Adv. Behav. Res. Ther. Vol. 10, pp. 179–199, 1988.
Printed in Great Britain.

0146-6402/88 $0.00+.50

ONCE IN CONTACT ALWAYS IN CONTACT: EVALUATIVE CONDITIONING IS RESISTANT TO EXTINCTION

Frank Baeyens,* Geert Crombez, Omer Van den Bergh and Paul Eelen
University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium

Abstract—The present study aimed at obtaining some further support for the hypothesis of a distinction between two basically different kinds of learning in a Pavlovian conditioning preparation: signal-learning and affective–evaluative learning (Baeyens *et al.*, 1988a,b; Levey and Martin, 1987). In this respect, we conducted an experiment to verify the Levey and Martin (1983, 1987) hypothesis that, unlike signal-learning, evaluative conditioning should be resistant to extinction. Mere contingent presentation of neutral with (dis)liked stimuli was sufficient to change the affective–evaluative tone of the originally neutral stimuli in a (negative) positive direction ($p<0.0001$). A subsequent extinction procedure did *not* have any influence on the acquired evaluative value of the originally neutral stimuli ($p<0.0001$). A follow-up study demonstrated that the evaluative discriminations were still present two months after the acquisition and extinction manipulations ($p<0.0001$). These findings provide full support for the resistance to extinction hypothesis. At a theoretical level, this is considered to be further evidence for the hypothesis that evaluative conditioning is not mediated by the acquisition of propositional–declarative knowledge about stimulus contingencies. Finally, we suggest an intriguing analogy between the evaluative conditioning phenomenon and the 'laws of sympathetic magic' (Rozin *et al.*, 1986).

INTRODUCTION

In previous studies (Baeyens *et al.*, 1988a,b) we proposed a theoretical distinction between two basically different kinds of learning in a classical conditioning situation: signal-learning and affective–evaluative learning. On the one hand, the subject can learn about the contingency relationship between the conditioned stimulus (CS) and the unconditioned stimulus (US): "CS predicts/is contingent with US". This kind of learning is referred to as *signal-learning*, which can be conceptualized as the acquisition of propositional–declarative knowledge about stimulus relationships in the environment, mediated by controlled (or capacity limited) cognitive processes (Baeyens *et al.*, 1988a,b; Dawson and Shell, 1985, 1987; Mackintosh, 1983). On the other hand, the subject's evaluation of the CS can be altered intrinsically in a classical conditioning situation: (s)he can learn to 'like or dislike' the CS. This kind of learning is referred to as

*Research Assistant of the National Fund for Scientific Research, Belgium. Correspondence may be addressed to Frank Baeyens, Department of Psychology, University of Leuven, B-3000, Belgium.

affective–evaluative learning or *evaluative conditioning* (Martin and Levey, 1978, 1985, 1987; Levey and Martin, 1975, 1983, 1987). Evaluative conditioning is used in shorthand to describe the process by which an affective evaluative reaction (ER) evoked by a significant stimulus is transferred to a previously neutral stimulus, presented contingently (contiguously) with the significant stimulus.

Over the last two decades, the theoretical work on human as well as on animal classical conditioning almost exclusively focused on signal-learning (Eelen *et al.*, 1988). This holds even to the degree that many prominents on the classical conditioning scene tend to *equate* classical conditioning with signal-learning (e.g. Rescorla and Wagner, 1972; Wagner, 1978; Rescorla, 1980; Mackintosh, 1987; Öhman, 1979, 1983, 1986; Davey, 1987; Dawson and Shell, 1985, 1987). As discussed in more detail in Baeyens *et al.* (1988b), this equation unnecessarily restricts the range of phenomena amenable to an analysis from the Pavlovian conditioning perspective (see also Garcia, 1977; Seligman and Hager, 1972; Rozin *et al.*, 1986; Rozin and Fallon, 1987). The proposed distinction can be partially justified on logical grounds: 'knowing that' the CS predicts the US logically is no sufficient condition for and is qualitatively different from '(dis)liking' the US. More importantly, there is experimental evidence in favor of the hypothesis of a distinction between signal-learning and affective–evaluative learning at the level of psychological processes. For example, we could demonstrate that human evaluative conditioning, unlike signal-learning, does not require and is not influenced by contingency awareness (Baeyens *et al.*, 1988a). This finding was interpreted as indicating that (a) evaluative conditioning does *not* imply the acquisition of a contingency proposition and (b) that the evaluative conditioning process might be an instance of automatic processing rather than of controlled, conscious processing (Schneider and Shiffrin, 1977; Shiffrin *et al.*, 1981; Shiffrin and Schneider, 1977; La Berge, 1981; Dawson and Shell, 1985; Öhman, 1983). The present study basically aimed at obtaining some further evidence for the hypothesis of a distinction between the two kinds of learning at the underlying process level. More specifically, we wanted to determine the validity of our interpretation that evaluative learning is not based on the acquisition of the contingency proposition "CS predicts US". Even though our experiments on contingency awareness and evaluative conditioning (Baeyens *et al.*, 1988a) demonstrated that evaluative conditioning does not presuppose explicit and verbalizable propositional knowledge on stimulus relationships, one could still argue that the evaluative shifts are ultimately based on *implicit* and automatically acquired propositional knowledge (CS predicts US; I (dis)like US; therefore I (dis)like CS).

In this respect, we were particularly interested in the Levey and Martin (1983, 1987) proposal that, unlike signal-learning, evaluative conditioning should be resistant to extinction. They hypothesize that ". . . the evaluative response, once it is conditioned to a neutral stimulus, cannot thereafter be extinguished through nonreinforcement . . ." (Levey and Martin, 1987, p.122). Although this statement was essentially based on "armchair speculation and evidence from the anecdotical level (*sic*)" (Levey and Martin, 1987, ibid.), we were quite willing to consider it as an hypothesis worth explicit experimental verification.

