Backward conditioning to a CS following an expected vs. a surprising UCS

ALLAN R. WAGNER and WILLIAM S. TERRY
Yale University, New Haven, Connecticut 06520

Rabbits were trained in eyelid conditioning with a "backward" arrangement of unconditioned stimulus (UCS) followed by conditioned stimulus (CS). When such a CS was tested alone it was observed to produce substantial conditioned responding if the UCS had been arranged to be "surprising" during the backward pairings, but not if it had been arranged to be "expected." The comparisons were made in a within-subjects design where the surprisingness of the UCS on the different pairing occasions was manipulated by preceding the UCS by discriminative CSs which were otherwise either never followed by the UCS (CS—) or consistently followed by the UCS (CS+). The results may have implications for the nonmonotonic course of responding seen during backward conditioning, as a UCS is at first surprising, but then expected on the basis of contextual cues.

It is frequently asserted that backward conditioning does not occur (e.g., Gormezano & Moore, 1969; Kimble, 1961; Osgood, 1953). In fact, if a conditioned stimulus (CS) and an unconditioned stimulus (UCS) are paired in a so-called backward relationship of UCS followed by CS, there are predictable response tendencies which may be seen to the CS. Over a small number of pairings, responding to the CS is likely to increase (e.g., Champion & Jones, 1961; Pavlov, 1928; Heth, 1974). With more extensive training, responding is likely to decrease to an untrained level (e.g., Smith, Coleman, & Gormezano, 1969; Spooner & Kellogg, 1947; Heth, 1974). And with sufficient training the CS is likely to be acted toward as an "inhibitory" stimulus, serving to depress the responding otherwise seen to cues with which it is compounded (e.g., Moscovitch & LoLordo, 1968) or showing retarded acquisition of conditioned response tendencies as a result of subsequent "forward" pairings (e.g., Siegel & Domjan, 1971).

One interpretation of this pattern is that the *associative* product of backward pairings is inhibitory, as is eventually seen after extensive training, but that nonassociative factors such as "sensitization" are responsible for the initial "pseudoconditioned" responding (e.g., Bernstein, 1934; Harris, 1941). The transient nature of the latter responding is indeed what marks it, according to this view, as different from conditioned responding (e.g., Plotkin & Oakley, 1975).

Alternatively, one might attempt to encompass both the initial responding and eventual inhibitory tendencies to a backward CS within an associative interpretation. The present paper will suggest one way this might be done and report some initial data consistent with the possibility.

In a series of recent papers (e.g., Terry & Wagner, 1975; Wagner, 1971; Wagner, Rudy, & Whitlow, 1973; Wagner, in press), we have proposed that CS-UCS association depends upon the postepisodic rehearsal of the two events in short-term memory, and that UCS rehearsal varies with the degree to which the UCS is predicted by the aggregation of cues which precede it. These notions, with some elaboration, may be useful in accounting for the effects of different temporal arrangements of CS and UCS, and specifically for the nonmonotonic variation in responding to a CS during backward conditioning.

We shall need to assume that the net associative tendency accruing to a CS during any interval of time embracing a CS and UCS will depend upon the amount of joint CS and UCS rehearsal and the amount of CS alone rehearsal that is afforded: Joint rehearsal should increase the tendency for the CS to be acted toward as a signal for the UCS while CS-alone rehearsal should decrease this tendency. According to this view, a forward CS-UCS pairing is more likely to produce conditioned responding than is a backward UCS-CS pairing, as in the latter case CS-alone rehearsal is likely to continue after UCS rehearsal has been terminated. However, it also suggests that any manipulation which changes the duration of UCS rehearsal should influence the degree of backward conditioning: The more protracted is UCS rehearsal, the more likely it should overlap with CS rehearsal to promote association between CS and UCS and deny CS-alone rehearsal that would diminish the association.

If it is assumed (e.g., Terry & Wagner, 1975; Wagner, 1971; Wagner, Rudy, & Whitlow, 1973) that the same UCS will be rehearsed longer when it is "surprising" than when it is "expected," it is possible to account for a transient phase of conditioned responding during backward conditioning. The initial UCS-CS pairings should involve a surprising UCS, favoring associative

learning. As the UCS comes to be expected on the basis of contextual cues, further UCS-CS pairings should more principally induce CS-alone rehearsal as opposed to joint CS and UCS rehearsal and lead to extinction (and eventually inhibitory tendencies). This analysis would lead to a further expectation. That is, that more persistent backward conditioning should be producible by maintaining the surprisingness of the UCS over the course of the UCS-CS trials.

