Behavioral Neuroscience
2003, Vol. 117, No. 2, 292–303

0735-7044/03/$12.00 DOI: 10.1037/0735-7044.117.2.292

Covariation of Alternative Measures of Responding in Rabbit (*Oryctolagus cuniculus*) Eyeblink Conditioning During Acquisition Training and Tone Generalization

Keith S. Garcia and Michael D. Mauk
University of Texas Medical School at Houston

Gabrielle Weidemann and E. James Kehoe
The University of New South Wales

The likelihood, size, and speed of eyelid movements are thought to covary during the acquisition and expression of conditioning in rabbits (*Oryctolagus cuniculus*) and are generally accepted as interchangeable measures of the associative strength activated by the conditioned stimulus (CS). To test this assumption, the authors examined the patterns of covariation in these eyelid movement measures in acquisition and stimulus generalization in the upper eyelid and nictitating membrane. Rather than the expected covariation among these measures, eyelid movement magnitudes during the CS were distributed in approximately a bimodal manner. That is, eyelid activity consisted largely of a mixture of very small (< 0.125 mm) baseline measurements and larger (> 1 mm) movements. The results are discussed with respect to their implications for real-time models of eyelid conditioning.

Pavlovian eyeblink conditioning involves pairing a neutral conditioned stimulus (CS), such as a tone, with a noxious unconditioned stimulus (US), such as a periorbital shock or an air puff directed at the eye. With repeated presentations of the CS at a fixed time before the US, the CS alone comes to elicit a timed closure of the eyelids. This closure starts during the interval between the onset of the CS and the onset of the US and tends to peak around the end of the interval, near the time of US onset. This pattern has been seen in both the upper eyelid (Sears, Baker, & Frey, 1979) and the nictitating membrane (NM; Smith, 1968). Across successive CS–US pairings, the likelihood, size, and speed of NM movements all show increases (e.g., Leonard, 1975; Smith, 1968).

Because the features of these eyelid movements generally change in the same direction, they are assumed to be largely interchangeable indices of underlying associative learning. This assumption of interchangeability has been convenient for both theoreticians and researchers. It has allowed theorists to focus on the process of learning without specifying the exact rules for its expression (Pearce & Hall, 1980; Rescorla & Wagner, 1972). Methodologically, this assumption has permitted researchers to report their results by using only one of these measures. Going beyond rabbit eyeblink conditioning, the assumption of interchangeability has permitted researchers using different species and response systems to compare the results of similar behavioral and neural manipulations.

Despite its appeal, the assumption of interchangeability is not as safe as it might seem. There are two reasons for being cautious. The first is methodological in nature; it concerns a subtle difference in the definition of conditioned responding. The second reason is empirical: The different measures of eyelid movements during the CS have been observed to diverge systematically.

Regarding the definitional issue, *conditioned responding* is usually defined as a change in behavioral activity beyond the level arising from both systematic sources (e.g., pseudoconditioning) and unsystematic sources (e.g., spontaneous movements). For example, in heart rate (Schneiderman, 1972; Winters, McCabe, & Schneiderman, 2002) and potentiated startle (Brown, Kalish, & Farber, 1951) paradigms, the detection of conditioned responding entails a comparison between the level of activity during the CS versus the level of activity during either an alternative, nonreinforced stimulus or the pre-CS period. Thus, in general, no single occurrence of a target behavior during a single presentation of the CS can be denoted as a conditioned response (CR) with absolute confidence. There is always the possibility that strong activity during any given presentation of the CS may arise from a nonassociative source.

The same is true for the rabbit eyelid preparations, albeit to a lesser degree than in many other preparations. The systematic acquisition of large eyelid movements appears confined to paired CS–US training. A suite of nonassociative controls—for example, no-stimulus, CS-alone, US-alone, and unpaired presentations of the CS and US—has consistently failed to produce large eyelid movements in both the upper eyelid and NM preparations (Gormezano, Kehoe, & Marshall, 1983; Kehoe & Macrae, 2002). Likewise, the level of unsystematic "noise" in the rabbit eyelid preparations is low. That is, the resting position of the rabbit eyelid provides a baseline that is largely free of spontaneous eyelid movement, artifacts from whole-body movements, and electrical

Keith S. Garcia and Michael D. Mauk, Department of Neurobiology and Anatomy, University of Texas Medical School at Houston; Gabrielle Weidemann and E. James Kehoe, School of Psychology, The University of New South Wales, Sydney, Australia.

This research was supported by a National Institute of Mental Health grant to Michael D. Mauk and Australian Research Council grants to E. James Kehoe.

Correspondence concerning this article should be addressed to E. James Kehoe, School of Psychology, The University of New South Wales, Sydney, NSW 2052, Australia. E-mail: j.kehoe@unsw.edu.au

artifacts in the recording system. Nevertheless, some large eyelid movements do occur that are unrelated to CS–US pairings; for a baseline observation period of a few hundred milliseconds, the likelihood of observing an NM movement greater than 0.5 mm has been estimated to be less than 5% (Gormezano et al., 1983; Kehoe, Marshall-Goodell, & Gormezano, 1987).

Given the low level of noise in the rabbit eyelid preparations, there have been two strategies for measuring conditioned responding. One has been to use the maximal movement of the eyelid following CS onset. This measure, known as *magnitude*, entails all movements, even when only a very small movement (or no measurable movement) occurs. By using this measure, there is no need to adopt a criterion for denoting some movements as CRs while considering other movements to be too small to count as CRs. Changes in magnitude across trials and differences in magnitude among training conditions are simply recorded and reported. Despite its methodological elegance, this measurement strategy has at least two practical limitations. First, the maximal eyelid movement following CS onset usually occurs around the time of US onset and the unconditioned response (UR; Coleman & Gormezano, 1971; Millenson, Kehoe, & Gormezano, 1977; Sears et al., 1979; Smith, 1968). Thus, the magnitude measure is usually recorded only on CS-alone test trials, in which the maximal movement cannot be masked by the UR. Second, because the magnitude scale includes movements of zero (no measurable movement), measuring the onset time of an eyelid movement becomes impossible unless one adopts a minimum criterion. Accordingly, the second strategy has been to adopt such a criterion. Typically, eyelid movements of 0.5 mm or larger are counted as CRs. Such a criterion excludes zero movements, small flutters, and any electrical noise in the recording apparatus. By dichotomizing the range of movements into CR versus non-CR regions, it becomes possible to measure the timing characteristics of the defined CRs, for example, the latency of CR onset and the temporal location of the CR's peak. It also becomes possible to compute CR likelihood, that is, the proportion of trials that contain a supercriterion movement. A truncated version of the magnitude measure is retained in this second strategy. This measure is labeled *amplitude*, which is the maximal movement of the eyelid, provided it exceeds the criterion for the CR (Gormezano, 1966; Marshall-Goodell, Schreurs, & Gormezano, 1982).

Although both strategies have been used for the measurement of eyelid responding, they have not been directly compared. Gormezano (1966), who first identified these strategies, suspected that the magnitude measure would be more variable than the likelihood or amplitude measures. He also suggested that the likelihood and amplitude measures would not be autocorrelated in the way magnitude and likelihood measures would be. In NM studies using multiple measures, Smith (1968) and Leonard (1975) examined likelihood, amplitude, and latency measures, but they did not include the magnitude measure. Hence, it is unknown how strongly correlated the magnitude measure is with the criterion-based measures. Moreover, it is unknown whether any specific magnitude used as the criterion for defining a CR is either a convenient point on a continuous distribution or a point of discontinuity between low-level noise versus the larger movements that arise from CS–US pairings.

Setting aside uncertainties about the magnitude measure versus the criterion-based measures, there is also uncertainty regarding how consistently the criterion-based measures of eyeblink conditioning covary. Selected lesions can dissociate different measures (Garcia & Mauk, 1998; Medina, Garcia, & Mauk, 2001; Perrett, Ruiz, & Mauk, 1993). For example, Medina et al. recently showed that infusion of picrotoxin into the anterior interpositus nucleus of the cerebellum at different stages of CS–US pairings dramatically reduced the onset latency and peak latency of upper eyelid CRs. However, this increase in the speed of the CR was never accompanied by an increase in CR likelihood. The CR likelihood remained the same as it had been prior to infusion of the picrotoxin.