Extinction refers to an experimental procedure containing two sequential phases. In the first phase (*acquisition*), the subject is exposed to a contingency between two stimuli — a conditioned stimulus (CS) and an unconditioned stimulus (US). The second phase (*extinction*) consists of the mere presentation of the CS, without contingent presentation of the US. We considered it a plausible hypothesis that this procedure entails radically different processes for signal-learning as compared with affective–evaluative learning. Our argument is as follows. In *signal-learning* the subject essentially acquires the proposition "CS predicts US" during the acquisition phase. Hence the CS acquires its significance through reference to the US-representation (Mackintosh, 1983; Davey, 1987; Dawson and Shell, 1985, 1987). At the behavior level, the acquired significance of the CS is reflected in conditioned orienting and preparatory responses (CRs) (see Öhman, 1983). In the extinction phase of the experiment, the CS-generated expectation of the US is disconfirmed: the propositional knowledge "CS predicts US" will be updated and changed into "CS does not (no longer) predict US" (Dawson and Shell, 1985). At the behavior level, conditioned orienting and preparatory responses will diminish and finally disappear. The same does not necessarily apply to *affective–evaluative* learning. If the hypothesis is correct that the acquired (dis)liking of the CS is not mediated by (implicit) reference to the US [is not based on the (implicit) propositional knowledge "CS signals US"], then the fact that the US is no longer contingent with the CS might well leave the acquired affective–evaluative value of the CS unaltered. If on the other hand the evaluative shift of the CS relies on the (implicit) referential value to the US, then the extinction procedure should change the acquired evaluative value of the CS into a neutral direction. In this way, an experiment on extinction could offer some more convincing arguments to disentangle the problem whether or not evaluative conditioning is based on (implicit) propositional knowledge about stimulus contingencies. "Armchair speculation and reference to anecdotes" not being the optimal way to test hypotheses, we conducted an experiment to clarify the evaluative conditioning and extinction question. Our experimental hypotheses were that (a) the mere contingent presentation of neutral (N) stimuli with disliked stimuli (D) or liked stimuli (L) would result in the originally N stimuli acquiring negative or positive evaluative value respectively; and (b) that the acquired evaluative shift would be uninfluenced by subsequent nonreinforced presentation of the N stimuli.

In addition, we wanted to obtain some preliminary answers to the question of which conditions promote the two kinds of learning. As we stated elsewhere (Baeyens *et al.*, 1988b), the nature of the US might be a plausible candidate as a determinant of the occurrence of signal-learning. We hypothesized that the typical use of intrusive, highly aversive stimuli as USs in the prototypical Pavlovian conditioning experiment 'forces' the subject to engage in a search for predictors. Hence, optimal conditions for observing signal-learning are created. In contrast, the use of liked or disliked pictures of human faces, landscapes, paintings and the like in the prototypical evaluative conditioning experiment (Levey and Martin, 1975; Baeyens *et al.*, 1988a) should not call for a capacity demanding, controlled search for predictors. If the nature of the US indeed is crucial, the use of intrusive, strongly aversive USs other than shock or aversive noise, should do as well to increase the likelihood of signal-learning. In order to test this hypothesis, we decided to compare the effect of pictures of human mutilated faces versus pictures of 'normal', disliked faces as USs. We hypothesized that more subjects would engage in signal-learning in the conditions with mutilated faces than in the conditions with 'normal', disliked faces as the US. This should become apparent in that the subjects confronted with mutilated faces should have a better score on a contingency-awareness measurement (recognition questionnaire; for a description see below) conducted immediately after the acquisition phase.

We also wondered whether signal-learning and affective–evaluative learning relate to each other in a *mutually exclusive* ("or . . . or") or in a *coordinate* ("and . . . and") manner. The first possibility would imply that when there is signal-learning, there is no affective–evaluative learning (and vice versa); the latter that evaluative conditioning takes place whether or not there is signal-learning. A comparison between the evaluative conditioning and the signal-learning results in both the 'mutilated faces' conditions and the 'normal faces' conditions, could provide some preliminary insights into this problem.

Finally, we attempted to corroborate our finding that evaluative conditioning does not require and is not influenced by contingency awareness (Baeyens *et al.*, 1988a).

THE MAIN EXPERIMENT

Method

Subjects

32 undergraduate students (psychology and educational sciences; 29

females, 3 males) participated as a partial fulfillment of course requirements. They were all uninformed as to the purpose of the experiment. Each subject was tested individually. Test duration was about 80 min.

Stimuli

In all conditions, 70 colour pictures and 70 identical colour slides of human faces were used as stimulus materials. The set of pictures and slides were varied in content (males, females, age variation) and form (profile vs front, size of faces, colour palette). In the 'mutilated faces' conditions, an additional set of 20 colour pictures and 20 identical colour slides of human mutilated faces was used. This set included a variety of human faces affected by severe dermatologic and veneric diseases.

Apparatus

The experimenter's room contained a slide projector connected with a timer controlling duration of stimulus presentation, interstimulus interval (ISI) and intertrial interval (ITI). Slides were projected (40 cm × 25 cm) from the rear on a glass transparent projection screen attached in an opening of the wall between the experimenter's room and the subject's room. The subject's room contained a table and two chairs. During stimulus presentation subjects turned their chair so that they faced the projection screen at a distance of 150 cm. In order to make the cover story (see below) as plausible as possible, an apparatus with two electrodes attached to it was also present. During the experiment the electrodes were attached to the subject's second and third fingers of the left hand. Apparently visible for the subject, the apparatus was connected by cable with the experimenter's room.

Procedure

The experiment consisted of three sequential phases: a *baseline phase*, an *acquisition phase* and an *extinction phase.*

Baseline phase. The subject was told that (s)he was participating in an experiment on depression. The experiment "aimed at increasing the validity of the diagnosis of depression by methods used in neuropsychophysiology". In a short didactic discourse the experimenter stated that "human skin conductance responses (SCR) are a function of both stimulus characteristics — neutral, liked or disliked — and the emotional condition of the subject. We predicted a different SCR to N, L or D stimuli; afterwards we wanted to compare their reactions with the reactions obtained in depressive patients". All this was told in order to minimalize demand effects. Subjects expressed their baseline evaluations *(evaluative response 1 = ER1)* by laying down the pictures one by one on a 21-category scale. The categories were denoted by −100, −90, . . ., 0, . . ., +90, +100 (−100 = very disliked, 0 = neutral, +100 = very liked). The experimenter strongly stressed the importance of relying on the first, immediate and spontaneous reactions. In the conditions with 'normal faces', the subjects evaluated the 70 pictures of normal human faces; in the conditions with 'mutilated faces', subjects evaluated the 20 pictures of mutilated faces randomly intermixed with the 70 pictures of normal human faces. The three pictures which received the highest and the lowest score were used as L and D stimuli respectively. For the 'mutilated faces' conditions, we expected the pictures receiving the lowest scores to be pictures of mutilated faces. The pictures from the categories –20, –10, 0, +10, +20 were taken separately to be used as N stimuli. The experimenter attached the electrodes and returned to the experimenter's room "to obtain a stable SCR baseline". Meanwhile we arranged 9 different stimulus pairings: 3 (N–L), 3 (N–D) and 3 (N–N) control pairs. As Martin and Levey (1978) demonstrated "similarity in form, content and colour palette" to facilitate evaluative conditioning, we tried as much as possible to use this criterion for matching the pictures.* As L (D) stimuli we used the three pictures most (dis)liked by the subject. For the N stimuli, the following rule was used: if the subject assigned 12 or more pictures to the 'truly neutral' (='0') category, only these pictures were used as N stimuli. If the subject assigned less than 12 pictures to the '0'-category, we used pictures from the categories '+10/+20' as N stimuli to be combined with L stimuli, and pictures from the '–10/–20'-categories for the (N–D) pairs. Colour slides corresponding to the pictures were put into the projector.