The study to be reported was addressed to this interpretation. In essence, it sought to evaluate whether backward conditioning would be favored by specifically arranging to make the UCS involved on UCS-CS trials surprising rather than expected. This was done by preceding the UCS on backward conditioning trials by either a cue (CS+) previously trained to predict the occurrence of the UCS, or by a cue (CS−) previously trained to predict the absence of the UCS. A CS− - UCS - CS sequence should produce more robust conditioned responding to the backward CS than should a CS+ - UCS - CS sequence. The investigation involved a within-subject comparison employing eyelid conditioning of the rabbit.

METHOD

Subjects

The subjects were 32 male New Zealand White rabbits weighing between 1.8 and 2.7 kg at the start of training. Each was individually housed and maintained with ad-lib food and water, except during the experimental sessions. Two rabbits died during the course of the investigation, and one rabbit was dropped prior to backward conditioning for failure to maintain the conditioned discrimination on which all subjects were trained. This left 29 subjects from which data are reported.

Apparatus

The subjects were trained in one of eight identical 66 x 48 x 48 cm isolation chambers. Each chamber was completely lined with aluminum foil to provide a homogeneous visual surround, dimly illuminated by a 15 W neon light. The chambers were maintained at a 67 dB (re. 20 μN/m^2) sound pressure level (General Radio Meter No. 1551-C, A-scale) by the presentation of white noise through a 12.7-cm loudspeaker. The rabbit was further confined within the chamber in a 51 x 18 x 14 cm Plexiglas restraining box, through which only its head protruded.

Four conditioned stimuli were used in the present study. There were two auditory stimuli: a 3500-Hz tone, presented through the same speaker as the white noise (located behind the subject on the rear wall of the chamber), and a train of 30/sec clicks produced by a relay mounted in a metal sounding box (located on the floor of the chamber, about 17 cm to the left of the subject's head). The intensity of the auditory stimuli was approximately 10 dB over the background noise level. A 15/sec flashing light CS was generated by a Knight KG-323 strobe lamp, also located behind the subject so as to diffusely reflect from the walls of the chamber. A vibrotactual CS could also be presented by means of a hand-massager (Valmour Model 880) mounted on the floor of the restraining box so as to maintain firm contact with the rabbit's chest. The unconditioned stimulus was a train of 100/sec 5-mA square-wave shock pulses produced by an Argonaut LRA-046 constant current generator, and delivered through stainless steel electrodes (Sklar surgical wire, 32 ga) sutured in the skin about the orbit of the rabbit's right eye. One electrode was implanted approximately 5 mm below the extreme nasal extent of the eye and the other approximately 5 mm above the extreme lateral extent.

Closures of the subject's right eyelid were monitored by a microtorque potentiometer taped to its head and communicating with the lid by a length of thread hooked to a small permanent suture loop. Movements of the eyelid turned a counterweighted wheel affixed to the axle of the potentiometer, and were graphically recorded on a Beckman Dynograph with an effective gain of four.

Procedure

On the day prior to the start of training, each subject was habituated to confinement in the restraining box outside of the conditioning chamber for 2 h. During this time, the subject's head was shaved and the electrodes were implanted.

All subjects then received training on a Pavlovian discrimination between the two auditory CSs, each 1100 msec in duration. For half the subjects, the tone CS was reinforced (CS+) by the presentation of the shock UCS during the last 100 msec of the CS, while the click CS was nonreinforced (CS−). This cue designation was reversed for the remaining subjects. Thirteen sessions of training were given, with 10 CS+ and 10 CS− trials presented daily in a sequence balanced for first-order transitions. The mean intertrial interval was four minutes.

