Computer-Assisted Behavioral Assessment of Pavlovian Fear Conditioning in Mice

Stephan G. Anagnostaras, Sheena A. Josselyn, Paul W. Frankland, and Alcino J. Silva[1]

Departments of Neurobiology, Psychiatry, and Psychology
Brain Research Institute
University of California
Los Angeles, California 90095 USA

Abstract

In Pavlovian fear conditioning, a conditional stimulus (CS, usually a tone) is paired with an aversive unconditional stimulus (US, usually a foot shock) in a novel context. After even a single pairing, the animal comes to exhibit a long-lasting fear to the CS and the conditioning context, which can be measured as freezing, an adaptive defense reaction in mice. Both context and tone conditioning depend on the integrity of the amygdala, and context conditioning further depends on the hippocampus. The reliability and efficiency of the fear conditioning assay makes it an excellent candidate for the screening of learning and memory deficits in mutant mice. One obstacle is that freezing in mice has been accurately quantified only by human observers, using a tedious method that can be subject to bias. In the present study we generated a simple, high-speed, and highly accurate algorithm that scores freezing of four mice simultaneously using NIH Image on an ordinary Macintosh computer. The algorithm yielded a high correlation and excellent linear fit between computer and human scores across a broad range of conditions. This included the ability to score low pretraining baseline scores and accurately mimic the effects of two independent variables (shock intensity and test modality) on fear. Because we used a computer and digital video, we were able to acquire a secondary index of fear, activity

suppression, as well as baseline activity scores. Moreover, we measured the unconditional response to shock. These additional measures can enhance the sensitivity of the assay to detect interesting memory phenotypes and control for possible confounds. Thus, this computer-assisted system for measuring behavior during fear conditioning allows for the standardized and carefully controlled assessment of multiple aspects of the fear conditioning experience.

In Pavlovian fear conditioning, a conditional stimulus (CS, usually a tone) is paired with an aversive unconditional stimulus (US, usually a shock) in a novel conditioning chamber. After even a single pairing, the animal comes to exhibit a long-lasting fear to the CS, which can be measured as freezing, an adaptive defense reaction, in rats and mice (Bolles 1970; Fanselow and Bolles 1979; Fanselow 1990; Bourtchuladze et al. 1994). Animals will also exhibit fear to the conditioning chamber and other circumstances surrounding the conditioning episode; this phenomenon has come to be known as context conditioning (Nadel 1980; Kim and Fanselow 1992; Phillips and LeDoux 1992; Bourtchuladze et al. 1994; Paylor et al. 1994).

In recent years, considerable progress has been made concerning the molecular, cellular, and neuroanatomical substrates of fear conditioning (Silva et al. 1998). For example, the amygdala has been found to play an essential role in fear conditioning, because lesions of this structure produce a general disruption of conditioned fear (Phillips and LeDoux 1992; Lee et al. 1996; Maren and Fanselow 1996; Maren et al. 1996b; McNish et al. 1997). Moreover, blockage of N-methyl-D-aspartate (NMDA) receptors in the amygdala, which disrupts the in-

[1]Corresponding author.

duction of long-term potentiation (LTP), blocks the acquisition of fear conditioning (Collingridge et al. 1983; Kim et al. 1991; Campeau et al. 1992; Maren et al. 1996a; Rogan et al. 1997; Lee and Kim 1998). In contrast, the hippocampus plays a more selective role in fear conditioning, because lesions of the hippocampus (or NMDA receptor disruption) produce a more selective deficit in contextual fear rather than tone fear (Phillips and LeDoux 1992; Chen et al. 1996; Logue et al. 1997; Maren and Fanselow 1997; Frankland et al. 1998; Gerlai 1998; see also Cho et al. 1999). Furthermore, post-training hippocampal lesions produce a temporally graded retrograde amnesia (RA) of contextual fear, in a manner analogous to the RA of declarative memory in humans (Scoville and Milner 1957; Kim and Fanselow 1992; Squire 1992; Maren et al. 1997; Anagnostaras et al. 1999). Thus, evidence has amassed that contextual fear conditioning may be an assay of function of the same memory system that mediates declarative learning in humans and that fear conditioning is dependent on the induction of LTP, a prominent synaptic model of learning.