Acquisition phase. We returned to the subject's room and said we would start the SCR measurement. All the subject had to do was watch the projected slides; SCRs would occur automatically and be registered in the experimenter's room. Stimulus pairs were presented in a completely randomized sequence [e.g. (N–L), (N–L), (N–D), (N–N), (N–L), (N–N) , . . .], and a different random sequence was used for each subject. Each stimulus pair was presented 10 times. The duration of each stimulus presentation was 1 sec, the interstimulus interval (ISI) (time between onset of first stimulus of a pair and onset of second stimulus of a pair) was fixed at 4 sec; the intertrial interval (ITI) (time between end of second stimulus of the previous trial and onset of the first stimulus of the following trial) was 8 sec. After stimulus presentation, the experimenter returned to the subject's room. He said that the essential data were gathered, but that he wanted some more information from the subject "in order to be able to control for some potentially relevant variables in the data analysis".

*The assessment of 'similarity in form, content and colour palette' was based on the experimenter's intuitive impression. As a matter of fact, we do not have any evidence to prove we were really successful in applying the 'similarity' criterion.

At that moment, the subject's evaluative reactions *(evaluative response 2 = ER2)* to the crucial stimulus materials were measured. The subject received a response booklet containing a rating scale for each picture. The scale was a line 200 mm long, labeled *very disliked* at the left, *neutral* in the middle and *very liked* at the right. In addition, there were numbers ranging from (–100/–80/–60/ . . . /0/ . . . /+60/+80/+100) under the scale. The subject was asked to mark the line at the place indicating her/his *actual* evaluation of the picture. Next the experimenter conducted the contingency awareness measurement. All projected stimuli were laid down to the left of the subject. One by one the experimenter took out the relevant N pictures [3(*N*–L); 3(*N*–D); 3(*N*–N)] and asked:

(a) "This picture was followed by
— always the same picture, namely (number . . .)
— not always the same picture
— I really could not say
during stimulus presentation".
(b) "I am
— completely sure
— rather sure
— rather unsure
— completely unsure
of my answer on the question above".

Subjects indicating correctly that a N stimulus was always followed by the same stimulus, but not being able to recognize which one, were prompted to answer whether they thought it to be a neutral, liked or disliked stimulus.

Extinction phase. The nine N stimuli [3(*N*–L); 3(*N*–D); 3(*N*–N)] were presented in a completely randomized order, *without* contingent presentation of L/D/N stimuli. Stimulus presentation duration was 1 sec, the ITI (end of stimulus of the previous trial to onset of stimulus of the following trial) was 8 sec. Subjects received 5 or 10 unreinforced presentations of each N stimulus, according to the experimental condition they participated in. As a cover story, it was explained to the subject that it was necessary to conduct a second SCR measurement in order to be able to interpret the results from the first and crucial measurement phase unambiguously. It was said that "factors other than her(his) evaluation of a stimulus might have codetermined the SCRs to the stimuli: sequence effects, due to fatigue and attentional decrements. The second measurement explicitly aimed at determining the importance of these factors and should allow us to correct the data from the previous measurement for these influences". Again, all this was said to distract the subject's attention from the real purpose of the experiment so that demand effects should be minimalized. After the stimulus presentations, the subject's evaluations *(evaluative response 3 = ER3)* of the 9 N stimuli were measured in the same way as after the acquisition phase (–100 to +100 score on response sheet). We once more stressed the importance of relying on her/his *actual* evaluative reactions.

Design: A 2 (mutilated faces/normal faces as D stimuli) × 2 (5/10 extinction trials) × 3 [type of stimulus pair: (N–L)/(N–N)/(N–D)] × 3 (moment of evaluative response measurement: baseline/acquisition/extinction) factorial design was used. The first and the second factors represent between-subject manipulations, the third and the fourth factors within-subject manipulations. Each subject was assigned randomly to one of the four conditions; 'mutilated faces, 5 extinction trials', 'mutilated faces, 10 extinction trials', 'normal faces, 5 extinction trials' or 'normal faces, 10 extinction trials'. All statistical analyses were conducted on mean ERs to N stimuli per subject [e.g. mean ER1 to N stimuli from three (N–L)-pairs; hence the data were reduced to 9 values per subject].

Results

Evaluative responses: baseline, acquisition and extinction phase

Table 1 summarizes mean ERs for different conditions at different moments of the experiment. A 2 × 2 × 3 × 3 analysis of variance demonstrated a main effect of the 'type of stimulus pair: (N–L)/(N–N)/(N–D)' factor [$F(2,56) = 25.28$; $p<0.0001$] and a main effect of the 'moment of evaluative response measurement' factor [$F\ (2,56) = 9.58$; $p<0.0001$]. A more detailed analysis with the Dunn test procedure for *a priori* contrasts made clear that the main effect of the 'type of stimulus pair' factor implied an overall difference between (N–L) and (N–D) conditions (=22.67), between (N–L) and (N–N) conditions (=12.01), and between (N–D) and (N–N) conditions (=10.66) ($d_i = 9.79$; $p<0.01$). The main effect of the 'moment of evaluative response measurement' factor implied an increase in overall ERs from baseline to acquisition phase (=5.19), from baseline to extinction phase (=7.44), but no difference between acquisition and extinction phases (=2.25) (Dunn test procedure for *a priori* contrasts; $d_i = 4.32$; $p<0.05$). No other main effects were obtained.