The subjects were then divided into two groups, each including an equal number of animals trained with tone and click as the CS+. For both groups the light and vibrotactual stimuli were introduced, and designated as "CS$_E$" and "CS$_S$." One group of 16 subjects received backward conditioning, involving training and test trials with these CSs embedded within the same sequence of discrimination trials as previously experienced. On certain days, these subjects received a CS+ - UCS - CS$_E$ episode as Trial 7 of the session, and a CS$_E$-alone test trial as Trial 16. On alternate days, the subjects received a CS− - UCS -CS$_S$ episode as Trial 7, with CS$_S$-alone on Trial 16. Thus, for the backward conditioned group, CS$_E$ consistently followed an "expected" UCS in the backward conditioning episode and CS$_S$ followed a "surprising" UCS. The presentation of either CS$_E$ or CS$_S$ alone on Trial 16 provided a test for the conditioning which had occurred to that backward CS presented earlier in the session.

For half of the subjects in the backward conditioned group, the flashing light stimulus was CS$_E$ and the vibrotactual stimulus was CS$_S$, this being reversed for the remaining subjects. These CSs were 500 msec in duration on the conditioning trials with their onsets following UCS offset by 500 msec. The duration of CS$_E$ and CS$_S$ on test trials was 500 msec for half of the subjects, and 1000 msec for the other half. The designation of discrimination cues (tone or click as CS+), backward cues (light or vibrotactual stimulus as CS$_E$), and duration of the last test trial CSs were appropriately counterbalanced.

The sequence of backward conditioning episodes over days followed, for different subjects, an abba or baab pattern, where a was the CS+ - UCS - CS$_E$ episode (with CS$_E$ on the test trial), and b was the CS− - UCS - CS$_S$ episode (with CS$_S$ on the test trial). The discrimination sequence also varied across sessions, so that the trials preceding Trials 7 and 16 were either CS+ - UCS or CS− episodes, on alternate days. Backward conditioning continued for 48 days, thus giving each subject 24 training and test trials with each of CS$_E$ and CS$_S$.

The remaining 16 subjects were assigned to a nonpaired group. This group experienced a similar course of events to the backward conditioned group, except that CS$_E$ and CS$_S$ were not presented following the expected or surprising UCS on Trial 7. These subjects simply received either CS+ - UCS on Trial 7 and CS$_E$ alone on Trial 16, or CS− - UCS on Trial 7 and CS$_S$ alone on Trial 16. This group was included to provide an assessment of the amount of responding to CS$_E$ and CS$_S$ in the absence of
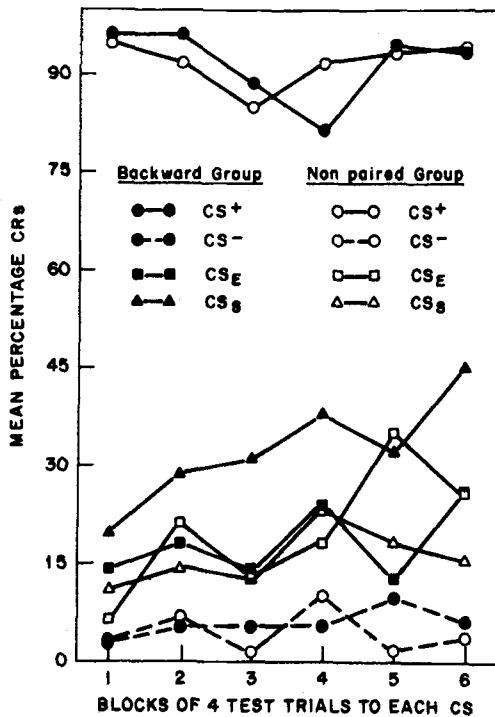
Figure 1. Mean percentage conditioned responses over blocks of eight test trials to each of CS+ and CS− on Trial 7, and to CS_E and CS_S on Trial 16, of daily sessions, plotted separately for the backward conditioned group and the nonpaired group.

backward pairings with the surprising and expected UCSs, in subjects otherwise trained identically to the backward conditioned group. As in the case of the backward conditioned subjects, for half the nonpaired subjects the flashing light was CS_E and the vibrotactual stimulus was CS_S, this being reversed for the other subjects. The test-CS duration on Trial 16 was 500 msec or 1000 msec for different halves of the nonpaired group. The sequence of expected and surprising UCS episodes (and thus also CS_E vs. CS_S presentations) across days followed the same schedule as that of the backward conditioned group. Training continued for 48 sessions, giving each subject 24 test trials on both CS_E and CS_S.