Advances in mouse molecular biology and genetics have generated many new tools that may be used to examine the molecular and cellular mechanisms that subserve learning and memory. Fear conditioning is an excellent candidate for the genetic analysis of learning and memory for several reasons: (1) Mice exhibit robust conditioned freezing even after a single training trial; (2) the assay is not labor intensive and for this reason may be useful in large screening studies; (3) fear conditioning is a form of Pavlovian conditioning that has been thoroughly characterized psychologically; (4) the equipment required is relatively standardized and widely available; (5) the learning episode is punctate and brief potentially allowing memory phases to be dissociated; (6) there is extensive information regarding the neuroanatomical substrates of fear; and (7) fear conditioning has already been used extensively in studies of mutant mice.

One limitation of fear conditioning has been that the measurement of freezing behavior depends on the direct observation by an experimentally blind human observer. Freezing is defined as the absence of any movement and is usually sampled every 5–10 sec or measured continuously using a stopwatch (Fanselow and Bolles 1979; Phillips and LeDoux 1992; Bourtchuladze et al. 1994; Paylor et al. 1994; Anagnostaras et al. 1999). Although freezing has proven to be a reliable index

of fear, direct observation of freezing is quite tedious and may be susceptible to observer bias. Moreover, it would be useful to have other indices of fear because it is possible that the freezing response, rather than conditioned fear, may be compromised by any particular experimental manipulation (Good and Honey 1997; McNish et al. 1997; Maren 1998; Anagnostaras et al. 1999).

In the present study, we first report an automated technique for the measurement of freezing and discuss criteria we believe that any computerized system should meet. Second, the use of an automated system allows us to measure a secondary index of fear, activity suppression, that expands the ability of this assay to detect deficits in fear. Third, we examine the unconditional response (UR) to shock in mice using the technique described more extensively in a forthcoming paper using rats (see B.P. Godsil, J.R. Spooner, S.G. Anagnostaras, G.D. Gale, and M.S. Fanselow, unpubl.). Finally, we verify the retrograde amnesia of contextual fear produced by dorsal hippocampal lesions using the new computerized system.

Materials and Methods

CONDITIONING CHAMBERS

Each conditioning chamber (32 cm wide × 25 cm high × 25 cm deep; Med-Associates Inc., ENV-008VP) was equipped with a speaker (ENV-224AM) in the side wall of the chamber and a stainless steel grid floor designed for mice (ENV-005AX; 36 rods, each rod 2-mm diameter, 8-mm center to center; the front and back anchoring walls were custom ordered as clear acrylic). The grid floor was over a stainless steel drop-pan (ENV-007A3), which was lightly cleaned with 95% ethyl alcohol to provide a background odor. The front, top, and back of the chamber were made of clear acrylic and the two sides were made of modular aluminum (ENV-008FP). The conditioning chambers were arranged in a 2 × 2 configuration on a steel rack. The rack was in an isolated room; the room walls were painted white, and, in particular, the wall behind the chambers was matte white. The room was lit by one overhead 100-W incandescent bulb in a translucent fixture and two tall 300-W halogen torch lamps, which were to the sides of the rack. Background noise (65-dB) was supplied by a HEPA air cleaner set to medium position (CFM-52500; Honeywell Inc.). The rack was directly in front of a wall-mounted color camera (Sony SSC-S20),

which captured all four chambers in its view and was connected to a computer and video equipment in an adjacent room (see below). Each chamber was connected (ENV-407A) to a solid-state scrambler (ENV-412) and an audio stimulus generator (ANL-926) located in a cabinet (SG-6500; SG-220; connected to power source SG-501) in the adjacent room. Each scrambler was connected (SG-219) to an electronic constant current shock source (ENV-413) located in 2 cabinets (SG-6080C), and all stimuli were controlled via an interface (DIG-729; DIG-700P1) connected to a Gateway Windows 98 computer running Windows Med-PC (v. 3.1) and Borland Turbo Pascal software (SOF-700W; SOF-303W).