We also observed an interaction between 'type of stimulus pair' × 'moment of evaluative response measurement' factors [$F(4,112) = 11.72$; $p<0.0001$]. Subsequent analyses of simple main effects revealed that this interaction was due to the fact that there was no effect of the factor 'type of stimulus pair' at the baseline measurement phase (=ER1) [$F\ (2,168) = 2.36$; ns], but a significant effect after acquisition (=ER2) [$F\ (2,168) = 30.82$; $p<0.0001$], and a significant effect after the extinction phase (=ER3) [$F\ (2,168) = 28.15$; $p<0.0001$]. After acquisition, we obtained a

Table 1. ERs at Baseline, after Acquisition and Extinction

		N-L			N-D			N-N		
		ER1	ER2	ER3	ER1	ER2	ER3	ER1	ER2	ER3
Faces	5 Ext. trials	5.41	25.25	25.40	-3.75	-16.34	-12.63	-1.66	-1.26	-7.75
	10 Ext. trials	4.58	15.98	19.43	-6.66	-10.34	-7.39	-2.08	-0.16	6.49
Mutilated faces	5 Ext. trials	5.41	30.75	33.39	-4.59	-1.40	-0.70	2.50	16.80	19.25
	10 Ext. trials	0.84	9.46	10.83	-2.50	-12.43	-6.61	0	3.46	7.00
Faces conditions		5.00	20.62	22.42	-5.21	-13.34	-10.01	-1.87	-0.71	-0.63
Mutilated faces conditions		3.13	20.11	22.11	-3.55	-6.92	-3.66	1.25	10.13	13.13
Total group		4.06	20.36	22.26	-4.38	-10.13	-6.83	-0.31	4.71	6.25

significant difference between (N–L) and (N–D) pairs (=30.49), between (N–L) and (N–N) (=control) pairs (=15.65), and between (N–D) and (N–N) (=control) pairs (=14.84) (Dunn test procedure for *a priori* contrasts; d_i = 11.56; $p<0.01$). After the extinction phase we observed an identical pattern of results: a significant difference between (N–L) and (N–D) pairs (= 29.09), between (N–L) and (N–N) (=control) pairs (=16.01) and also between (N–D) and (N–N) (=control) pairs (=13.08) (Dunn test procedure for *a priori* contrasts; d_i = 11.56). A comparison between ERs after acquisition versus those after the extinction phase revealed no significant difference for N stimuli from (N–L) pairs (difference = 1.9); the same comparison revealed no difference for (N–D) pairs (difference = 2.45) nor for (N–N) pairs (difference = 1.54) (Dunn test procedures for all comparisons; d_i = 6.58 for $p<0.05$) (see Fig. 1).

Besides the interaction between 'moment of evaluative response measurement' × 'type of stimulus pair' discussed above, we observed a significant interaction between '5/10 extinction trials' × 'mutilated faces/normal faces as D stimuli' factors [$F(1,28) = 4.39$; $p<0.05$], and between '5/10 extinction trials' × 'mutilated faces/normal faces as D stimuli' × 'moment of evaluative response measurement' factors [$F(2,56) = 3.64$; $p<0.05$). Both interactions were apparently due to the absolute high, and theoretically difficult to explain, ER scores after acquisition and after extinction, in the 'mutilated faces, 5 extinction trials' condition. No other interaction reached significance.

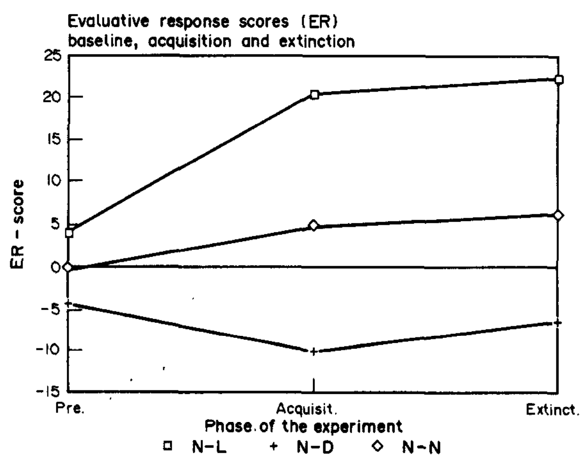

FIG. 1. Mean Evaluative Responses (ER) to originally neutral stimuli at baseline (PRE), after the acquisition phase (ACQUISIT.) and after the extinction phase (EXTINCT.) [N–L = neutral followed by liked; N–D = neutral followed by disliked; N–N = neutral followed by neutral (= control).] Mean over all conditions.

Contingency awareness

The answers obtained with the contingency awareness questionnaire were scored according to the following criteria. If a subject indicated correctly that a particular N stimulus was always followed by the particular L, D or N stimulus presented contingently with the N stimulus *or* by another L, D or N stimulus *and* if (s)he expressed that (s)he was (rather) sure of this answer, her/his answer received a score of 2.* An answer that met the same conditions, but about which a subject felt (rather) unsure, received a score of 1. All other answers received a score of 0. Hence, only correct recognition of a factually presented neutral–dislike/like/neutral contingency was required, *not* recognition of the particular L, D or N stimulus involved. Next the awareness score were summated per subject per type of stimulus pair, yielding for each subject general indices of contingency awareness for (N–L), (N–D) and (N–N) stimulus pairs.

As we were especially interested in the effect of the factor 'mutilated face/normal human face as D picture', awareness data were subjected to a 2 × 3 analysis of variance ['mutilated face/normal human face as D picture' × 'type of stimulus pair: (N–L)/(N–N)/(N–D)']; the first variable is a between-subject factor, the latter a within-subject factor). No main effects were obtained ['mutilated face/normal face' factor: $F(1,30) = 1.21$; ns and 'type of stimulus pair': $F(2,60) = 0.91$; ns]. The interaction between the two factors was significant at the 0.10 level [$F(2,60) = 2.58$; $p<0.10$]. Analysis of simple main effects made clear this interaction was due to the fact that the 'mutilated faces/normal faces' factor did have an influence on contingency awareness in the (N–L) case [$F(1,90) = 5.19$; $p<0.05$], but not in the (N–D) nor in the (N–N) cases [$F(1,90) = 0.30$; ns and $F(1,90) = 0.30$; ns, respectively]. The 'type of stimulus pair' factor did have an influence on contingency awareness in the 'normal faces' conditions [$F(2,60) = 3.22$; $p<0.05$], but not in the 'mutilated faces' conditions [$F(2,60) = 0.28$; ns] (see Fig. 2).

Finally, we calculated an 'evaluative response differentiation score' for each subject, according to the formula $D = [ER2_{(N-L)} - ER1_{(N-L)}] - [ER2_{(N-D)} - ER1_{(N-D)}]$. This 'differentiation score' provides an index of the degree of absolute divergence of the evaluations of the N stimuli due to the contingent presentation of L or D stimuli; hence it can be considered a general index of evaluative conditioning in a particular subject. On the

*In our previous experiments on contingency awareness and evaluative conditioning (Baeyens *et al.*, 1988a), only the subjects receiving a score of 2 [correct and (rather)sure] were considered 'contingency-aware' of a particular contingency. According to this criterion, subjects were aware of 36% of all (288) contingencies presented in the present experiment, which is slightly higher than the 25% (Experiment 3) and the 18% (Experiment 4) observed (under similar conditions) in Baeyens *et al.* (1988a).