Other dissociations among criterion-based CR features have been achieved even without pharmacological interventions. The timing of the CR's peak in the NM preparation can be altered dramatically by changing the CS–US interval, but with little effect on likelihood or amplitude (Coleman & Gormezano, 1971). In experiments in which direct electrical stimulation of the CS pathway is used in place of a physical stimulus, timing of the upper eyelid can be shifted in either direction by modifying the stimulation frequency or intensity of the CS (Svensson, Ivarsson, & Hesslow, 1997). Conversely, the time course of the CR can be preserved even when its likelihood is dramatically altered. Kehoe, Horne, and Macrae (1995) completely extinguished an NM CR to one CS and then began training with a new CS using both a different sensory modality and a different CS–US interval. There was no immediate generalization across modalities, but the first CR acquired to the new CS showed a time course appropriate to the original CS–US interval rather than the new interval.

In order to determine to what extent the alternative measures of eyeblink conditioning are interchangeable, we examined their pattern of covariation in two distinct situations in which associative strength and its expression can be manipulated, namely, during acquisition training and during stimulus generalization. We chose to investigate stimulus generalization because, in theory, manipulating the similarity of the test stimuli to the original CS should affect behavioral expression without altering underlying associative strength. In addition, little is known about how measures of eyeblink conditioning covary during generalization testing. Previous studies of generalization report only changes in CR likelihood (Moore, 1972; Moore & Newman, 1966; Siegel, Hearst, George, & O'Neal, 1968; Solomon & Moore, 1975).

Method

Subjects

For Experiment 1, the subjects were 9 male, albino rabbits (*Oryctolagus cuniculus*), weighing 2.0–2.5 kg on arrival. Experiment 2 used 16 female, albino rabbits that weighed 1.5–2.0 kg on arrival. All rabbits had free access to food and water in their home cages. Treatment of the rabbits and surgical procedures complied with approved animal welfare protocols at each institution.

Surgical Procedures

In Experiment 1, subjects were prepared with a head bolt cemented to the skull. Rabbits were anesthetized with 5 mg/kg acepromazine. Anesthesia was maintained with halothane (1%–2% mixed in oxygen), and sterile procedures were used during the placement of the head bolt. After exposing the skull, four holes were drilled to accommodate screws that would be used to affix a bolt to the skull. The head bolts were secured to the skull with dental acrylic, and the skin was sutured. Finally, two stainless steel stimulating electrodes were chronically implanted in the

periorbital muscles rostral and caudal to the eye. Antibiotics, intravenous fluids, and analgesics were administered postsurgically as needed, and the rabbits were allowed approximately 1 week to recover.

In Experiment 2, hair posterior to the rabbit's right eye was shaved, and a 2-mm loop of silk (000 Dynex) was sutured into the right NM with a local anesthetic (proxymetacaine hydrochloride). The rabbits were returned to their home cages for 2 days.

Experimental Apparatus

The apparatus and recording procedure for both experiments were patterned after those of Gormezano (1966). The apparatus used in Experiment 1 is described by Garcia and Mauk (1998), and the apparatus used in Experiment 2 is described by Kehoe, Horne, Horne, and Macrae (1994).

In Experiment 1, movement of the unrestrained eyelid was recorded by measuring the reflectance of an infrared LED aimed at the eyelid. Voltage responses were linearly related to eyelid movement and were calibrated for each rabbit daily. Eyelid position was sampled at intervals of 1 ms and analyzed with custom-written software.

In Experiment 2, movements of the unrestrained NM were measured with a photoelectric transducer (Gormezano & Gibbs, 1988) connected by means of an L-shaped crank to the suture loop in the NM of the rabbit's right eye. The signal from the transducer was sampled every 5 ms and analyzed with custom-written software (Marshall-Goodell et al., 1982; Scandrett & Gormezano, 1980).

The CS was a 1000-Hz, 550-ms pure tone. In Experiment 1, the tone intensity was ~75 dB (SPL, re: 2×10^{-5} N/m^2) superimposed on an ambient noise level of ~30 dB. In Experiment 2, the intensity was measured at ~89 dB superimposed on an ambient noise level of ~81 dB, provided by a ventilation fan in each chamber. The test tones were all 550 ms and had frequencies of 1260, 1590, 2000, 2520, 3170, 4000, and 5040 Hz, respectively. The frequencies of the CS and test tones were selected so that adjacent stimuli were separated by one third of an octave.

The amplitude of the sinusoidal wave generated by the tone oscillators was held constant when frequency was manipulated. For the range of frequencies used in this study, the average intensity was constant when measured with C scale weightings. According to the instructions received with our sound-level meters, the C scale weightings are nearly constant relative to the 1000 Hz reference tone, for example, −0.8 dB at 4000 Hz. (Similarly, for the A scale weightings, which more closely match human hearing characteristics, the weightings are relatively constant across our frequency range, for example, +1.0 dB at 4000 Hz). For assessing perceived intensity (loudness), the small amount of available data indicate that rabbit's hearing characteristics are similar to those of humans (Martin, Lonsbury-Martin, & Kimm, 1980). Accordingly, equal-loudness contours for humans indicate that the loudness would have risen relative to the 1000-Hz reference tone as frequency was increased. For example, the increases would have been the equivalent of 8 dB at 4000 Hz and 5 dB at 5000 Hz (e.g., Kling & Riggs, 1972, p. 246). Thus, psychophysically, reductions in responding as a function of increases in tone frequency would reflect generalization decrements based on perceived frequency (pitch), but with some offsetting contributions from increases in perceived intensity (loudness).

In Experiment 1, the US was a 50-ms, 4-mA train of constant-current cathode pulses (200 Hz, 1-ms pulse width) delivered through two stainless steel wires implanted subdermally in the dorsolateral aspect of the left eye. In Experiment 2, the US was 50-ms, 4-mA, 50-Hz AC electric current delivered through two stainless steel Autoclip wound clips positioned ~10 mm apart, ~15 mm posterior to the dorsal canthus of the right eye.

Procedure

In both experiments, rabbits were first trained over 4 days to an asymptotic level of responding in a delay conditioning protocol with a 500-ms interval between the onset of the CS and the onset of the US. Each of the four training sessions consisted of 12 nine-trial blocks. Each block comprised eight paired presentations of the CS and US and one presentation of the CS only. The CS was presented for 550 ms during CS-alone trials and coterminated with the US during paired trials. Throughout the experiment, trials were separated by a 30-s intertrial interval.

Following initial acquisition training, stimulus generalization was assessed over eight daily test sessions. Each session contained 18 blocks of five trials each. Each block consisted of four CS–US pairings followed by one nonreinforced test trial using either one of the test tones, the trained CS, or a "blank" containing no tone, used to assess baseline blink rate. In each session, each of the eight tones and the blank were presented twice.

Response Definition

In both experiments, eyelid position was measured for 200 ms prior to each trial to provide a baseline for detecting movement. In both experiments, the *magnitude* was the maximal movement of the eyelid within an observation interval that was from 40 ms to 1800 ms after tone onset on test trials or during a corresponding period in the blank trials. For the criterion-based measures, a CR was defined as any movement that exceeded either 0.3 mm of upper eyelid movement (Experiment 1) or 0.5 mm of NM movement (Experiment 2) during the same observation interval as used for the magnitude measure. For movements that met the CR criterion, four measures were computed: *likelihood*, which was the percentage of trials on which a CR occurred; *amplitude*, which was the maximal movement of the eyelid provided it met the criterion for the CR; *onset latency*, which was the point at which the movement departed from the eyelid's resting baseline by 0.0625 mm; and *peak latency*, which was the point in time at which the movement reached its maximal extension. For all the measures, trials were excluded from the analysis if a movement greater than the criterion occurred during the baseline period.

Results

Statistical tests were conducted with planned contrasts in multivariate analyses of variance for repeated measures, using a Type I error rate of .01 (O'Brien & Kaiser, 1985). Although an error rate of .05 is conventional, the more conservative error rate was adopted to guard against an inflation of the overall error rate that otherwise would have occurred by using five different dependent variables.

Acquisition Training

Figure 1 summarizes the results for the five response measures during Days 1–4 of training in both experiments. Each panel of the figure shows the mean level of each measure aggregated across successive blocks of four CS-alone trials. There were two such blocks within each day of training. The bars around each mean represent 2 *SE*. In both experiments, amplitude, onset latency, and peak latency measures could not be calculated for each rabbit until it began to show movements that met the criterion for a CR. Accordingly, when only some rabbits contributed to a mean, the number that did contribute is listed next to the symbol for each mean. Recall that Experiment 1 contained 9 rabbits and Experiment 2 contained 16 rabbits.