VIDEO/COMPUTER EQUIPMENT

The camera was connected via an S video cable to an inexpensive Iomega Buz video digitizer (Iomega Inc.) in a Power Macintosh G3/266 (Apple Computer Inc.) equipped with 288 MB of RAM. The software used was NIH Image (v. 1.62), Microsoft Excel (Office 98), and Quickeys (v. 3.52; CE Software Inc., West Des Moines, IA; Quickeys macros were used for convenience; we have subsequently rewritten them into NIH Image). One important note is that at the time of this writing, NIH Image is incompatible with the Macintosh G3 series built-in digitizer supplied by Apple (non-G3 digitizers will work). However, we achieved satisfactory compatibility and performance using the Iomega Buz and NIH should be consulted regarding compatibility with other digitizers. The camera was also connected via an RCA cable to a standard VHS VCR (Sony SLV-788HF) and displayed on a 20-inch video display for human observers. The VCR output was also connected to the digitizer via an RCA cable. For activity burst analyses (see below), a Turbomouse trackball (Kensington Microware Limited, San Mateo, CA) with four programmable buttons was particularly useful.

SUBJECTS

In the pilot experiment, eight female C57BL/6J mice (8 months old at the time of testing) purchased from the Jackson Laboratory (Bar Harbor, ME) were used. For Figure 1 four female C57BL/6NTacfBr mice (4–5 months old) purchased from Taconic Farms (Germantown, NY) were used. In experiments 1 and 2, 40 and 20 female C57BL/

6NTacfBr mice (4–5 months old), respectively, purchased from Taconic Farms were used. All mice were housed in groups of four, were maintained on a 14:10 light/dark cycle, and had unrestricted access to food and water. All testing was done during the mid-light cycle.

OBSERVER SCORING

Two blind observers scored freezing by viewing a video monitor in a room separate from the mice during the test period. Freezing was defined as the absence of visible movement except that required for respiration (fluctuation in the volume of the thorax) and was scored according to an 8-sec instantaneous time-sampling procedure (Fanselow and Bolles 1979). Every 2 sec a 250-msec computer-generated beep signaled the observer to score; the observer scored the animal as freezing or not at that instant and then proceeded to the next chamber when the next beep occurred; after chamber 4, the observer started again with chamber 1. Thus each animal was scored every 8-sec, yielding eight observations for each mouse for each minute of each test. These were converted into percent time freezing scores by dividing the number of freezing observations by the total number of observations for each mouse. The two observers (S.A. and S.J.) had many years of experience quantifying behavior in animals.

COMPUTER SCORED-FREEZING

The principle feature of any computerized freezing system is simple: Because freezing is defined as the absence of any visible movement, movement must be measured, and then near-zero movement can be equated with freezing. The movement threshold required to yield good freezing scores can be determined empirically by examining the correspondence (correlation and linear fit) between computerized and observer-based scoring. One problem is that the movement index must be sensitive enough to detect small movements such as grooming that are clearly not scored as freezing by human observers. Moreover, the movement scores provided by the computer for these small movements need to be substantially greater than the normal noise that occurs in all analog signals. That is, the freezing threshold must be both below the signal for small movements but well above that of the normal video noise. Finally,

the movement index must not be particularly computing intensive, to allow the assessment of 5 min of freezing data from a squad of four mice in a practical amount of time (1 or 2 min).