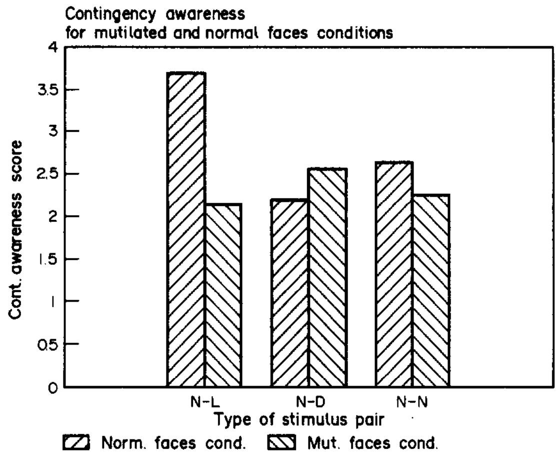

FIG. 2. Contingency awareness scores in 'normal faces' and 'mutilated faces' conditions, for neutral–like (N–L), neutral–dislike (N–D), and neutral–neutral (N–N) stimulus pairs.

other hand, the formula [awareness $\text{score}_{(N-L)}$ + awareness $\text{score}_{(N-D)}$] provides a general index of contingency awareness for a particular subject. If contingency awareness is a necessary condition for evaluative conditioning, we should predict a positive correlation between the number of contingencies a subject is aware of (as indicated by the 'general index of contingency awareness') and the extent of the evaluative conditioning effect (as indicated by the 'evaluative response differentiation score'). We obtained no correlation between the 'differentiation scores' and the 'general awareness scores' ($r = 0.03$; ns).

Discussion

Firstly, we were able to replicate the basic evaluative conditioning phenomenon. The mere contingent presentation of affective–evaluative neutral stimuli with actively liked or disliked stimuli was demonstrated to be sufficient to change the affective–evaluative tone of the originally neutral stimuli into the positive or negative direction, respectively. At first glance, inspection of Fig. 1 might give the impression that there is some asymmetry between the influence of (N–L) and (N–D) stimulus pairs; this impression disappears however when one recognizes the importance of evaluating the influence of (N–L) and (N–D) pairs *relative to the influence of the control (N–N) pair.* As a matter of fact, mere repeated presentation of an evaluatively neutral stimulus contingent with another neutral stimulus resulted in the first N stimulus acquiring some positive evaluative value. We interpret this finding as coming close to the 'mere exposure' phenomenon, in which the mere repeated exposure of a stimulus results in an increased liking for the stimulus (Zajonc, 1968; Kunst-Wilson and Zajonc, 1980; Vanbeselaere, 1983).

Secondly, and of crucial importance, our data are clearly supportive to the hypothesis that evaluative conditioning is resistant to extinction. In the standard Pavlovian conditioning preparation, a few unreinforced presentations of the CS are generally sufficient to cause abolishment or at least a strong diminution of the conditioned response (Dawson and Shell, 1985; Atkinson *et al.*, 1987). By contrast, we observed that 5 or even 10 unreinforced presentations of the N stimuli did not have any influence on the previously acquired evaluative value of the stimuli. The crucial affective–evaluative differentiations between N stimuli that participated in (N–L), (N–D) or (N–N) acquisition pairs could still be observed after the extinction procedure, and this at a highly significant level. Questions could be raised however concerning the validity of the data (ER3 measurement) in support of the resistance-to-extinction hypothesis. It cannot be excluded that the evaluative response measurement after the extinction phase (ER3) was influenced by demand characteristics. In fact, the time interval between the ER2 measurement and the ER3 measurement was only 15 to 20 min, so that subjects might well have remembered their ER2 scores at the moment of ER3 measurement. If one believes subjects hold the hypothesis that it is important to demonstrate some consistency in one's (evaluative) reactions to be true, and if one accepts that subjects try to behave according to this hypothesis, subjects might well have 'evaluated' the N stimuli (after the extinction phase) based on their memory of their ER2 scores. In order to be able to exclude this interpretation, we conducted a follow-up experiment about 2 months after the main experiment (see below).

Thirdly, the contingency awareness data gave no support for the hypothesis that the nature of the US is a crucial determinant for the occurrence of signal-learning. Contrary to our expectations, subjects in the 'mutilated faces' conditions did not demonstrate higher contingency awareness scores. We only observed some tendency for a better contingency awareness score in the 'normal faces' conditions in respect of the (N–L) contingencies than in respect of the (N–N) and (N–D) contingencies; also, the contingency awareness score for the (N–L) pairs was higher in the 'normal faces' conditions than in the 'mutilated faces' conditions. *Post hoc*, this could be interpreted as follows: in a context in which no strongly negative events occur, subjects are especially attentive to predictors of the occurrence of positive (or liked) events; this tendency is disrupted by the occurrence of strongly negative events. Alternative explanations are equally plausible however: maybe the 'mutilated faces' were not intrusive and/or aversive enough to 'force' the subject to engage in a controlled search for predictors. At present it seems best to consider the proposition which states that the nature of the US is a determinant for the occurrence of signal-learning as an hypothesis worth further experimental investigation. As a consequence of these results, the further question of whether evaluative learning and signal-learning relate to each other in a mutually exclusive or in a coordinate manner, could not be answered.

Fourthly, the near zero correlation between the general index of contingency awareness and the general index of evaluative conditioning, can be considered a further corroboration of our previous findings (Baeyens *et al.*, 1988a) that evaluative conditioning does not require and is not influenced by contingency awareness.

THE FOLLOW-UP STUDY

As stated in the discussion section, our data suggesting that evaluative conditioning is indeed resistant to extinction, are vulnerable to alternative interpretations (memory effects and demand characteristics). In order to rule out the possibility of memory effects and demand characteristics, we decided to retest *[evaluative response, (ER4)]* the subjects about two months after the experiment had taken place. If we could still obtain the affective–evaluative discriminations two months later, it would be at least implausible to attribute this effect to recall of the ER2 or ER3 scores. As a consequence, to the extent we could obtain ER4 results comparable to the ER3 results, a much stronger argument would be available in favor of the resistance-to-extinction hypothesis.