Inspection of the figure indicates that, in both experiments, the magnitude, likelihood, and amplitude measures all showed increases across training, whereas onset latency and peak latency decreased. Linear trend tests were conducted to determine whether the observed changes across training were statistically significant.

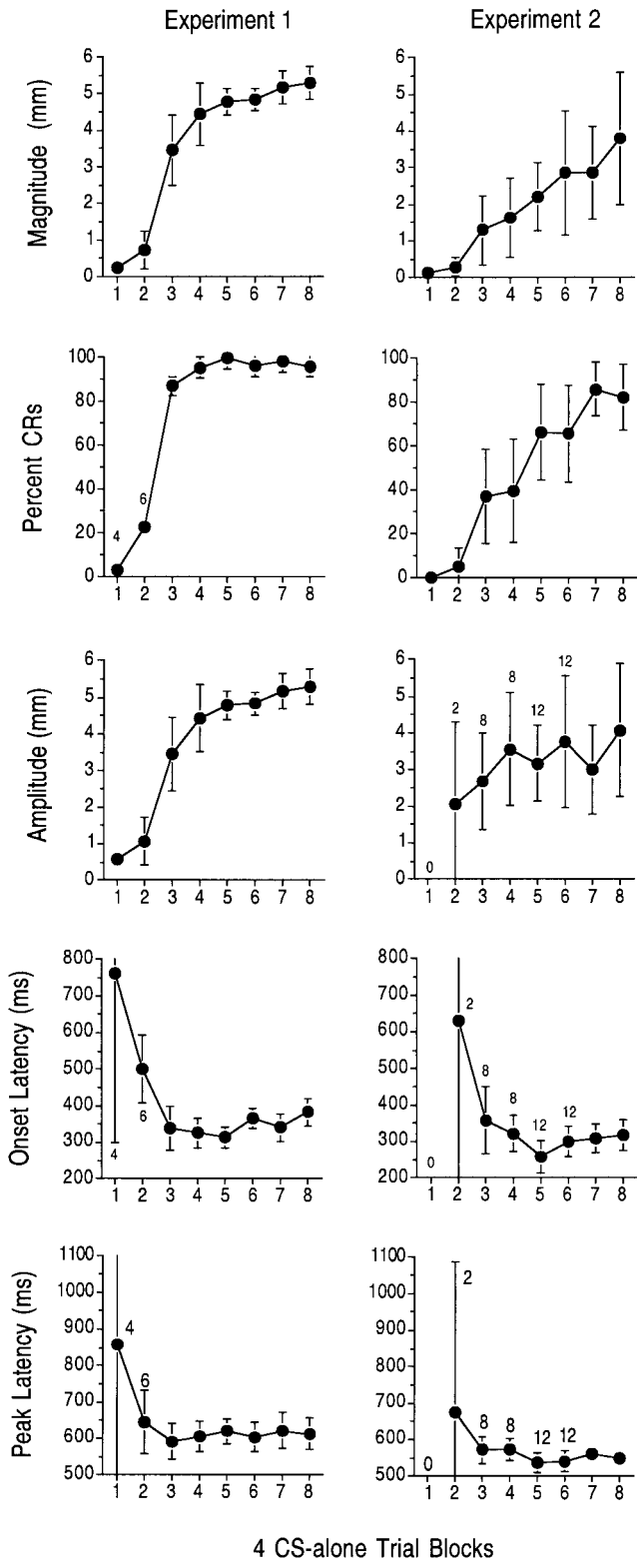

Figure 1. Mean level of each response measure for successive blocks of four conditioned stimulus (CS)-alone trials. When not all rabbits contributed to a mean, the number that did contribute is listed above the symbol for each mean ($n = 9$ for Experiment 1; $n = 16$ for Experiment 2). Bars represent 2 *SE*. CR = conditioned response.

For the magnitude measure, there were significant upward linear trends in both experiments, $F(1, 8) = 302.47$, $p < .01$; $F(1, 15) = 15.26$, $p < .01$. Likewise, for the likelihood measure, the upward trends were significant in both experiments, $F(1, 8) = 607.99$, $p < .01$; $F(1, 15) = 112.61$, $p < .01$. (The large F ratios of Experiment 1 reflect its low error variance, despite its having fewer rabbits than Experiment 2.)

For the amplitude, onset latency, and peak latency measures, the big changes tended to occur between the second and third blocks of test trials. However, relatively few rabbits contributed to the measures for the first and second blocks, thus making it difficult to conduct trend tests with so much missing data. Two approaches were tried to ascertain whether there were significant changes in those three measures.

First, we tested whether there were any significant changes after the first and second blocks. Hence, linear trend tests were conducted across Blocks 3–6. The data from all 9 rabbits in Experiment 1 could be used, but, only the data from 8 rabbits from Experiment 2 could be used. For Experiment 1, there were significant linear trends across Blocks 3–6 in amplitude, $F(1, 8) = 13.30$, $p < .01$, but not for onset latency, $F(1, 8) = 3.92$, $p > .05$, or peak latency ($F < 1$). For Experiment 2, there were no significant linear trends in amplitude, onset latency, or peak latency, largest $F(1, 7) = 3.31$, $p > .05$.

Second, we tested whether there were changes in the early CRs displayed by each rabbit whenever they first appeared. To do this test, the data were realigned: The first block in which CRs were displayed by each rabbit was designated as Block 1, and so forth for the remaining blocks. For Experiment 1, there were six blocks of data for all 9 rabbits. In agreement with the first analysis, there was a significant upward linear trend in amplitude, from a mean of 1.4 mm in the first block to 5.0 mm in the sixth block, $F(1, 8) = 58.37$, $p < .01$. Onset latencies showed decreases from 578 ms to 353 ms, and peak latencies showed decreases from 725 ms to 597 ms. However, neither of these decreases were statistically significant, largest $F(1, 8) = 3.72$, $p > .05$. For Experiment 2, there were four blocks of data for 12 rabbits. Analyses of these data failed to reveal significant linear trends across blocks. Specifically, amplitude increased nonsignificantly from 2.2 to 4.0 mm, $F(1, 11) = 3.30$, $p > .05$; onset latency decreased nonsignificantly from 342 ms to 290 ms, $F(1, 11) = 4.79$, $p > .05$; and peak latency barely changed, from 549 ms to 547 ms ($F < 1$).

Generalization Testing

Figure 2 summarizes the results for the five response measures in both experiments during generalization testing. Each panel of the figure shows the mean level of responding for a measure as a function of tone frequency. The bars around each mean represent 2 *SE*. For Experiment 1, the means and *SE*s for the amplitude and latency measures at the three highest test frequencies were based on *n*s of 7, 7, and 6, respectively. These reduced sample sizes are listed next to the symbol for the means in the figure. At these higher frequencies, the other rabbits did not show CRs and therefore could not be included in the computations for those measures. For Experiment 2, CRs appeared at all frequencies in all rabbits, and hence, all measures could be computed for all tone frequencies.