Baseline noise data from our video system was collected by digitizing 122 sec of full-screen live video of the empty conditioning chambers (hereafter referred to as the "noise video") at 1 Hz in 8-bit greyscale using NIH Image with the Iomega Buz operating as a Quicktime digitizer. This video was separated into four separate stacks cropped to the portions of the chambers where a mouse could be found (we have subsequently developed software to skip this step, using the NIH Image region of interest (ROI) techique). Three candidate-dependent variables were taken to index movement from frame-to-frame. First, we could directly compare frame-to-frame by thresholding each frame to a 1-bit image, taking the pixels that differed with the next frame (exclusive of function), and then counting them. This computing-intensive technique has been used previously by Kim et al. (1993). Second, activity could be indexed simply by taking some measure of each frame and comparing with the same measure in the next frame. Two measures were taken, which are both automated and high speed in NIH Image: mean 8-bit grayscale optical density of the picture and standard deviation of the optical density. The variability of these measures was examined for several noise videos; to reduce noise overall, several modifications were made, including reducing the reflective surfaces in the room (reducing the amount of gloss paint behind the chambers and reducing metal surfaces), altering the camera angle and digitized space to reduce visible metal surfaces, and improving the evenness of lighting by reducing direct lighting and increasing brightness overall. However, modifications were not exhaustive in this respect, in that, for example, a single camera was not restricted to a single conditioning chamber (which would increase the cost) and a more expensive digitizer was not used. This placed some chambers at a disadvantage because of the camera angle and lighting (the lower chambers in particular), but this was used to test the resiliency of the system rather than altering the conditioning set-up any further.

After examination of the noise data it was immediately apparent that the "number-of-pixels" method was not fast enough to be practical. It took > 10 min to obtain measures of a single squad (four mice) of 5-min data (even on the high-end computing equipment), whereas the other two dependent measures could be computed in <60 sec. Therefore, we primarily concentrated on these more practical measures of digital video movement.

After this, two squads of four mice each were placed into a simple conditioning paradigm so that we could obtain pilot data (Pilot experiment). Prior to placement of the mice, a final 122-sec noise video was digitized (as above) and was used in all further noise analyses. Mice were then placed into the chambers and after 122 sec they received a single 2-sec, 1-mA scrambled footshock. Sixty sec later they were returned to their home cages, and 30 min later we placed them back in the chambers for a 122-sec contextual freezing test. Freezing was scored by human observers for both the 122-sec baseline (BL) and 122-sec context (test) periods as is shown in Table 1A. Human measurements showed that these mice exhibited the characteristically near-zero baseline freezing (0.4%) that is usually observed in a novel context and typical average freezing (32.1%) for the test period. Live full-screen 8-bit greyscale video during both periods was also digitized at 1 Hz using NIH Image. This video was cropped into four separate stacks (this step was subsequently eliminated). Mean density and standard deviation of the density (disturbance) were measured (to eight decimal places to increase sensitivity) for every frame of all four stacks using a Quickeys macro (subsequently replaced with an Image macro) and NIH Image; this generated 122 density and 122 disturbance values for each mouse for each test. We pasted these values into Microsoft Excel to process them into freezing scores using several algorithms to determine which method could yield accurate freezing estimates.

In all cases, the basic strategy was the same: We used the change in the density or disturbance as an index of movement from frame to frame and set a low threshold (above video noise); any observations below this threshold are defined as freezing. These freezing observations were computed into a percent time freezing for each mouse by dividing them over the total number of observations. These percent time freezing scores were compared with human observers for validity and calibration.

Retrospectively, several criteria were important. (1) The algorithm should use identical thresholds for every test, every chamber, and every mouse, as having many user-defined free parameters could actually increase subjective bias in the