Secondly, we wanted to test for the possibility of a *phenomenal dissociation* between recognition memory and affective–evaluative discriminations made by the subjects (see for example Zajonc, 1980, 1984; Bornstein *et al.*, 1987). More specifically, we asked whether there would be instances of subjects not being able to recognize pictures as previously seen or not seen (old/new), but nevertheless demonstrating persistence of the evaluative conditioning effect. As there is evidence for recognition and recall memory being context-dependent (e.g. Smith *et al.*, 1978; Eich, 1980; Eich and Birnbaum, 1982), we tried to manipulate the context factor as follows. Half of the subjects were retested in the room in which they participated in the main experiment *(same context)*, whereas half of the subjects were retested in a completely different room *(different context)*. We hypothesized that subjects retested in the same context should do better at a recognition task (seen/not seen before) than subjects retested in a completely different context, but that there would be no difference between the two groups with respect to the evaluative conditioning effect.

Method

Subjects and design

Two months after the main experiment, 31 out of the 32 subjects who participated in the main experiment were available for retesting. Of each subset of 8 subjects participating in one of the four conditions of the main experiment (e.g. '5 extinction trials/mutilated faces condition'), 4 subjects were retested in the same context and 4 subjects were retested in the different context. The subjects of each subset were assigned randomly to one of the two context conditions.

Stimuli and test rooms

The set of 70 normal human face pictures and the set of 20 mutilated face pictures were used. In addition, a new set of 18 normal human face pictures was constructed. The new set of normal human faces was comparable to the set of 70 human faces described before. The subjects in the 'same context' condition were retested in the room referred to as the 'subject's room', as described earlier. The subjects participating in the 'different context' condition were retested in a psychophysiology laboratory, containing the standard psychophysiological measurement equipment.

Procedure

To continue the cover story used in the main experiment, subjects were told that ". . . at present, we had some convincing evidence demonstrating that the SCR measurement procedure used, provided a valid tool for the diagnosis of depression. In order to be able to use this procedure for a follow-up of the depressive patients, we should obtain some additional data concerning the question whether the evaluative reactions towards the stimulus materials showed cross-temporal variability or stability". Next, we presented the subject with a set of 36 pictures, containing the (standard) set of 18 new pictures randomly intermixed with the (idiosyncratically determined: 6 (N–L), 6 (N–D) and (N–N) pictures) 18 pictures used in the main experiment. The subject was asked to evaluate each picture, by putting a mark on the –100 . . . 0 . . . +100 evaluative response measurement line. We again stressed the importance of relying on the actual, immediate and spontaneous evaluative reactions towards the picture stimuli. When the ER4 measurement was completed, the experimenter conducted the recognition measurement. For each of the 36 stimuli, the experimenter asked:

(a) "Do you think whether or not you have seen this picture the previous time?"

(b) "When I would ask you to express the degree of confidence on the previous question by choosing between 'completely sure', 'half sure' or 'guess', what would you say?"

When a subject indicated having seen the picture before [question (a)], the following question was added:

(c) "Do you think your present evaluation of the picture has changed in comparison with the very first time you evaluated this picture? If yes, in what direction (more positive, more negative . . .)?"

When a subject expressed evaluative change, the following question was added:

(d) "What do you think is the reason for this change?"

All answers were registered by the experimenter on a pre-prepared response sheet.

Results

Evaluative responses two months after acquisition

All statistical analyses were conducted on mean ERs to N stimuli per subject [e.g. mean ER4 to N stimuli that participated (main experiment) in the three (N–L)-pairs]. A 2 × 3 × 2 analysis of variance ['same/different context' × 'type of stimulus pair: (N–L)/(N–N)/(N–D)' × 'moment of evaluative response measurement: baseline (ER1)/two months later (ER4)'; the first variable is a between-subject factor, the second and the third variables are within-subject factors] demonstrated a main effect of the 'type of stimulus pair: (N–L)/(N–N)/(N–D)' factor [$F(2,58) = 21.03$; $p<0.0001$].* No other main effects were obtained. We also observed an interaction between the 'type of stimulus pairs' × 'moment of evaluative response measurement' factors [$F(2,58) = 9.98$; $p<0.0005$].* A subsequent analysis of simple main effects demonstrated that the interaction was due to the fact that there was no effect of the variable 'type of stimulus pair' at the baseline (ER1) level [$F(2,116) = 2.85$; ns], but a highly significant effect at the ER4 (two months later) level [$F(2,116) = 31.94$; $p=0.0001$]. The main effect of the 'type of stimulus pair' at the ER4 level implied a difference between (N–L) and (N–D) pairs (=28.93), between (N–L) and (N–N) pairs (=16.74), and also between (N–D) and (N–N) pairs (=12.91) ($d_i = 11.07$; $p<0.01$) (Dunn test procedure for *a priori* comparisons) (see Fig. 3). No other interactions were observed.

*As the 'same context' condition contained the data of 15 subjects, and the 'different context' condition the data of 16 subjects, and as the drop-out of one subject in the 'same context' condition could be safely considered to be unrelated to the experimental manipulations, an *unweighted means solution* was used for the analysis of variance.

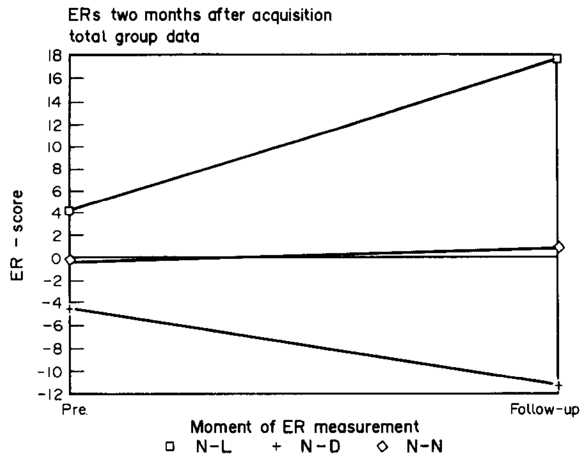

FIG. 3. Mean Evaluative Responses (ER) to originally neutral stimuli at baseline (PRE), and two months after the acquisition and extinction manipulations (FOLLOW-UP). [N-L = neutral followed by liked; N-D = neutral followed by disliked; N-N = neutral followed by neutral (= control)]. Mean over all conditions.

