment of valence but presented in random order, were photographs of: *negative:* a ''dogdoo,'' wounded child, a cockroach; *neutral:* pencil sharpener; *positive:* Mae West, US flag, puppies, John F. Kennedy.

The two groups differed only in that they experienced two different random assignments of odors to objects. On the first round, each picture was presented along the bottom of a plastic box imbued with the appropriate odor.

TABLE 3
Study 5: Change in Rating of Odors Associated with Negative, Neutral or Positive Objects

Object	Mean rating of object	Mean odor rating change (*SD*)	*t*[a]
Puppies	155	−12.46 (44.66)	−1.42
U. S. flag	128	10.35 (44.83)	1.17
President JFK	114	−12.38 (35.57)	−1.77
Mae West	108	−19.69 (35.93)	−2.79*
Pencil sharpener	108	−3.73 (27.15)	−0.70
Wounded child	26	−9.27 (51.39)	−0.91
Dogdoo	26	−2.81 (27.86)	−0.51
Cockroach	14	−27.58 (38.78)	−3.63**

Note. $n = 26$; ratings in mm on a 200-mm Scale.
[a] Dependent *t*-test, two-tailed.
* $p < 0.5$.
** $p < .01$.

Subjects sniffed the opened box while looking at the object and repeating the name of the object. The subsequent seven rounds were identical to the first in mode of presentation (8 exposures, one of each pairing), except that: (a) the order of presentation of the picture-odor pairs was varied randomly; and (b) subjects were asked to close their eyes and guess the object on smelling each odor. Immediate feedback (they were allowed to open their eyes and look at the picture while still smelling the odor) followed each guess. Trials lasted 8 seconds, with a 4 second intertrial interval. Following the eight rounds, each consisting of eight odor-picture pairs, the post-test odor and object ratings were carried out (see above). This was followed by a debriefing.

Results and Discussion

In the debriefing, not a single subject indicated awareness of the true purpose of the experiment (gauging changes in liking for the odors). The data from both groups were combined such that all trials with a specific US (e.g., puppies) were merged. Table 3 presents the mean rating of each US, along with the mean change in rating of the odor paired with each of the USs. Note that our *a priori* assumptions about the actual attractiveness of the various USs were not accurate. One might reasonably characterize the puppies and US Flag as positive, and the wounded child, dogdoo and cockroach as negative. The JFK and Mae West pictures, which we presumed to be positive, emerged as quite close to neutral (though on the positive side). There were two significant changes in CS ratings. One, as expected, was a substantial 28 mm drop in the rating for the odor paired with the cockroach. The second was a not predicted drop of 19 mm in the odor paired with the close to neutral picture of Mae West.

A second individualized measure of ''net'' conditioning for each subject was calculated, using the correlation (Pearson *r*) of US ratings with change in corresponding CS ratings. For each subject, the correlation was made up of eight pairs of measures: for each CS–US pair, the two points entering into the correlation were the mean of the pre- and post-picture rating and the posttest minus pretest change in the rating of the corresponding odor. The mean of the 26 *r*s is .13 (*S.D.* = .36), in the direction predicted by EC, but not significant ($t = 1.841$).

The results of this study can, at best, be considered mildly encouraging. By far the most effective US was the photograph of a cockroach.

STUDY 6: DISGUST EVALUATIVE CONDITIONING IN THE LABORATORY—ANALYSIS OF INDIVIDUAL DIFFERENCES AND ATTEMPTS TO ENHANCE THE EFFECT

This experiment had two aims. The first was to capitalize on the suggestive results of Study 5, in which a photograph of a cockroach was the most effective US. In this study, an EC experiment similar to that in Study 5 was conducted, using a more potent cockroach stimulus: a real, dead cockroach. In addition, another strong, negative US was included: a photograph of Adolph Hitler. Second, we explore the possibility that occurrence of EC might have an individual difference component. Some subjects in all of the first 5 studies showed substantial effects; these effects could result from the variation we expect in performance on any task in the absence of relevant individual differences, but they could also be caused by stable individual differences in the tendency to undergo EC. The variation could also result from an interaction between the types of stimuli and parameters of conditioning and individual differences in the subjects. In this study, each subject experienced two consecutive EC ''experiments.'' The two experimental rounds involve the same set of USs, but a totally different set of CSs. The question we ask is whether the degree of EC in one phase predicts the degree in the other phase. In addition, an extinction procedure was added to the protocol.