Table 1: *Pilot data comparing various freezing measures*

		Algorithm	Thr	BL%	Test%	Noise%	Thr/25 %
A.	*Human*	visual sampling	subj.	0.4	32.1	(empty chambers)	
B.	*Density [mean (μ) optical density for each frame]*						
		$\|(\mu_1 - \mu_2)\|$	0.05	9.6	31.2	62.5	48.1
		$\|(\mu_1 - \mu_2)\|$ & adj	0.08	5.6	34.2	77.3	65.6
		$\sigma(\mu_1, \mu_2)$	0.04	10.2	34.9	55.0	42.5
		$\sigma(\mu_1, \mu_2)$ & adj	0.05	4.4	30.1	72.9	59.6
		$\sigma(\mu_1, \mu_2, \mu_3)$	0.06	3.8	31.9	87.1	69.2
		$\sigma(\mu_1, \mu_2, \mu_3)$ & adj	0.07	2.0	31.2	92.1	75
C.	*Disturbance [standard deviation (σ) of density for each frame]*						
		$\|(\sigma_1 - \sigma_2)\|$	0.05	7.1	33.1	90.2	80.6
		$\|(\sigma_1 - \sigma_2)\|$ & adj	0.06	1.6	31.0	94.8	85.4
		$\sigma(\sigma_1, \sigma_2)$	0.03	6.3	29.2	93.1	84.2
		$\sigma(\sigma_1, \sigma_2)$ & adj	0.05	2.5	33.4	97.1	91.0
		$\sigma(\sigma_1, \sigma_2, \sigma_3)$	0.06	2.7	31.9	100	98.5
		$>> \sigma(\sigma_1, \sigma_2, \sigma_3)$ & adj	0.07	0.8	31.4	100	100

Table 1 depicts data from the Pilot experiment using NIH Image to score movement and index freezing behavior. (*A*) Average measurements taken by visual sampling by two human observers. (BL%) Freezing during the 122-sec baseline prior to any shock. (Test%) Freezing during the 122-sec contextual fear test. (*B*) Measurements using changes in mean optical density from video frame to frame. (Algorithm) Comparison made of successive frames [(μ) mean optical density; (σ) s.d. of the optical density; subscripts 1, 2, 3 refer to successive frames]. (Thr) Threshold for this algorithm that yielded the best fit with human-scored freezing during the context test. (Noise%) Ability to reject noise as nonmovement, from video taken of the chambers without any mice in them. (Thr/25%) Ability to still reject noise if the threshold value is arbitrarily decreased by 25%. (*C*) Measurements using changes in the disturbance (s.d. of the optical density) from frame to frame. The measure (disturbance) and algorithm chosen [$\sigma(\sigma_1, \sigma_2, \sigma_3)$ & adj)] for use in experiment 1 yielded equivalent context scores to human observers, exhibited low baseline, and robustly rejected noise.

measurement of freezing rather than decrease it. (2) The method should score low baseline freezing using the exact same criteria that it scores context test freezing equivalent to human observers. High baseline freezing reflects a movement threshold that is set too high and is scoring small movements (such as grooming) as freezing; this high baseline would "clip" the overall dynamic range of the freezing assay by introducing an artificial floor (e.g., Valentinuzzi et al. 1998). (3) The method should score all video noise as nonmovement. That is, if there is no mouse in the chamber, or if a mouse shows 100% freezing, the computer should consider this freezing as nothing is moving inside the chamber. We refer to this as noise rejection. If any noise is accepted as nonfreezing movement the method would be incapable of scoring very high freezing, because some proportion of video noise is interpreted by the computer as animal movement. This would clip the dynamic range of the freezing assay by introducing an artificial ceiling. That is, the dynamic range of human-observed freezing is 0–100%; the dynamic range of an automated system will be defined by the baseline (de-

termined primarily by the ability to detect small movements) to the ceiling (determined by the ability to reject noise).

We used these criteria to evaluate several algorithms for scoring freezing from density and disturbance data. In each case, a particular frame-to-frame comparison was made. Then, a global threshold value was set (to 2 decimal points) that yielded the best fit (estimated by the difference between computer and human means) between the computerized measurement and the freezing data obtained during the context test by human observers. This measurement was the average of both human observers. Then, baseline and noise video data were examined using this same threshold value and algorithm. These are shown in Table 1B,C. An acceptable automated system should accurately score both freezing during the context test, during baseline, and during episodes of 100% freezing. That is, the video noise should not be scored as movement. After going through an evolutionary series of algorithms that systematically improved their ability to meet these criteria (Table 1), we examined the overall correspondence be-

tween human and automated freezing and then performed additional calibration using a larger and more systematic set of data (Experiment 1). We were aware that there may have been a need to adjust the threshold value for freezing. Therefore, we subjected video noise data to an artificial decrease in threshold of 25%.