Recognition questionnaire data

Several measures were constructed out of the recognition questionnaire data [questions (a) and (b)]. These included: (1) *number of errors* (number of pictures seen before but not recognized as 'seen' + number of pictures not seen before but erroneously believed to be 'seen'); (2) *not recognized* (number of pictures seen before but not recognized as 'seen'); (3) *correct and completely sure* (number of pictures seen before and recognized as 'seen' + number of pictures not seen before and recognized as 'not seen'; only answers at the 'totally sure' confidence level were included); (4) *error and completely sure* (number of pictures not seen before but recognized as 'seen' + number of pictures seen before but not recognized as 'seen'; only answers at the 'totally sure' confidence level were included); (5) *general confidence level* (sum of confidence scores divided by number of responses; confidence levels were scored according to the criterion 'completely sure' = score of 2; 'half sure' = score of 1; 'guess' = score of 0). Means for these four measures are represented in Table 2, separately for the 'same context' and for the 'different context' conditions (columns 1 and 2); the results of *t*-tests (*t*-tests for independent data) for each of these measures are represented in column 3. Although there was a consistent tendency in the results of these analyses indicating superior performance in the 'same context' conditions, none of them reached significance (see Table 2).

Analysis of the subjective-awareness-of-evaluative-change data [ques-

Table 2. Recognition Questionnaire Data

	Same context	Different context	t
Number of errors	1.73	3.13	1.24
Not recognized	0.13	0.44	1.23
Correct and completely sure	29.33	27.50	1.07
Error and completely sure	0.33	0.75	0.89
General confidence level	1.78	1.75	0.49

tion (c)] revealed the following. With respect to the neutral stimuli that participated in (N–L) stimulus pairs, we obtained 76.3% 'no evaluative change' answers, 4.3% 'positive change', 17.2% 'negative change' and 2.1% 'do not know' answers. The factually observed evaluative shifts (ER4 – ER1) were +13.3, +28.3, +8 and +34, respectively. A similar picture emerged with respect to the N stimuli from (N–D) pairs: subjects expressed 68.8% 'no evaluative change', 11.8% 'positive change', 14% 'negative change' and 5.4% 'do not know' answers. The mean changes in evaluative responses were –7.9, –4, –10.3 and 8.6, respectively. In the (N–N) case, we obtained 72% 'no evaluative change', 7.2% 'positive change', 15% 'negative change' and 5.7% 'do not know' answers. The observed evaluative changes were 0.2, 22, –5.8 and 7.2, respectively. In the case where subjects were *not* aware of a change in their evaluative reactions, (N–L) and (N–D) stimulus pairs clearly *did* have a different effect ['no change' (N–L) = +13.3, 'no change' (N–D) = –7.9; $t(133) = 4.27$; $p<0.001$].* None of the subjects expressing a positive or negative change attributed this to the contingent stimulus presentations.

Discussion

Two months after the acquisition and extinction manipulations, the evaluative conditioning effect could still be observed at a highly significant level. This finding was interpreted as having two important consequences. Firstly, this is considered as being strong evidence for the validity of our interpretation of our ER3 data (evaluative responses after extinction procedure): evaluative conditioning indeed is resistant to extinction. Secondly, the results of the follow-up study clearly demonstrate that the mere passage of time has no influence on the acquired affective–evaluative value of a stimulus.

Recognition memory for the pictures presented two months before was of a very high accuracy. As most of the pictures were correctly recognized

*Data analysis based on assumption of independence; this assumption only partially holds: some data are from the same subject, some data are from different subjects.

as 'seen before', we were not able to make any reliable observations concerning the possibility of a dissociation between recognition memory and conditioned affective–evaluative reactions. Moreover, we could not observe any significant influence of the context manipulation (same/different context groups) on the accuracy of recognition memory.

Finally, the subjects were generally unaware of the (factually observed) shifts in their affective–evaluative reactions towards the stimulus materials.

GENERAL DISCUSSION

This section is restricted to a brief account of the implications of the single major finding of the present study. The evidence presented in this study allows us to fully subscribe to the Levey and Martin proposal which holds that ". . . the evaluative response, once it is conditioned to a neutral stimulus, cannot thereafter be extinguished through nonreinforcement. . ." (Levey and Martin, 1987, p.122). At a theoretical level, this can be interpreted as further evidence for our position which holds that evaluative conditioning is *not* based on the acquisition of propositional–declarative knowledge about the contingency between CS and US. We have already demonstrated that evaluative conditioning does not require and is not influenced by contingency-awareness (Baeyens *et al.*, 1988a). This finding did still allow, however, for the possibility of *implicit* (=not available for verbal awareness) propositional knowledge about the CS–US contingency being at the core of the evaluative conditioning phenomena. As explained in the introductory part, the resistance to extinction finding makes this interpretation very implausible.

Finally, we believe that there exists an intriguing analogy between this finding and the description of the *'laws of sympathetic magic'*. The laws of sympathetic magic were described by Frazer and Maus at the beginning of this century to account for magical belief systems in traditional cultures (Rozin *et al.*, 1986, 1987). As discussed more extensively in Rozin (1986, 1987), one of these laws, *contagion*, holds that there can be a permanent transfer of properties from one object to another by brief contact; hence, *"once in contact, always in contact"*. Rozin *et al.* (1986) provided some convincing evidence that this law fits well a variety of affective–evaluative behaviors in Western (American) culture (disgust, food preference, liking and disliking of clothes, and other personal objects). Moreover, Rozin *et al.* (1986, 1987) conceive of the apparent similarities between the laws of sympathetic magic and the laws of association (more particularly, the phenomenon of evaluative conditioning). The point to be made here is that the findings of the present study seem to provide an explicit experimental demonstration of the validity of the "once in contact, always in contact" intuition, expressed in the 'law of contagion'.

REFERENCES

Atkinson, R. L., Atkinson, R. C., Smith, E. E. and Hilgard, E. R. (1987) *Introduction to Psychology*. Chapter 7: Learning and conditioning. Harcourt Brace Jovanovich, New York.

Baeyens, F., Eelen, P. and Van den Bergh, O. (1988a) Contingency awareness in evaluative conditioning: a case for unaware affective–evaluative learning. *Cognit. Emot.* (in press).

Baeyens, F., Eelen, P., Crombez, G. and Van den Bergh, O. (1988b) Weten dat en houden van: signaal-leren in confrontatie met affectief–evaluatief leren. *Gedragstherapie* **21**, 197–216.

Bornstein, R. F., Leone, D. R. and Galley, D. J. (1987) The generalizability of subliminal mere exposure effects: Influence of stimuli perceived without awareness on social behavior. *J. Personality soc. Psychol.* **53**, 1070–1079.

Davey, G. (1987) *Cognitive Processes and Pavlovian Conditioning in Humans*. John Wiley and Sons, New York.