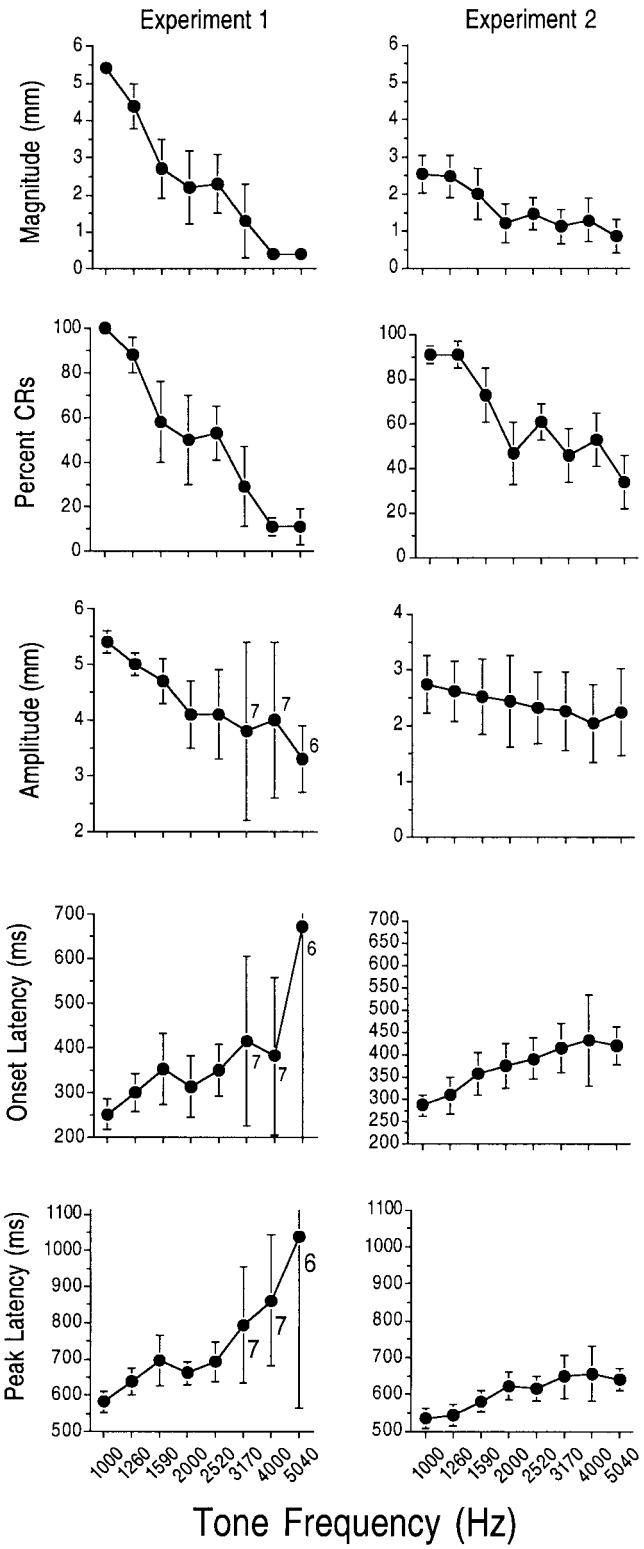

Figure 2. Mean level of each response measure in generalization testing as a function of tone frequency. The frequency used on conditioned stimulus–unconditioned stimulus trials was 1000 Hz. When not all rabbits contributed to a mean, the number that did contribute is listed adjacent to the symbol for each mean ($n = 9$ for Experiment 1; $n = 16$ for Experiment 2). Bars represent 2 *SE*. CR = conditioned response.

In both experiments, all measures showed a generalization gradient. That is, the magnitude, likelihood, and amplitude all showed declines as the tone frequency increased away from the training value of 1000 Hz. Conversely, the onset and peak latencies increased as tone frequency increased.

For Experiment 1, there was a significant linear trend for each measure except onset latency. For each measure, a linear trend test was conducted on the five lower frequencies, which allowed the data for all 9 rabbits to be included. The *F* ratios ($df = 1, 8$) for magnitude, likelihood, amplitude, onset latency, and peak latency were 70.33, 49.06, 14.68, 7.67, and 14.91, respectively ($ps < .01$). A quadratic trend test for curvature in the generalization gradient was also conducted for each measure. These tests yielded *F* ratios ($df = 1, 8$) of 12.51, 5.88, 1.01, 3.87, and 6.76, respectively. In these tests, only the *F* ratio for the magnitude measure achieved the declared level of statistical significance ($p < .01$).

A second set of tests was conducted with all eight frequencies for just the magnitude and likelihood measures, for which scores of zero could be properly included. For both measures, there was a significant linear trend, smaller $F(1, 8) = 878.59$, $p < .01$, but no quadratic trend, larger $F(1, 8) = 3.24$, $p > .10$.

For Experiment 2, the generalization gradients were relatively shallow. Nevertheless, the statistical analysis revealed that, with the exception of the amplitude measure, there were significant linear trends in all measures across frequencies. The *F* ratios ($df = 1, 15$) for magnitude, likelihood, amplitude, onset latency, and peak latency were, respectively, 42.64, 58.68, 5.16, 16.99, and 21.90 ($ps < .01$). There was some evidence of curvature, but none of the tests for quadratic trend attained the declared level of significance. Specifically, the *F* ratios ($df = 1, 15$) for quadratic trend were 7.46, 5.02, 0.64, 3.65, and 4.30, respectively.

Covariation Among Measures

All measures showed some changes as a function of training trials and tone frequency, but some more so than others. If the *F* ratios for the linear trends can be construed as signal/noise ratios, the magnitude and likelihood measures showed the greatest sensitivity in both stages in both experiments. On this basis, it might be tempting to infer that they are interchangeable. Conversely, amplitude, onset latency, and peak latency appear to be relatively insensitive and not easily interchangeable with magnitude and likelihood. To obtain a more detailed picture of the degree of covariation and hence their potential interchangeability, two sets of correlational analyses were conducted for all 10 pairs of the measures.

The first set of correlational analyses were conducted among the group means shown in Figures 1 and 2. As pictorial examples, Figure 3 shows a set of scattergrams for (a) likelihood versus magnitude (filled circle) and (b) likelihood versus amplitude (open circle). The upper panels show the scattergrams for acquisition training in the two experiments, and the lower panels show the scattergrams for generalization testing. In the lower panels, each symbol represents the pairing of means for each of the eight tone frequencies in generalization testing. Each panel also includes the plot of two linear regression lines, one for each pairing of measures.

Table 1 summarizes the correlational analyses for all 10 pairs of response measures. There is a separate panel for acquisition train-

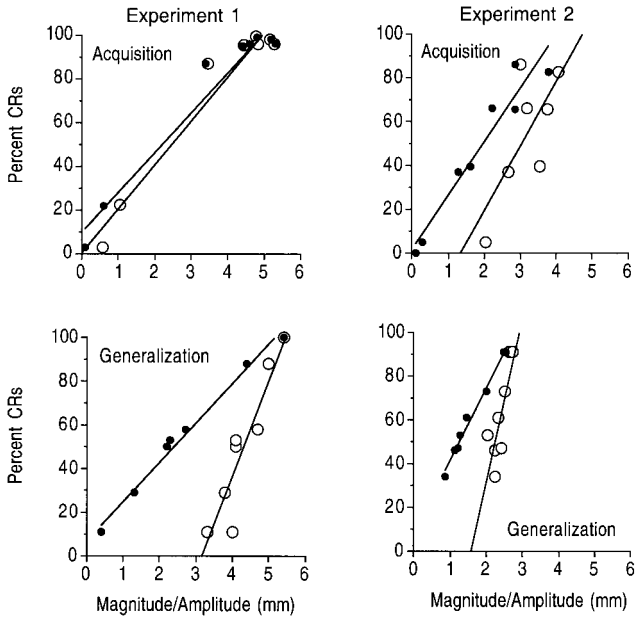

Figure 3. Scattergrams for likelihood versus magnitude (filled circles) and likelihood versus amplitude (open circles) for each block of conditioned stimulus-alone test trials in acquisition training (upper panels) and for each tone frequency in generalization testing (lower panels). Each panel also includes the plot of two linear regression lines, one for each pairing of measures. CR = conditioned response.

ing in Experiment 1, acquisition training in Experiment 2, generalization testing in Experiment 1, and generalization testing in Experiment 2. Each cell shows two numbers. The left-hand number indicates the proportion of the total variance explained by linear regression, and the right-hand number indicates the proportion of the total variance explained by curvilinear regression based on a second-order polynomial function. (Other curvilinear fits using logarithmic and exponential functions were also examined. They explained about the same amount of variance as the second-order polynomial function.) The fit for a curvilinear relationship was included because, in some cases, one measure appeared to reach either its asymptote or natural limit while the other measure was still changing. Where the amount of explained variance exceeds 75%, the number is printed in bold type. (An explained variance of 75% corresponds to a Pearson correlation coefficient of .870, which is required for statistical significance if a stringent criterion of .005 is adopted to take into account the 10 coefficients being calculated.)