We began by using the changes in mean density as our measure of freezing (Table 1B) (Maren et al. 1996; Richmond et al. 1998). The first comparison that we evaluated was the absolute difference in mean density from one frame to the next $[|(\mu_1 - \mu_2)|]$. Although it could score freezing during the context test, this algorithm yielded a high baseline freezing and did not reject video noise well. To reduce baseline, we introduced the criterion that two adjacent nonmoving (below threshold) observations needed to be made before freezing was scored $[|(\mu_1 - \mu_2)|]$ & adj; hereafter referred to as the adjacent observations criteria). This could reduce false baseline freezing scores by eliminating instantaneous momentary inactivity that may not reflect fear (Maren 1998; Richmond et al. 1998). This reduced baseline and improved noise rejection, but not to acceptable levels. Thus, a simple difference in mean density is not an acceptable index of freezing because the threshold required to score accurate context freezing is both above the signal given by small movements (yielding a false high baseline) and below the normal variability of the noise signal (yielding poor video noise rejection). Because the poor signal-to-noise ratio, under conditions of small movement, seemed to be the primary problem, we began to focus our efforts on how to control for it.

We changed the frame-to-frame comparison from a simple difference score of mean densities to the standard deviation of the mean density values for adjacent frames. This change allowed us to expand the comparison to more than two frames. First, we examined the standard deviation of mean densities between two frames $[\sigma(\mu_1,\mu_2)]$. This yielded similar results to the difference score of mean densities. Therefore, we then used this same algorithm with the adjacent observations criteria added. This modification, however, resulted only in minor improvements. Finally, we determined whether an algorithm that incorporate the standard deviation of three adjacent frames would improve the measure of freezing $[\sigma(\mu_1,\mu_2,\mu_3)]$. For example, observation 1 was $\sigma(\mu_1,\mu_2,\mu_3)$ and observation 2 was $\sigma(\mu_2,\mu_3,\mu_4)$. These changes improved noise rejection, but not to acceptable levels. Add-

ing the adjacent observations criteria improved both baseline and video noise rejection. Nevertheless, it was apparent that frame-to-frame comparisons of mean density could not be used to score freezing accurately in mice. Although this approach could yield a high correlation with human scoring within certain ranges of freezing (above the baseline and below the video noise level), it would randomly introduce noise in the measurement. This noise is especially problematic at low and high levels of freezing.

These problems led us to change our strategy. We decided to use a strategy in which the noise is measured rather than the actual signal itself. For example, a similar strategy has been used to measure miniature excitatory post-synaptic potentials (minis). The signal from these minis is smaller than the electric noise recorded. We assumed that the video noise is relatively constant and that changes in the video noise could be used to measure movement by the animal. To implement this idea, we measured the standard deviation of the density for each video frame (Table 1C). This measure resulted in lower baselines and much better video noise rejection. The algorithm finally chosen used the standard deviation of the disturbance values from three successive frames, as well as the adjacent-observations criterion $[\sigma(\sigma_1,\sigma_2,\sigma_3)$ & adj]. The implementation of this algorithm resulted in a near-zero baseline and 100% noise rejection under the same criteria that it accurately scored context freezing. Importantly, it also completely rejected video noise when subjected to large (up to 50%) decrease in threshold (Table 1). This is important because it shows that the system can tolerate a wide variety of noise conditions.

Compared with human measurements, this algorithm $[\sigma(\sigma_1,\sigma_2,\sigma_3)$ & adj] yielded accurate baseline (Wilcoxon nonparametric paired comparison, $P = 0.6$) and context [univariate ANOVA, $F(1,14) = 0.001$, $P > 0.99$] measurements of freezing data that were not significantly different from human-scored freezing. Therefore, this algorithm was used in all subsequent experiments to score freezing.

One important additional note is two additional frames of video were added at the end of each captured movie. This allows the system to generate, for example, 120 comparisons for 122 frames of data.