The two subjects that died were both in the backward conditioned group. The one subject that was dropped was a nonpaired animal. These losses did not materially affect the results, in that inspection of the resulting data showed no notable differences among the means of the several subgroups of subjects through which counterbalancing of cues and CS-test durations was effected.

Scoring

A single criterion for recording a conditioned response (CR) was used for all CSs. Eyelid closures ⩾ 1.25 mm (5-mm pen deflection) occurring between 140 and 1000 msec after CS onset were counted as CRs. Conditioned responding to the backward-paired and nonpaired CSs was monitored only on the CS-alone test trial of each session, with the above criterion being used for subjects receiving a 500-msec test CS and for those receiving a 1000-msec test CS.

RESULTS

The initial phase of discrimination training was sufficient to produce differential eyelid responding to CS+ and CS− in all subjects. On the first day of the

backward conditioning phase of the study, each subject showed at least a 50% difference in the mean percentage responding to the two CSs, with a mean over all subjects of 94% CRs to CS+ and 4% to CS−. Discriminative responding to CS+ and CS− was not appreciably affected by the interpolated training and test trials with CS_E and CS_S. This can be seen in Figure 1, which plots over blocks of 8 sessions the mean percentage conditioned responses to CS+ and CS− on Trial 7 of the several sessions when the cues were always followed by a UCS, and, in the case of the backward conditioned group, more remotely by CS_E or CS_S. The differential responding to CS+ and CS− on these trials offers assurance that the subsequent UCS may be considered differentially expected throughout the course of training. During this phase of the experiment, the backward conditioned and the nonpaired groups did not differ in their mean level of response to either CS+ or CS−, Fs < 1.

Also shown in Figure 1 are the data of primary interest, involving the mean percentage responding to CS_E and CS_S over successive blocks of 8 total test trials for the backward conditioned and nonpaired groups. The first point to be noted concerning these data is that the percentage responding of the nonpaired group to CS_E and CS_S was consistently above zero, even though the stimuli were never "paired" with reinforcement. Responding to one or both of the nonpaired CSs was shown by every subject, with no consistent difference in responding to CSs occurring on expected or surprising UCS days. This level of responding may be taken to indicate the net generalized response tendencies from the previously trained CS+ and CS−.

For the backward conditioned subjects, responding also occurred to CS_S and CS_E. But, in this case, there was a substantially higher level of responding to the CS which had followed a surprising UCS on conditioning trials (CS_S) than to the CS which had followed an expected UCS (CS_E). Responding to the CS_E did not differ from that of the nonpaired groups to either CS_E or CS_S. However, responding to CS_S was greater than that of the nonpaired group to either cue. (The only deviation in this pattern being the inexplicably high level of responding of the nonpaired group to CS_E on Block 5.) These observations were verified statistically by an unweighted means analysis of variance of the mean percentage CRs across all test trials. The analysis showed a significant interaction between the two groups of subjects (Backward Conditioned vs. Nonpaired) and the two test stimuli (CS_S vs. CS_E), [$F(1/20) = 4.58$, $p < .05$]. Thus, for the backward conditioned group responding was reliably greater to the CS which had followed a surprising UCS than to the CS which had followed an expected UCS [$t(20) = 2.22$, $p < .05$], while for the nonpaired group, responding did not reliably differ to those CSs tested on surprising vs. expected UCS days, [$t(20) = .81$]. Comparisons using Scheffé's test further substantiated these findings,

indicating that the responding to CS_S in the backward conditioned group was higher than to CS_E in the backward conditioned group, and to CS_E and CS_S of the nonpaired group (ps < .05), while the latter three conditions did not differ among themselves. There were no significant differences attributable to the CS durations on test trials (500 or 1000 msec), or cue designations for CS_E and CS_S (flashing light or vibrotactual stimuli), and no reliable interactions of these variables with either the treatment groups or the CS_E vs. CS_S treatment conditions.