Inspection of Figure 3 and Table 1 reveals that the magnitude and likelihood measures had a near-perfect linear relationship in both stages of both experiments. The relationship of amplitude to the magnitude and likelihood measures, however, was not as strong, particularly in Experiment 2. For example, in acquisition training, a linear relationship between CR likelihood and CR amplitude explained only 51% of the covariance. Similarly, the linear relationships of the speed measures—onset latency and peak latency—to the other measures varied across the experiments and their stages. The curvilinear relationships appeared to be stronger, which means that there was a portion of each relationship in which one variable—usually the speed measure—was static while the other variable was still changing in value. Finally, onset latency and peak latency had a very strong relationship, in which at least 89% of the variance was explained by a linear relationship across both stages of both experiments.

Although there were solid linear relationships among the group means for some pairs of measures, there is no guarantee that these relationships were equally strong for individual rabbits. To determine the strength of the relationships within individual rabbits, analyses were conducted for the 10 pairs of measures for individual rabbits. Just as was done with the group means, the linear relationship between each pair of measures was computed by using the eight blocks of test trials in acquisition and, similarly, the eight tone frequencies in generalization testing.

Table 2 summarizes the subject-by-subject correlational analyses for the 10 pairs of response measures. The cell for each pair of measures shows (a) the explained variance averaged across rabbits, (b) the lowest explained variance among the rabbits, and (c) the highest explained variance among rabbits. For the magnitude and likelihood measures, all trial blocks for all rabbits could be used. However, for the amplitude, onset latency, and peak latency measures, the number of points was reduced, because those measures

Table 1
Percentage of Covariance Among Group Means Explained by Linear and Second-Order Polynomial Fitted Lines

Measure	Magnitude	Amplitude	Onset latency	Peak latency
		Experiment 1: Acquisition		
Likelihood	**.95/.99**	**.95/.99**	**.86/.92**	.67/**.75**
Magnitude	—	**.99/.99**	.73/**.89**	.54/.74
Amplitude		—	.72/**.91**	.54/**.76**
Onset latency			—	**.89/.89**
Peak latency				—
		Experiment 2: Acquisition		
Likelihood	**.93/.96**	.51/.60	.65/.66	.69/.70
Magnitude	—	.72/.73	.53/**.95**	.62/**.93**
Amplitude		—	.56/**.96**	.62/**.90**
Onset latency			—	**.97/.97**
Peak latency				—
		Experiment 1: Generalization		
Likelihood	**.99/.99**	**.86/.87**	.56/**.84**	**.78/.94**
Magnitude	—	**.88/.88**	.53/.60	.72/**.86**
Amplitude		—	.66/**.88**	.72/**.84**
Onset latency			—	**.90/.92**
Peak latency				—
		Experiment 2: Generalization		
Likelihood	**.99/.99**	.65/**.80**	**.82/.82**	**.86/.88**
Magnitude	—	.72/**.86**	**.87/.87**	**.91/.94**
Amplitude		—	**.94/.98**	**.90/.94**
Onset latency			—	**.97/.97**
Peak latency				—

Note. Left-hand number is the percentage of covariance explained by linear regression. Right-hand number is the percentage of covariance explained by second-order polynomial regression. Numbers in bold type are statistically significant at a Bonferroni-corrected level of .005.

Table 2
Averaged Percentage of Covariance in Individual Rabbits Explained by Linear Regression

Measure	Magnitude	Amplitude	Onset latency	Peak latency
		Experiment 1: Acquisition		
Likelihood	**.89 (.72–.96)**	.66 (.01–.97)	.35 (.08–.87)	.61 (.04–.96)
Magnitude	—	**.98 (.91–1.00)**	.60 (.21–.80)	.37 (.09–.63)
Amplitude		—	.61 (.32–.83)	.36 (.07–.61)
Onset latency			—	.44 (.02–.91)
Peak latency				—
		Experiment 2: Acquisition		
Likelihood	**.80 (.51–.99)**	.48 (.02–1.00)	.49 (.00–1.00)	.39 (.00–1.00)
Magnitude	—	**.94 (.72–1.00)**	.51 (.00–1.00)	.54 (.01–1.00)
Amplitude		—	.50 (.00–1.00)	.53 (.06–1.00)
Onset latency			—	.48 (.01–1.00)
Peak latency				—
		Experiment 1: Generalization		
Likelihood	**.97 (.92–.99)**	.39 (.00–.75)	.38 (.00–.77)	.47 (.01–.78)
Magnitude	—	.49 (.02–.81)	.49 (.00–.77)	.37 (.00–.77)
Amplitude		—	.32 (.00–.75)	.36 (.01–.91)
Onset latency			—	.65 (.32–1.00)
Peak latency				—
		Experiment 2: Generalization		
Likelihood	**.86 (.62–1.00)**	.44 (.02–.89)	.41 (.00–.85)	.44 (.03–.83)
Magnitude	—	.57 (.00–.96)	.48 (.06–.88)	.47 (.04–.86)
Amplitude		—	.44 (.03–.83)	.40 (.03–.90)
Onset latency			—	**.80 (.15–.97)**
Peak latency				—

Note. Left-hand number is the mean percentage of covariance explained by linear regression. Numbers in parentheses show the range of explained variance among rabbits. Numbers in bold type are statistically significant at a Bonferroni-corrected level of .005.

were restricted to trials containing movements that equaled or exceeded the criterion for a CR.

This latter restriction had two impacts on the correlational analyses. First, in blocks of trials in which no eyelid movements met the criterion, the CR was undefined, and hence that point had to be excluded from computations for the criterion-based measures. Second, in trial blocks in which at least one movement met the CR criterion, points for amplitude, onset latency, and peak latency could be computed. However, the mean amplitude, which was criterion based, would necessarily be larger than the mean magnitude, which included subcriterion movements. Only in trial blocks in which all movements met the CR criterion were the mean amplitude and mean magnitude identical.

Inspection of Table 2 reveals that only magnitude and likelihood were well correlated on a subject-by-subject basis in both stages of both experiments. The range reveals that, in most cases, the mean amount of explained variance was reasonably representative of the individual rabbits. In acquisition, magnitude and amplitude were strongly correlated. However, this correlation diminished severely in generalization testing. Moreover, in all stages, the average correlation between likelihood and amplitude was modest and highly variable across rabbits. Similarly, among the other pairings of variables, there were moderate correlations, but the average amount of variance explained was low and often highly variable across rabbits.

Distribution of Eyelid Movements

The pattern—or, more correctly, the lack of a consistent pattern—in the relationship among the magnitude, likelihood, and amplitude measures seems to defy the intuitive hypothesis that the size of the eyelid movement varies continuously. An alternative hypothesis is that eyelid movements vary discontinuously between a negligible baseline noise level and a larger magnitude that is relatively fixed for any given rabbit. Hence, acquisition would be characterized by a growth in the proportion of trials containing the larger movements superimposed on the baseline level. According to this hypothesis, both the magnitude and likelihood measures across a set of trials would reflect the averaging of the small and large eyelid movements. Accordingly, as has been seen, these two measures would be well correlated. Conversely, the amplitude measure, which includes only the large, supercriterion movements, would necessarily remain constant at all times. Hence, it would be uncorrelated with the magnitude and likelihood measures.

To test these alternative hypotheses, we examined the distribution of magnitudes in both stages of both experiments. Figure 4 shows the mean proportion of movements that fell into each of eight bins. The bin for the smallest movements included those measured between 0 and 0.125 mm. This bin encompassed subcriterion movements that could not be readily distinguished from noise in the recording system. The next two bins (0.126–0.250