An exact example of how this final algorithm was used to score freezing is depicted in Figure 1. Figure 1A shows one frame of four mice in four
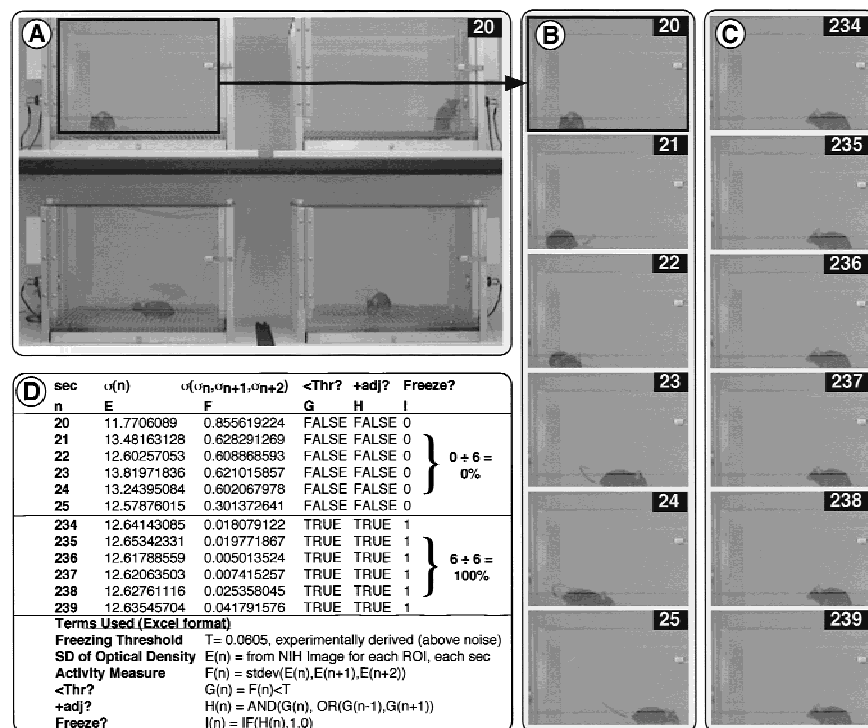
Anagnostaras et al.

Figure 1: Visual depiction of the technique. (*A*) Full-screen video of all four chambers was captured at 1 Hz by NIH Image. The mice shown here were recorded for 302 sec and given a shock at 150 sec. The full-screen frame shown is from 20 sec (black numbers indicate time). (*B*) Video was analyzed by separating the video in four ROIs corresponding to the inside of the chambers. Shown here is an example of a naive mouse, 20–25 sec after being placed in the chamber. (*C*) Shown here is the same mouse exhibiting post-shock freezing during seconds 234–239. (*D*) Values and processing required to compute freezing from the video frames shown in *B* and *C*. The standard deviation of the optical density from the region of interest (chamber) for each frame (second) was obtained from NIH Image (shown in column E). The standard deviation of this value from three adjacent seconds was then computed as a movement score (F). This value (F) was used as the activity measure itself. It was determined if this was below the threshold for freezing (G), set just above the noise video signal (T). Then a freezing observation was only made if there was also an adjacent observation (i.e., the next or previous observation was also freezing; H). Finally, this was converted to a 1 for freezing and 0 for nonfreezing so these could be counted (I).

chambers captured simultaneously for 5 min, during which a 2-sec 0.5-mA shock was delivered at 150 sec. Figure 1B shows seconds 20–25 from the mouse in the upper-left chamber, during which it was quite active. Figure 1C shows this mouse during seconds 234–239 when it was exhibiting post-shock freezing. Figure 1D shows standard deviation of optical density reported from Image for these frames and how it was used to compute freezing.