There was some tendency for responding to increase to both CS_E and CS_S in the two groups over the course of testing. However, this tendency was especially prominent only in the case of CS_S of the backward conditioned group which had a level of 19% responding on the first block of trials and a level of 45% responding on the final block of trials, as shown in Figure 1. Analysis of the mean level of responding on the first and last blocks of trials revealed a reliable Groups by CS by Trials interaction [$F(1/27) = 16.02$, $p < .01$]. Subsequent comparisons indicated that the increase in responding on the final as compared to the first block of testing was greater to CS_S of the backward conditioned group than to CS_E of the backward group, [$t(27) = 3.63$, $p < .01$], than to CS_S of the nonpaired groups, [$t(27) = 6.07$, $p < .01$], and than to CS_E of the nonpaired group, [$t(27) = 1.70$, $p = .05$, one tailed].

DISCUSSION

It should first be noted that there are several procedural differences between the present experimental design and previous studies of backward conditioning, which may have contributed to the amount of responding observed to CS_S of the backward conditioned group. For example, the discrimination pretraining may have been important in increasing the detectability of conditioning, as differential backward conditioning was observed within a few trials, compared with the more extensive training needed to form the original discrimination. Also, the spacing of the backward conditioning trials, one per daily session, may have allowed more conditioning to be observed with a relatively small number of pairings than would have been seen by using a more massed schedule of presentations.

The principal findings, however, are clear in indicating that one determiner of the conditioning to a CS which is arranged to follow a UCS is the degree of "surprisingness" of the UCS. Demonstrable backward conditioning was obtained when the UCS on conditioning trials was made surprising by preceding it with a CS−, while no conditioning was found when the UCS was made expected by preceding it with a CS+. In the former case more responding was observed than in a comparable nonpaired condition, while in the latter case no more responding was observed. The general

suggestion which has been offered is that evidence of "excitatory" conditioning to a CS will depend upon the relative amounts of joint, CS and UCS, rehearsal vs. CS-alone rehearsal. If a surprising UCS is rehearsed longer in short-term memory than is an expected UCS (e.g., Terry & Wagner, 1975; Wagner et al., 1973), then it is reasonable to suspect that it will provoke more overlap in short-term memory with a subsequently presented CS, and hence produce more "backward" conditioning. The present results are consistent with this reasoning.

The associative product of backward conditioning pairings is commonly inferred from the asymptotic performance to the backward CS. Since the latter performance often indicates the acquisition of inhibitory properties by the CS (e.g., Moscovitch & LoLordo, 1968; Siegel & Domjan, 1971), responding on early conditioning trials (e.g., Pavlov, 1928; Spooner & Kellogg, 1947), which might otherwise be taken to reflect excitatory conditioning, is often attributed (e.g., Bernstein, 1934; Harris, 1941; Plotkin & Oakley, 1975) to nonassociative factors. While sensitization or other pseudoconditioning processes may indeed contribute to the responding seen to a backward (or forward) trained CS, the present data would invite more serious consideration of an associative basis for early trial responding in spite of the ultimate asymptotic performance. Excitatory conditioning may occur on the initial backward trials, as the UCS on such trials is relatively surprising. Extinction and inhibitory conditioning may subsequently occur, with more extended training, as the later trials involve a UCS which has become more expected due to contextual associations. Steady-state performance in this instance may simply not be very illuminating of the dynamics of acquisition.

Enhanced conditioning, as presently observed, with a surprising as opposed to an expected UCS is not a novel observation. For example, it is well known (Kamin, 1968) that a conditioned stimulus is less responded to alone when it has been reinforced in simultaneous compound with another stimulus previously trained to predict the occurrence of the UCS (i.e., a CS+), than when it has been reinforced alone. And it has been found (Wagner, 1971) that a conditioned stimulus is *more* responded to alone when it has been reinforced in simultaneous compound with another stimulus previously trained to predict the absence of the UCS (i.e., a CS−) than when it has been reinforced alone. The modulating effects of having a CS+ or a CS− precede the UCS with which a target CS is paired are notable in the present case only in that the target CS was not presented in compound with a CS+ or CS−, preceding the UCS, but was presented after the UCS. It thus appears that the differential effects on conditioning of expected or surprising UCSs are not limited to a CS which precedes the UCS, but rather occur to a CS which is, in some

more general fashion, contemporaneous with the UCS or its processing in memory.