Method

Subjects ($n = 20$) came from the same undergraduate population as in the prior studies. The explanation of the study, olfactory competence screening and payment of subjects, number of conditioning cycles and parameters of conditioning were the same as in Study 5.

The odors employed in this study were: lavender, sweet birch, sandalwood, kiwi, and rose in the first phase; clove/coconut mixture, papaya, coriander, jasmine, and ginger/cinammon mix in the second phase. The USs were selected, *a priori,* to represent two negative, two positive, and one neutral event. They were the same in both phases (referred to by name for the subjects, as indicated by underline): actual dead *cockroach* (1″), picture of

TABLE 4
Study 6: Change in Liking for Odors Paired with Affectively Charged Objects

US object	Session order	Conditioning-like change post–pre (mm)[a]	Extinction-like change post–pre (mm)[a]
Roach	1	−11 (33)	12 (40)
Roach	2	−15 (32)	−16 (49)
Hitler	1	9 (30)	24 (41)
Hitler	2	−3 (25)	−11 (33)
Rainbow	1	10 (45)	10 (37)
Rainbow	2	−9 (30)	4 (43)
Connery–Pfeiffer	1	5 (31)	8 (45)
Connery–Pfeiffer	2	6 (15)	15 (35)
Nothing	1	11 (21)	30 (34)
Nothing	2	1 (21)	1 (39)

Note. Mean (*SD*) rating change for 20 subjects in mm on a 200-mm like–dislike analog scale.

[a] Post-(conditioning or extinction) rating on a 200-mm like–dislike scale minus pre-(original) rating on same scale, measured in mm.

Hitler, picture of attractive nature scene (*rainbow*), picture of Sean Connery with Michelle Pfeiffer (*Connery-Pfeiffer*), or *nothing.* The same pretest, posttest and presentation conditions were in force, as in Study 5. As in Study 5, there were 8 rounds of exposure, each consisting of a pairing of an odor with the US randomly assigned to it. In the pretest and posttest, only a single rating (like–dislike) was made for each CS and US. Following the posttest, an extinction procedure was instituted. For eight rounds, subjects sniffed each of the five odorants without the USs, rating the CS on a different scale in each round (e.g., fruity-non-fruity). This was followed by another posttest of both the odorants and the objects.

This entire sequence was repeated with a new set of five odorants, and the same USs. In both phases of this study, CS-US pairings were determined randomly for each subject.

Results

The roach paired CS was the only CS that showed a net drop in both the first and second rounds (only the second round effect is significant, $t = 2.09$, $df = 19$, $p < .05$). Examination of Table 4 indicates no other consistent trend in the direction predicted by EC. The biggest, yet very modest positive change (mean 11 mm in the first round) is for the *neutral* (nothing) US. ''Extinction'' effects are highly variable, and sometimes in the wrong direction, but since conditioning was not clearly established, extinction measures have limited relevance.

The second issue of concern in this study was to determine if there is a consistent subset of subjects who show EC. The correlation (Pearson r, $n =$

20) for size (including direction) of the conditioning effect on roach rounds 1 and 2 was .33 (n.s. for $n = 20$, but in the predicted direction). Another approach is to examine any large effect ($> = 30$ mm) in the predicted direction for the four valenced USs. For the 40 roach conditioning situations (20 subjects, 2 conditioning sessions per subject), there were 11 cases of decreases in liking for the paired odor of greater than 30mm. These occurred in 10 different subjects; only one subject showed a substantial liking decrease in both conditioning phases. Results are similar, with overall smaller effects for the other USs: for Hitler, there were 5 large decreases, over 4 subjects, while for rainbow there were 5 large increases over 5 subjects and for Connery-Pfeiffer there were 4 large increases over 4 subjects. Thus, considering the 20 subjects across the 4 USs and 2 trial sets for each subject (160 data points), there were 24 large (>30mm) effects; only 2 cases occurred in which the same subject showed a large effect in the predicted direction on both phases. This is not impressive evidence.

GENERAL DISCUSSION

Overall, the results we present argue for the operation of some yet-to-be-understood factors that influence whether EC will take place or not. The results suggest that novelty of the CS, as expected, may be critical, but this was really not an issue in the last four (experimental) studies. Temporal contiguity, individual differences in susceptibility, and a variety of vague contextual effects may also be involved. In general, EC study results do not constitute a homogenous set of weakly supportive studies. Rather, there is a group of studies which vary from no net effect to substantial effects. All together, the results argue for an effect. But thinking about the important role of EC in daily life, the results under controlled conditioning are disappointing. Although there are many cases of failure in real world instances, the robustness of the real world EC effects discussed is impressive. Hence, it may be the case that the laboratory environment, in general, operates to reduce the effect or make it more sensitive to parameters that normally vary within and between experiments.