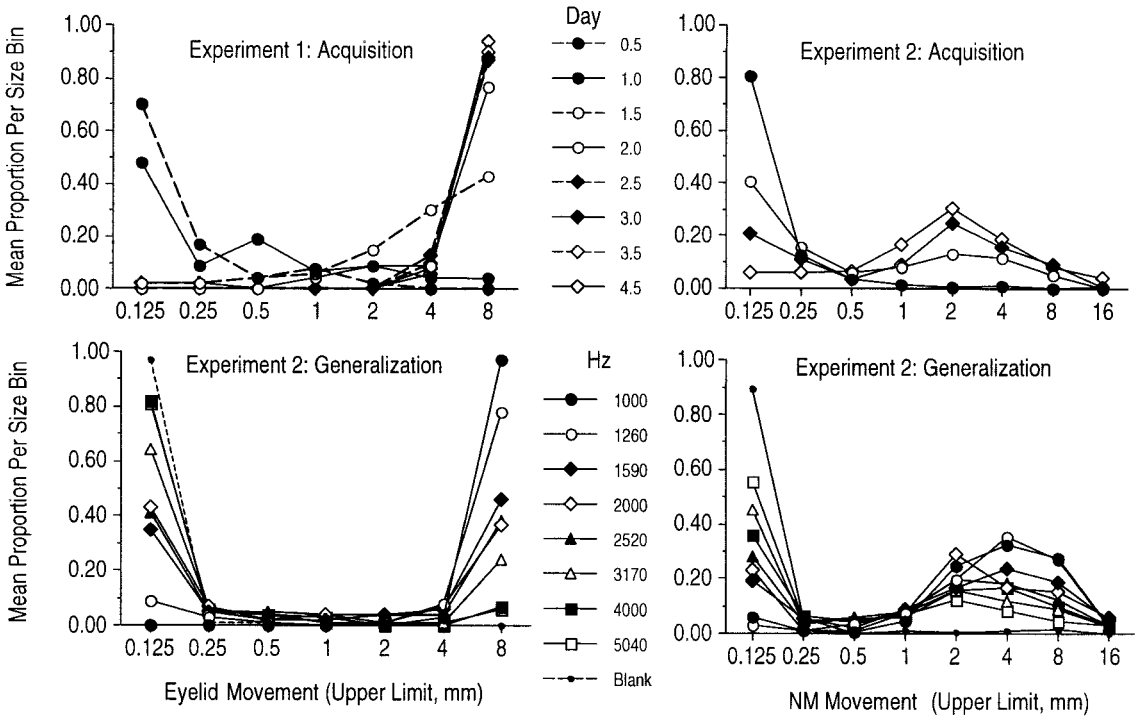

Figure 4. Mean proportion of response magnitudes that fell into each of eight size bins. Each line in the upper panels depicts the distribution on either half-day blocks (Experiment 1) or full-day blocks (Experiment 2). Each line in the lower panels depicts the distribution for each tone frequency plus blank trials, which contained no tone. NM = nictitating membrane.

mm, 0.251–0.500 mm), included movements that could be distinguished from background noise but were still less than or equal to the criterion for a CR. All other bins included movements that were counted as CRs.

The top two panels of Figure 4 shows the mean proportion of magnitudes that fell into each of the bins on successive days of initial training. For Experiment 1, in which acquisition was rapid, each day was divided into two halves. For Experiment 2, in which acquisition was slower, all the trials in each day were grouped together. The bottom two panels of Figure 4 show the mean proportion of magnitudes that fell into each bin for each tone frequency. In addition, a corresponding distribution is plotted for the baseline magnitude measurements observed on the blank trials, which contained no tone.

Inspection of the distributions during acquisition reveals that, across days, the distributions shifted rapidly. On the 1st day, the vast majority of measurements (> 70%) were concentrated in the baseline range of 0–0.125 mm. By the end of the 2nd day and certainly by the end of the 3rd day, the movements were concentrated in the higher ranges, specifically, 4.01 mm and greater in Experiment 1 and 1.01–2.00 mm in Experiment 2.

For present purposes, however, the key question is whether the eyelid movements passed through the intermediate sizes. In acquisition, there is some evidence that the movements passed, albeit rapidly, through intermediate sizes. Examination of the data from individual rabbits revealed that most showed at least one block of trials in which their mean magnitude was greater than 0.125 mm but less than 1 mm. Specifically, 7 of 9 rabbits in Experiment 1 and 14 of 16 rabbits in Experiment 2 showed movements within the 0.126–1.00 mm range.

Trend tests were conducted to determine whether there were any statistically reliable changes in each bin during acquisition training. Across days of acquisition in Experiment 1, these tests revealed that there was a significant downward linear trend in the proportion of magnitude measurements in the range of 0–0.125 mm, $F(1, 8) = 42.39$, $p < .01$. Conversely, there was a significant upward linear trend in the proportion of movements in the range of 4.01–8 mm, $F(1, 8) = 148.21$, $p < .01$. Most important, for testing whether the movements passed through intermediate magnitudes during acquisition, there was a significant quadratic trend for two intermediate ranges, specifically, 0.501–1.00 mm, $F(1, 8) = 5.63$, $p < .05$, and 2.01–4.00 mm, $F(1, 8) = 9.47$, $p < .05$. That is, the proportion of movements within those two ranges increased and then decreased during acquisition. However, no significant trends were observed for subcriterial movements in the 0.126–0.250 mm or 0.251–0.500 mm bins.

Unfortunately, for Experiment 2, in which the distributions were less well defined, it was not possible to ascertain whether or not the distributions passed through intermediate values. Statistically, there was a significant downward linear trend across days in the proportion of movements in the range of 0–0.125 mm, $F(1, 15) = 127.26$, $p < .01$. There were also significant upward trends in the proportion of movements in the ranges 0.501–1.00 mm, 1.01–2.00 mm, and 2.01–4.00 mm, largest $F(1, 15) = 12.66$, $p < .01$. There was no evidence of a significant quadratic trend in the proportion of movements with intermediate sizes.

Although there was some evidence that the magnitude of eyelid movements was distributed continuously in acquisition, the results of generalization testing revealed, once acquisition had occurred, the movements were distributed bimodally. For Experiment 1, the results are striking. The movements were either negligible in size (0–0.125 mm) or substantial (> 4.01 mm). Trend tests revealed that, across tone frequencies, there were significant linear trends in the opposite direction for the 0–0.125-mm and 4.01–8.00-mm bins, smaller $F(1, 8) = 535.38$, $p < .01$. With regard to the intermediate sizes, there were few movements, and there were no significant trends across tone frequencies.

For Experiment 2, the results are less striking but still relatively clear cut. For the training frequency of 1000 Hz, the distribution was centered in the range of 2.01–4.00 mm. As the test frequency was increased away from the 1000-Hz value, the distribution became progressively bimodal. First, the proportion of measurements in the range of 0–0.125 mm increased. Second, as the proportion of magnitudes greater than the CR criterion decreased, the second mode shifted to the range of 1.01–2.00 mm. Despite this shift, there was no increase in the proportion of movements in the bins just below the CR criterion. The proportion of movements in the range of 0.251–0.500 mm hovered between 1% and 6%. Statistically, as the test frequency increased away from the training value of 1000 Hz, there was a significant upward linear trend in the range of 0–0.125 mm, $F(1, 15) = 67.72$, $p < .01$, while there were significant downward linear trends in the ranges of 2.01–4.00 mm and 4.01–8.00 mm, largest $F(1, 15) = 16.70$, $p < .01$. For the other bins, any apparent differences in generalization testing in Experiment 2 failed to reach significance.

Discussion

When aggregated in group means, all the measures changed roughly in parallel in both acquisition and generalization testing. Not all measures, however, changed to the same degree. In both experiments, magnitude and likelihood changed in a large and consistent fashion. Amplitude showed similar changes, but these were dampened by the exclusion of subcriterion movements. In acquisition, onset latency and peak latency quickly settled at stable values, but these measures did show moderate changes across tone frequencies in generalization testing.

The findings in acquisition confirm what has previously been observed in rabbit eyeblink conditioning (e.g., Smith, 1968). The failure to see significant decreases in onset latency might seem to contradict the usual description of the emergence of the CR. When examined on a trial-by-trial basis, early CRs are usually initiated just before the US, and subsequent CRs move away from the US and toward the middle of the CS–US interval (Gormezano et al., 1983). In the present experiments, CRs were measured on only a few CS-alone test trials distributed among the CS–US trials. This periodic nature of the sampling plus the apparent rapid decrease in onset latency made it difficult to detect any early changes.

With regard to the group means, the new findings from these experiments concern generalization. The reductions in CR likelihood across test frequencies agree with previous findings in the NM preparation (Moore, 1972). However, the reductions in the magnitude and amplitude are new, as are the increases in onset latency and peak latency. For example, the increase in tone frequency from the training value of 1000 Hz to a test value of 2000 Hz produced increases of approximately 30% in onset latency and 15% in peak latency. The increase in onset latency as a measure of the reduced expression of associative strength is perhaps not surprising. However, the increase in peak latency was not expected, given the usual depiction of peak latency as being strongly tied to the CS–US interval.