In discussing the notion that conditioning depends upon the relative amounts of CS-UCS vs. CS-alone rehearsal generated by a training episode, we have presently concentrated upon the phenomenon of "backward" conditioning. However, the utility of such a view must be measured by its ability to integrate a variety of data involving the different possible arrangements of CS and UCS. In this context it is notable that Krane and Wagner (1975) appealed to a congenial set of propositions in attempting to account for the fact that a saccharin-taste CS was well conditioned by an electric shock UCS when a long (forward) CS-UCS interval was used, but not when a shorter interval was employed. The lack of conditioning to a gustatory CS with a short CS-UCS interval may resemble what occurs in many backward conditioning studies. If a novel taste CS is processed for some considerable duration of time after the stimulus has been removed, it may, with a short CS-UCS interval, be processed beyond the time of UCS rehearsal, to the detriment of conditioning. Some of the apparent complexities of associative learning may be better understood in terms of the processing that occurs during and after stimulation, as opposed to the simple sequencing of stimulus onsets.

REFERENCES

Bernstein, A. L. Temporal factors in the formation of conditioned eyelid reactions in human subjects. Journal of General Psychology, 1934, 10, 173-197.

Champion, R. A., & Jones, J. E. Forward, backward, and pseudoconditioning of the GSR. Journal of Experimental Psychology, 1961, 62, 58-61.

Gormezano, I., & Moore, J. W. Classical conditioning. In M. H. Marx (Ed.), Learning: processes. Toronto: Macmillan, 1969.

Harris, J. D. Forward conditioning, backward conditioning, pseudoconditioning and adaptation to the conditioned stimulus. Journal of Experimental Psychology, 1941, 28, 491-502.

Heth, C. P. Simultaneous and backward Pavlovian conditioning as a function of number of CS-UCS pairings. Unpublished doctoral dissertation, Yale University, 1974.

Kamin, L. J. "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), Miami symposium on the prediction of behavior: Aversive stimulation. Miami: University of Miami Press, 1968.

Kimble, G. A. Hilgard and Marquis' conditioning and learning. New York: Appleton-Century-Crofts, 1961.

Krane, R. V., & Wagner, A. R. Taste aversion learning with a delayed shock US: Implications for the "generality of the laws of learning." Journal of Comparative and Physiological Psychology, 1975, 88, 882-889.

Moscovitch, A., & LoLordo, V. M. Role of safety in the Pavlovian backward fear conditioning procedure. Journal of Comparative and Physiological Psychology, 1968, 66, 673-678.

Osgood, C. E. Method and theory in experimental psychology. New York: Oxford University Press, 1953.

Pavlov, I. P. Lectures on conditioned reflexes. New York: International Publishers, 1928.

Plotkin, R. C., & Oakley, P. A. Backward conditioning in the rabbit (Oryctolagus cuniculus). Journal of Comparative and Physiological Psychology, 1975, 88, 586-590.

Siegel, S., & Domjan, M. Backward conditioning as an inhibitory procedure. Learning and Motivation, 1971, 2, 1-11.

Smith, M. C., Coleman, S. R., & Gormezano, I. Classical conditioning of the rabbit's nictating membrane response at backward, simultaneous, and forward CS-UCS intervals. Journal of Comparative and Physiological Psychology, 1969, 69, 226-231.

Spooner, A., & Kellogg, W. N. The backward conditioning curve. American Journal of Psychology, 1947, 60, 321-334.

Terry, W. S., & Wagner, A. R. Short-term memory for "surprising" vs. "expected" USs in Pavlovian conditioning. Journal of Experimental Psychology: Animal Behavior Processes, 1975, 1, 122-133.

Wagner, A. R. Priming in STM: An information processing mechanism for self-generated or retreival-generated depression in performance. In T. J. Tighe and R. N. Leaton (Eds.), Habituation: Perspectives from child development, animal behavior, and neurophysiology. Hillsdale, N. J.: Lawrence Erlbaum Associates (in press).

Wagner, A. R. Elementary associations. In H. H. Kendler & J. T. Spence (Eds.), Essays in neobehaviorism: A memorial volume to Kenneth W. Spence. New York: Appleton-Century-Crofts, 1971.

Wagner, A. R., Rudy, J. W., & Whitlow, J. W. Rehearsal in animal conditioning. Journal of Experimental Psychology, 1973, 97, 407-426.