It is notable that our four disappointing experimental studies all employed odor as a CS (although odor is an important component of CTAs, one of the robust real-world examples). On the one hand, it is possible that the generally high affective loading of odors prior to the study makes odors less susceptible than other stimuli to affective manipulation. We were able to select relatively neutral odor CSs, but this was not easy. It is also possible that the difficult-to-find neutral odors have special properties, to wit, a resistance to becoming attached to affectively significant events. On the other hand, odors seem to be highly labile in affective interpretation; for example, the same odor can be highly pleasant or unpleasant depending on its perceived source (e.g., cheese versus feces). It is our guess that odors are affectively laden, but also affectively labile. However, in the pilot studies on

olfactory CS EC by Todrank (unpublished), a series of unsuccessful attempts were made. When Todrank switched to using odor as a US (Todrank *et al.,* 1995), a substantial and reasonably stable effect emerged.

In the last two laboratory studies, the most effective US was a cockroach, either in pictorial representation or in real life. Taking note of this, we performed a few pilot studies in which we made the roach US more potent (even though much of the Leuven work suggests that the strength of US is not a critical variable). In one study, an unfamiliar Japanese cracker, the CS, was initially rated for palatability (based on appearance) on the 200mm scale. Then, the cracker was dropped into a large bottle filled with live, large cockroaches, who scurried all over and around the cracker. Subjects were then given another clean instance of the same cracker, from a commercial cracker box, and asked again to rate their liking. No attempt was made in this study to hide the obvious centrality of the contingency. No overall effect was obtained, although about 10% of the subjects showed a substantial drop in rating.

Perhaps a 10% or so ''yield,'' especially after one or a few conditioning trials, should be considered encouraging. In our prior questionnaire study of disgust conditioning (Rozin, 1986), we reported only 22 cases (from well over 400 surveyed subjects) of aversion based on pairing of some object or event with a disgusting US in real life. Surely, such pairings occur rather often in real life. Our 1986 study indicates that such experiences sometimes lead to enduring and substantial aversions, but it would seem that most of the time either nothing happens, or there is just a transitory effect.

Our conclusion from these studies is that EC is important and worth serious study, but that part of the study should include isolation of the conditions that promote EC. It may well be that something like what is loosely called ''framing'' may be a critical variable. Within procedure, we have weak evidence for an individual difference variable: this could be explored more systematically, including perhaps a linkage between susceptibility to EC in the real world and in the laboratory. Such research would have the dual advantage of contributing to our understanding of the properties of EC, and allowing us to devise more robust and reliable laboratory models to study it.

REFERENCES

Baeyens, F., Crombez, G., van den Bergh, O., & Eelen, P. (1988). Once in contact always in contact: Evaluative conditioning is resistant to extinction. *Advances in Behaviour Research and Therapy,* **10**(4), 179–199.

Baeyens, F., Eelen, P., van den Bergh, O., & Crombez, G. (1990). Flavor-flavor and color-flavor conditioning in humans. *Learning & Motivation,* **21,** 434–455.

Baeyens, F., Eelen, P., van den Bergh, O., & Crombez, G. (1991). Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour Research & Therapy,* **30,** 133–142.

Baeyens, F. (1996). Intrinsic or implicitly referential? On the representation of acquired evaluative value. (unpublished manuscript).

Baeyens, F., Kaes, B., Eelen, P., & Silverans, P. (1996). Observational evaluative conditioning of an embedded stimulus element. *European Journal of Social Psychology,* **26,** 15–28.

Baeyens, F., Vansteenwegen, D., DeHouwer, J., & Crombez, G. (1996). Observational conditioning of food valence in humans. *Appetite,* **27,** 235–250.

Baeyens, F., Wrzesniewski, A., De Houwer, J., & Eelen, P. (1996). Toilet rooms, body massages, and smells: Two field studies on human evaluative odor conditioning. *Current Psychology,* **15,** 77–96.

Bernstein, I. L. (1978). Learned taste aversions in children receiving chemotherapy. *Science,* **200,** 1302–1303.