The pattern of data from generalization testing is consistent with a spreading activation model of generalization and the timing of eyelid movements (Desmond, 1990). This model assumes that CS onset initiates activity in an array of hypothetical stimulus elements arranged in successive, highly interconnected layers. At first, only a few elements in the first layer are activated. These elements in turn activate more elements in the second layer, these new elements activate still more elements in the next layer, and so forth. However, this progressive activation does not continue indefinitely. Each element is activated for only a brief period, and gradually all activation fades away. Thus, a wave of activation ripples over the layers of elements, widening and then contracting. Elements in whichever layers are active at the time of US presentation are assumed to acquire associative strength. Hence, the initiation of a CR-sized eyelid movement would occur when just enough elements with associative strength were activated during a CS presentation. The CR peak occurs when the largest number of elements having associative strength is activated.

According to the spreading activation model, generalization occurs because the array activated by a test tone shares elements with the array of conditioned elements. The proportion of shared elements would determine the magnitude of the eyelid movement to a test tone. However, the proportion of shared elements would vary across time. The early elements activated by the test tone would not overlap those activated by the training tone. The array of elements activated by the test tone would only overlap the array of conditioned elements in their large middle portions. Consequently, as seen in the present results, the initiation and peak of the eyelid movement would tend to be delayed until a sufficient number of elements with associative strength were activated.

Although the group means tended to show parallel changes, their degree of covariation itself was variable. On a group basis, the magnitude and likelihood measures were closely coupled across experiments and stages. In all cases, over 90% of the variance could be explained by a linear relationship. The only other pair of variables to be equally well coupled were onset latency and peak latency. Other pairs of variables showed less consistent linear covariance. With the exception of the relationship between magnitude and likelihood, the linear covariation among measures substantially decreased and, in some cases, disappeared when the data were considered on a subject-by-subject basis.

This pattern of findings runs contrary to the common wisdom that features of conditioned responding are readily interchangeable measures of the underlying associative strength of the CS, particularly on a subject-by-subject basis. The good news, however, is that, in eyeblink conditioning, the magnitude and likelihood measures do appear to be interchangeable, certainly on a group basis. Even among individual rabbits, the lowest correlation ($r = .71$) for an individual rabbit explained 51% of the variance between magnitude and likelihood.

The strong correlation between magnitude and likelihood measures can be understood by considering the approximately bimodal distributions of magnitudes. These distributions indicate that the

acquisition and expression of CRs is discontinuous. That is, when a CR does occur, it occurs with a relatively fixed size. Hence, across a block of several trials, both the likelihood of a CR and the mean magnitude of all movements would reflect the mixture of trials containing background noise (< 0.125 mm) and trials containing CR-sized movements (1–4 mm). By the same token, measures that depend on there being a CR-sized movement—namely, amplitude, onset latency, and peak latency—would vary less and be less well correlated with the magnitude and likelihood measures.

The present study was prompted by methodological and empirical questions as to whether different measurements of eyelid responding are interchangeable. The answer is a mixture of good news and bad news. The good news is that two of the most widely reported measures—magnitude and likelihood—appear to be interchangeable. The bad news is they are not interchangeable with amplitude, onset latency, or peak latency. Given the lack of correlation between the magnitude and amplitude, there is a particular risk in confusing these two measures. In everyday usage, the words *magnitude* and *amplitude* are largely synonymous. They, however, have been distinguished here as technical labels in line with Gormezano's (1966) usage. More important than the specific label, reports of data simply need to be clear as to whether measurements included the full range of movements, including trials with zero movement (magnitude), versus a truncated range based on a minimal criterion for movement (amplitude). Confusing the two could produce artifactually divergent outcomes.

From a theoretical perspective, the present results challenge the assumption that different features of conditioned responding reflect the expression of associative strength in a uniform fashion. However, by specifying the rules of CR elicitation, modern theories of conditioning can explain the different features of conditioned responding. In particular, a spreading activation model (Desmond, 1990) and other "real-time" models (e.g., Buhusi & Schmajuk, 1999; Buonomano & Mauk, 1994; Grossberg & Schmajuk, 1989; Moore & Choi, 1997; Sutton & Barto, 1990) are well suited to explaining the time course of the CR. However, these models have not typically made precise assumptions about the expression of the underlying associative strength in overt eyelid movement.

In broad terms, there are two basic hypotheses about associative expression. First, one could use a noisy threshold as a horizon; overt eyelid movement would occur as the total associative strength gradually and more consistently rose above this threshold. This horizon-type mechanism would predict incorrectly that the amplitude of overt movements should show a continuous distribution rather than a discontinuous distribution. Moreover, under this hypothesis, amplitude should have been more strongly correlated with magnitude and likelihood than what was actually observed. However, it is conceivable that the rise of associative strength above the horizon was so fast that test trials were too infrequent to detect the "dawn" period in which intermediate amplitudes would have been observed. Second, as an alternative to the horizon-type mechanism, a noisy threshold could be depicted as a trigger for firing the overt eyelid CR (Baker & Frey, 1979; Frey & Sears, 1978). Under this hypothesis, the timing and amplitude of the CR would be established during early CS–US pairings prior to its overt expression. Whenever associative strength exceeded the threshold, the full-blown CR would appear. This trigger-type hypothesis would predict that, as seen in the present results, the overt CR would tend to have a fixed time course with a fixed amplitude.

There is certainly enough complexity in the pathways that govern eyelid conditioning to support a wide variety of possible mechanisms for associative acquisition and expression. For example, in cerebellar pathways alone, the CR output is governed by at least two routes. Specifically, the interpositus nucleus (*nuc*) receives CS input through two routes. One route entails connections between mossy fibers (*mf*) and the interpositus nucleus (*mf–nuc*). The second route entails indirect connections from the mossy fibers through the cerebellar cortex, specifically, to granule cells (*gr*) which connect to Purkinje cells (*Pkj*) that inhibit the interpositus nucleus (*mf–gr–Pkj–nuc*) (Clark & Lavond, 1993; Hesslow, Svensson, & Ivarsson, 1999; Mauk & Donegan, 1997; Medina et al., 2001). In addition, the CR output from the red nucleus, which receives inputs from the interpositus nucleus, appears to feed back to the cortical granule cells (Rosenfield & Moore, 1995). Attempts to locate the points of plasticity affected by CS–US pairings have provided evidence for plasticity in both the *mf–nuc* synapses (e.g., Lavond, Kanzawa, Ivkovich, & Clark, 1994; Lavond, Kim, & Thompson, 1993; Thompson et al., 1987) and the *gr–Pkj* synapses (e.g., Gruart & Yeo, 1995; Hesslow et al., 1999). As a first hypothesis related to the present findings, the all-or-none elicitation of the CR may be governed by the *mf–nuc* connections, whereas the amplitude and timing of the CR may be controlled by the pathways through the *gr–Pkj* synapses in the cerebellar cortex (Mauk & Donegan, 1997; Medina et al., 2001).

As a final note, extreme caution should be exercised in reaching any general conclusions about differences between the upper eyelid response used in Experiment 1 versus the NM response in Experiment 2 with regard to either their mean levels or variances. Although we attempted to run the two experiments under identical conditions, there were numerous differences between the two laboratories as detailed in the methods section, for example, response transducers, CS intensity, background noise intensity, and US parameters. In addition, the rabbits came from different sources. Even when laboratory conditions are identical, animals from different suppliers can yield significantly different rates of acquisition (Kehoe, Horne, Kingham, Martin, & Roach, 1995). In the face of so many possible sources of variation, the similarities in the results of the present two experiments are all the more impressive.

References

Baker, J. S., & Frey, P. W. (1979). A cusp catastrophe: Hysteresis, bimodality, and inaccessibility in rabbit eyelid conditioning. *Learning & Motivation, 10,* 520–535.

Brown, J. S., Kalish, H. I., & Farber, I. E. (1951). Conditioned fear as revealed by magnitude of startle response to an auditory stimulus. *Journal of Experimental Psychology, 41,* 317–328.

Buhusi, C. V., & Schmajuk, N. A. (1999). Timing in simple conditioning and occasion setting: A neural network approach. *Behavioural Processes, 45,* 33–57.

Buonomano, D. V., & Mauk, M. D. (1994). Neural network model of the cerebellum: Temporal discrimination and the timing of motor responses. *Neural Computation, 6,* 38–55.

Clark, R. E., & Lavond, D. G. (1993). Reversible lesions of the red nucleus during acquisition and retention of a classically conditioned behavior in rabbits. *Behavioral Neuroscience, 107,* 264–270.