Dawson, M. E. and Shell, A. M. (1985) Information processing and human autonomic classical conditioning. In: P. K. Ackles, J. R. Jennings and M. G. H. Coles (Eds) *Advances in Psychophysiology*, Vol. 1, pp. 89–165. JAI Press, Greenwich, CT.

Dawson, M. E. and Shell, A. M. (1987) Human autonomic and skeletal classical conditioning: The role of conscious factors. In: F. Davey (Ed.) *Cognitive Processes and Pavlovian Conditioning in Humans*, pp. 27–57. John Wiley and Sons, New York.

Eelen, P., Van den Bergh, O. and Baeyens, F. (1988) Leerpsychologie en gedragstherapie. In: J. W. G. Orlemans, W. Brinkman, W. P. Haaijman and E. J. Zwaan (Eds) *Handboek voor Gedragstherapie*. Van Loghum Slaterus, Deventer (in press).

Eich, J. E. (1980) The cue-dependent nature of state-dependent retrieval. *Mem. Cognit.* **8**, 157–173.

Eich, J. E. and Birnbaum, I. M. (1982) Repetition, cuing and state-dependent memory. *Mem. Cognit.* **10**, 103–114.

Garcia, J., Rusimak, K. W. and Brett, I. P. (1977) Conditioning food-illness aversions in wild animals: Caveant Canonici. In: H. Davis and H. M. B. Hurwitz (Eds) *Operant–Pavlovian Interactions*. Erlbaum, Hillsdale, NJ.

Kunst-Wilson, W. R. and Zajonc, R. B. (1980) Affective discrimination of stimuli that cannot be recognized. *Science*, **207**, 557–558.

La Berge, D. (1981). Automatic information processing: A review. In: J. Long and A. Baddeley (Eds) *Attention and Performance IX*, pp. 173–185. Erlbaum, Hillsdale, NJ.

Levey, A. B. and Martin, I. (1975) Classical conditioning of human 'evaluative' responses. *Behav. Res. Ther.* **13**, 221–226.

Levey, A. B. and Martin, I. (1983) Cognitions, evaluations and conditioning. Rules of sequence and rules of consequence. *Adv. Behav. Res. Ther.* **4**, 181–195.

Levey, A. B. and Martin, I. (1987) Evaluative conditioning: a case for hedonic transfer. In: H. J. Eysenck and I. Martin (Eds) *Theoretical Foundations of Behavior Therapy*. Plenum Press, New York.

Mackintosh, N. J. (1983). *Conditioning and Associative Learning*. Oxford University Press, New York.

Mackintosh, N. J. (1987). Animal minds. In: C. Blakemore and S. Greenfield (Eds) *Mindwaves: Thoughts on Intelligence, Identity and Consciousness*. Basil Blackwell, Oxford.

Martin, I. and Levey, A. B. (1978) Evaluative Conditioning. *Adv. Behav. Res. Ther.* **1**, 57–102.

Martin, I. and Levey, A. B. (1985). Conditioning, evaluations and cognitions: an axis of integration. *Behav. Res. Ther.* **23**, 167–175.

Martin, I. and Levey, A. B. (1987) Learning what will happen next: conditioning, evaluation and cognitive processes. In: G. Davey (Ed.) *Cognitive Processes and Pavlovian Conditioning in Humans*, pp. 57–82. Wiley and Sons, New York.

Öhman, A. (1979) The orienting response, attention and learning: an information-processing perspective. In: H. D. Kimmel, E. H. van Holst and J. F. Orlebeke (Eds) *The Orienting Reflex in Humans*. Erlbaum, Hillsdale, NJ.

Öhman, A. (1983) The orienting response during Pavlovian conditioning. In: D. A. T. Siddle (Ed.) *Orienting and Habituation: Perspectives in Human Research*, pp. 315–369. John Wiley and Sons, Chichester, England.

Öhman, A. (1986) Face the beast and fear the face: Animal and social fears as prototypes for evolutionary analyses of emotion. (Presidential Address, 1985). *Psychophysiology,* **23,** 123–145.

Rescorla, R. A. (1980) *Pavlovian Second-order Conditioning: Studies in Associative Learning*. Erlbaum, Hillsdale, NJ.

Rescorla, R. A. and Wagner, A. R. (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: A. H. Black and W. F. Prokasy (Eds) *Classical Conditioning*, Vol. 2. Appleton-Century-Crofts, New York.

Rozin, P. (1986) One-trial acquired likes and dislikes in humans: Disgust as a US, food predominance and negative learning predominance. *Learn. Motiv.* **17,** 180–189.

Rozin, P. and Fallon, A. E. (1987) A perspective on disgust. *Psychol. Rev.* **94,** 23–41.

Rozin, P., Millman, L. and Nemeroff, C. (1986) Operations of the laws of sympathetic magic in disgust and other domains. *J. Personality soc. Psychol.* **50**, 703–712.

Schneider, W. and Shiffrin, R. M. (1977) Controlled and automatic human information processing. I. Detection, search and attention. *Psychol. Rev.* **84**, 1–66.

Seligman, M. E. P. and Hager, J. L. (Eds) (1972) *Biological Boundaries of Learning.* Appleton-Century-Crofts, New York.

Shiffrin, R. M. and Schneider, W. (1977) Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychol. Rev.* **84,** 127–190.

Shiffrin, R. M., Dumais, S. T. and Schneider, W. (1981) Characteristics of automatism. In: J. Long and A. Baddeley (Eds) *Attention and Performance IX*. Erlbaum, Hillsdale, NJ.

Smith, S. M., Glenberg, A. and Bjork, R. A. (1978) Environmental context and human memory. *Mem. Cognit.* **6**, 342–352.

Vanbeselaere, N. (1983) Mere exposure: a search for an explanation. In: W. Doise and S. Moscovici (Eds) *Current Issues in European Social Psychology*, Vol. 1. Cambridge University Press, Cambridge.

Wagner, A. R. (1978) Expectancies and the priming of STM. In: S. H. Hulse, H. Fowler, and W. K. Honig (Eds) *Cognitive Processes in Animal Behavior.* Erlbaum, Hillsdale, NJ.

Zajonc, R. B. (1968) Attitudinal effects of mere exposure. *J. Personality soc. Psychol. Monogr.* **9** (2,Pt.2), 1–27.

Zajonc, R. B. (1980) Preferences need no inferences. *Am. Psychol.* **35,** 151–175.

Zajonc, R. B. (1984). On the primacy of affect. *Am. Psychol.* **39,** 117–123.