Garb, J., & Stunkard, A. J. (1974). Taste aversions in man. *American Journal of Psychiatry,* **131,** 1204–1207.

Garcia, J. (1989). Food for Tolman: Cognition and cathexis in concert. In T. Archer & L.-G. Nilsson (Eds.). *Aversion, avoidance, and anxiety* (pp. 45–85). Hillsdale, N.J.: Lawrence Erlbaum.

Garcia, J., Hankins, W. G., & Rusiniak, K. W. (1974). Behavioral regulation of the milieu interne in man and rat. *Science,* **185,** 824–831.

Garcia, J., & Koelling, R. A. (1966). The relation of cue to consequence in avoidance learning. *Psychonomic Science,* **5,** 123–124.

Levey, A. B., & Martin, I. (1975). Classical conditioning of human ''evaluation'' responses. *Behavior Research & Therapy,* **13,** 221–226.

Levey, A. B., & Martin, I. (1983). Cognitions, evaluations and conditioning. Rules of sequence and rules of consequence. *Advances in Behavior Research & Therapy,* **4,** 181–195.

Levey, A. B., & Martin, I. (1990). Evaluative conditioning: Overview and further options. *Cognition and Emotion,* **4,** 31–37.

Logue, A. W. (1985). Conditioned food aversion learning in humans. *Annals of the New York Academy of Sciences,* **443,** 316–329. [Experimental assessments and clinical applications of conditioned food aversions. N. S. Braveman & P. Bronstein (Eds.)]

Martin, I., & Levey, A. B. (1978). Evaluative conditioning. *Advances in Behavior Research & Therapy,* **1,** 57–102.

Mineka, S., & Cook, M. (1988). Social learning and the acquisition of snake fear in monkeys. In T. R. Zentall, & B. G. Galef, Jr. (Eds.) *Comparative social learning* (pp. 51–74). Hillsdale, N.J.: Lawrence Erlbaum.

Ohman, A. (1979). Fear relevance, autonomic conditioning and phobias: A laboratory model. In S. Bates, W. K. Dockens, K. G. Blotesharm, L. Melin, & P. O. Sjoden (Eds.), *Trends in behavior therapy.* New York: Academic Press.

Parker, L. A. (1982). Nonconsummatory and consummatory behavioral CRs elicited by lithium- and amphetamine-paired flavors. *Learning & Motivation,* **13,** 281–303.

Pelchat, M. L., & Rozin, P. (1982). The special role of nausea in the acquisition of food dislikes by humans. *Appetite,* **3,** 341–351.

Pelchat, M. L., Grill, H. J., Rozin, P., & Jacobs, J. (1983). Quality of acquired responses to tastes by Rattus norvegicus depends on type of associated discomfort. *Journal of Comparative Psychology,* **97,** 140–153.

Rachman, S., & Hodgson, R. J. (1968). Experimentally-induced ''sexual fetishism'': Replication and development. *Psychological Record,* **18,** 25–27.

Rozin, P. (1986). One-trial acquired likes and dislikes in humans: Disgust as a US, food

predominance and negative learning predominance. *Learning and Motviation,* **17,** 180–189.

Rozin, P., & Kalat, J. W. (1971). Specific hungers and poison avoidance as adaptive specializations of learning. *Psychological Review,* **78,** 459–486.

Rozin, P., & Zellner, D. A. (1985). The role of Pavlovian conditioning in the acquisition of food likes and dislikes. *Annals of the New York Academy of Sciences,* **443,** 189–202.

Seligman, M. E. P. (1970). On the generality of the laws of learning. *Psychological Review,* **77,** 406–418.

Stuart, E. W., Shimp, T. A., & Engle, R. W. (1987). Classical conditioning of consumer attitudes: Four experiments in an advertising context. *Journal of Consumer Research,* **14,** 334–349.

Todrank, J., Byrnes, D., Wrzesniewski, A., & Rozin, P. (1995). Odors can change preferences for people in photographs: A cross-modal evaluative conditioning study with olfactory USs and Visual CSs. *Learning & Motivation,* **26,** 116–140.

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology,* **3,** 1–14.

Zellner, D. A., Rozin, P., Aron, M., & Kulish, C. (1983). Conditioned enhancement of liking for flavors by pairing with sweetness. *Learning & Motivation,* **14,** 338–350.

Received August 26, 1997
Revised June 2, 1998