Coleman, S. R., & Gormezano, I. (1971). Classical conditioning of the rabbit's (*Oryctolagus cuniculus*) nictitating membrane response under symmetrical CS–US interval shifts. *Journal of Comparative and Physiological Psychology, 77,* 447–455.

Desmond, J. E. (1990). Temporal adaptive responses in neural models: The stimulus trace. In M. Gabriel & J. W. Moore (Eds.), *Learning and computational neuroscience* (pp. 421–456). Cambridge, MA: MIT Press.

Frey, P. W., & Sears, R. J. (1978). Model of conditioning incorporating the Rescorla-Wagner associative axiom, a dynamic attention process, and a catastrophe rule. *Psychological Review, 85,* 321–340.

Garcia, K. S., & Mauk, M. D. (1998). Pharmacological analysis of cerebellar contributions to the timing and expression of conditioned eyelid responses. *Neuropharmacology, 37,* 471–480.

Gormezano, I. (1966). Classical conditioning. In J. B. Sidowski (Ed.), *Experimental methods and instrumentation in psychology* (pp. 385–420). New York: McGraw-Hill.

Gormezano, I., & Gibbs, C. M. (1988). Transduction of the rabbit's nictitating membrane response. *Behavior Research Methods, Instrumentation, and Computers, 20,* 18–21.

Gormezano, I., Kehoe, E. J., & Marshall, B. S. (1983). Twenty years of classical conditioning research with the rabbit. In J. M. Sprague & A. N. Epstein (Eds.), *Progress in psychobiology and physiological psychology* (Vol. 10, pp. 197–275). New York: Academic Press.

Grossberg, S., & Schmajuk, N. A. (1989). Neural dynamics of adaptive timing and temporal discrimination during associative learning. *Neural Networks, 2,* 79–102.

Gruart, A., & Yeo, C. H. (1995). Cerebellar cortex and eyeblink conditioning: Bilateral regulation of conditioned responses. *Experimental Brain Research, 104,* 431–448.

Hesslow, G., Svensson, P., & Ivarsson, M. (1999). Learned movements elicited by direct stimulation of cerebellar mossy fiber afferents. *Neuron, 24,* 179–185.

Kehoe, E. J., Horne, A. J., Horne, P. S., & Macrae, M. (1994). Summation and configuration between and within sensory modalities in classical conditioning of the rabbit. *Animal Learning & Behavior, 22,* 19–26.

Kehoe, E. J., Horne, A. J., Kingham, J., Martin, T., & Roach, W. (1995). Acquisition of a conditioned reflex in New Zealand White rabbits from three sources. *Laboratory Animals, 29,* 394–399.

Kehoe, E. J., Horne, A. J., & Macrae, M. (1995). Learning to learn: Real-time features and a connectionist model. *Adaptive Behavior, 3,* 235–271.

Kehoe, E. J., & Macrae, M. (2002). Fundamental behavioral methods and findings in classical conditioning. In J. W. Moore (Ed.), *A neuroscientist's guide to classical conditioning* (pp. 171–231). New York: Springer.

Kehoe, E. J., Marshall-Goodell, B., & Gormezano, I. (1987). Differential conditioning of the rabbit's nictitating membrane response to serial compound stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 13,* 17–30.

Kling, J. W., & Riggs, L. A. (1972). *Woodworth & Schlosberg's experimental psychology* (3rd ed., Vol. 1). New York: Holt, Rinehart and Winston.

Lavond, D. G., Kanzawa, S. A., Ivkovich, D., & Clark, R. E. (1994). Transfer of learning but not memory after unilateral cerebellar lesions in rabbits. *Behavioral Neuroscience, 108,* 284–293.

Lavond, D. G., Kim, J. J., & Thompson, R. F. (1993). Mammalian brain substrates of aversive classical conditioning. *Annual Review of Psychology, 44,* 317–342.

Leonard, D. W. (1975). Partial reinforcement effects in classical aversive conditioning in rabbits and human beings. *Journal of Comparative and Physiological Psychology, 88,* 596–608.

Marshall-Goodell, B., Schreurs, B. G., & Gormezano, I. (1982). Ruler vs. the Apple II/FIRST system analysis of analog signals in classical conditioning. *Behavior Research, Methods and Instrumentation, 14,* 519–525.

Martin, G. K., Lonsbury-Martin, B. L., & Kimm, J. (1980). A rabbit preparation for neuro-behavioral auditory research. *Hearing Research, 2,* 65–78.

Mauk, M. D., & Donegan, N. H. (1997). A model of Pavlovian eyelid conditioning based on the synaptic organization of the cerebellum. *Learning and Memory, 3,* 130–158.

Medina, J. F., Garcia, K. S., & Mauk, M. D. (2001). A mechanism for savings in the cerebellum. *Journal of Neuroscience, 21,* 4081–4089.

Millenson, J. R., Kehoe, E. J., & Gormezano, I. (1977). Classical conditioning of the rabbit's nictitating membrane response under fixed and mixed CS–US intervals. *Learning and Motivation, 8,* 351–366.

Moore, J. W. (1972). Stimulus control: Studies of auditory generalization in rabbits. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 206–230). New York: Appleton-Century-Crofts.

Moore, J. W., & Choi, J.-S. (1997). Conditioned response timing and integration in the cerebellum. *Learning & Memory, 4,* 116–129.

Moore, J. W., & Newman, F. L. (1966). Intertrial stimuli and generalization of the conditioned eyelid response. *Journal of Experimental Psychology, 71,* 414–419.

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin, 97,* 316–333.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532–552.

Perrett, S. P., Ruiz, B. P., & Mauk, M. D. (1993). Cerebellar cortex lesions disrupt learning-dependent timing of conditioned eyelid responses. *Journal of Neuroscience, 13,* 1708–1718.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Rosenfield, M. E., & Moore, J. W. (1995). Connections in cerebellar cortex (Larsell's HVI) in the rabbit: A WGA-HRP study with implications for classical eyeblink conditioning. *Behavioural Neuroscience, 109,* 1106–1118.

Scandrett, J., & Gormezano, I. (1980). Microprocessor control and A/D data acquisition in classical conditioning. *Behavior Research, Methods and Instrumentation, 12,* 120–125.

Schneiderman, N. (1972). Response system divergencies in aversive classical conditioning. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research* (pp. 341–376). New York: Appleton-Century-Crofts.

Sears, R. J., Baker, J. S., & Frey, P. W. (1979). The eyeblink as a time-locked response: Implications for serial and second-order conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 5,* 43–64.

Siegel, S., Hearst, E., George, N., & O'Neal, E. (1968). Generalization gradients obtained from individual subjects following classical conditioning. *Journal of Experimental Psychology, 78,* 171–174.

Smith, M. C. (1968). CS-US interval and US intensity in classical conditioning of the rabbit's nictitating membrane response. *Journal of Comparative and Physiological Psychology, 66,* 679–687.

Solomon, P. R., & Moore, J. W. (1975). Latent inhibition and stimulus generalization of the classically conditioned nictitating membrane response in rabbits (*Oryctolagus cuniculus*) following dorsal hippocampal ablation. *Journal of Comparative and Physiological Psychology, 89,* 1192–1203.

Sutton, R. S., & Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel & J. W. Moore (Eds.), *Learning and*

computational neuroscience (pp. 497–537). Cambridge, MA: MIT Press.

Svensson, P., Ivarsson, M., & Hesslow, G. (1997). Effect of varying the intensity and train frequency of forelimb and cerebellar mossy fiber conditioned stimuli on the latency of conditioned eye-blink responses in decerebrate ferrets. *Learning and Memory, 4,* 105–115.

Thompson, R. F., Donegan, N. H., Clark, G. A., Lavond, D. G., Lincoln, J. S., Madden, J., et al. (1987). Neuronal substrates of discrete, defensive conditioned reflexes, conditioned fear states, and their interactions in the rabbit. In I. Gormezano, W. F. Prokasy, & R. F. Thompson (Eds.), *Classical conditioning* (3rd ed., pp. 371–399). Hillsdale, NJ: Erlbaum.

Winters, R. W., McCabe, P. M., & Schneiderman, N. (2002). Functional utility and neurobiology of conditioned autonomic responses. In J. W. Moore (Ed.), *A neuroscientist's guide to classical conditioning* (pp. 46–85). New York: Springer.

Received February 6, 2002
Revision received July 23, 2002
Accepted August 2, 2002 ■