Experiment 1: The Effects of Shock Intensity and Test Modality

To examine the validity of the algorithm used to score freezing in the Pilot experiment, we de-cided to systematically vary fear by examining the effects of shock intensity (0.25, 0.50, 0.75, or 1.0 mA; between-groups variable) and the effects of test modality (context or tone; within-subjects variable). This allowed us to examine several criteria important for any automated measure of freezing. (1) The system should accurately reflect the low baseline freezing observed in a novel context. Mice and rats do not freeze on placement into a novel context (e.g., Frankland et al. 1998; Anagnostaras et al. 1999). (2) The system should accurately mimic the effects of independent variables known to manipulate fear (such as shock intensity or test modality). (3) The system should exhibit a high correlation with human observer-based scoring, indicating that it robustly captures the variance in

human-scored freezing. (4) The system measuring freezing should exhibit a nearly perfect linear fit with human-scored freezing such that qualitatively high or low values can be compared with laboratory norms in previously obtained (human-scored) data. Moreover, we sought to exploit the additional capabilities of the computer to score movement, by examining whether activity could be a useful secondary index of fear. Finally, we also used the UR to the shock as an index of pain reactivity in mice.

CONDITIONING

Forty mice were placed into the conditioning chambers (see above) and after a 120-sec baseline period were given a 30-sec tone (2.8-kHz, 85-dB/A-scale) that coterminated with a 2-sec scrambled footshock. Mice were randomly assigned to one of four shock intensities: 0.25, 0.50, 0.75, or 1.0 mA ($n = 10$/group). The animals were returned to their home cages 150 sec after the shock.

CONTEXT TEST

One day after conditioning, all of the mice were returned to the same conditioning chambers for a contextual fear test. They were placed into the conditioning chambers and, 5 min later, returned to their home cages.

TONE TEST

One day after the context test, the animals were returned to the conditioning chambers for an on-baseline contextual fear test. After a 2-min baseline period, the training tone was presented continuously for 3 min. Freezing was scored by human observers for both 5-min tests; their scores were converted to 80 percent time freezing observations for each mouse (40 context and 40 tone) for each 5-min test period.

UR MEASUREMENT

This measure will be described in more detail elsewhere (see B.P. Godsil, J.R. Spooner, S.G. Anagnostaras, G.D. Gale, and M.S. Fanselow, unpubl.). Briefly, full-screen video for the 2-sec period immediately before the shock (baseline) and 2-sec period during the shock were digitized at 10 Hz using NIH Image. X–Y coordinates were ob-

tained for each frame for each mouse by a blind observer using the wand auto-measure tool; these were imported into Microsoft Excel. Distance traveled measured in pixels between successive frames was computed using the distance formula $[v((x_n - x_{n+1})^2 + (y_n - y_{n+1})^2)]$; these values were converted into real distance in centimeters using known landmark distances in the video frame. This was then converted into velocity in cm/sec by dividing by time.

Results

THRESHOLD ADJUSTMENTS

To optimize the fit between automated and human freezing measurements, we adjusted the movement threshold below which freezing was scored. Threshold adjustment was performed by examining the analyses below under a number of threshold values that were systematically varied from that obtained from pilot data (the original threshold was 0.07; Table 1). Although the original value provided excellent results, we were able to obtain qualitatively superior data through minor adjustment, because in the pilot experiment thresholds were only adjusted to 2 decimal points. The final threshold value used for all analyses below was 0.065. This value still rejected 100% of the noise video from the Pilot experiment.

BASELINE FREEZING

As with our pilot data, we also examined baseline freezing data in the 120 sec prior to the first tone–shock pairing on the conditioning day. Computer-scored data yielded a low and typical baseline ($n = 40$; mean = 1.4 ± 0.4%; range = 0–10%) indicating that our system accurately scored the near-zero freezing baseline that is typically observed prior to conditioning (see, e.g., Table 1; Fig 6A; Bourtchuladze et al. 1994; Frankland et al. 1998; Anagnostaras et al. 1999).

CORRELATIVE ANALYSIS

To determine what level of correspondence we could expect between the computer and human observers, we first examined the correlation between two human observers. The correlation for all 80 observations (40 context and 40 tone scores for the 5-min tests from each observer) was examined first (Fig. 2). Scores recorded by S.A. and S.J. correlated to a high and typical degree ($r = 0.90$, $P < 0.0001$) and yielded an excellent linear fit
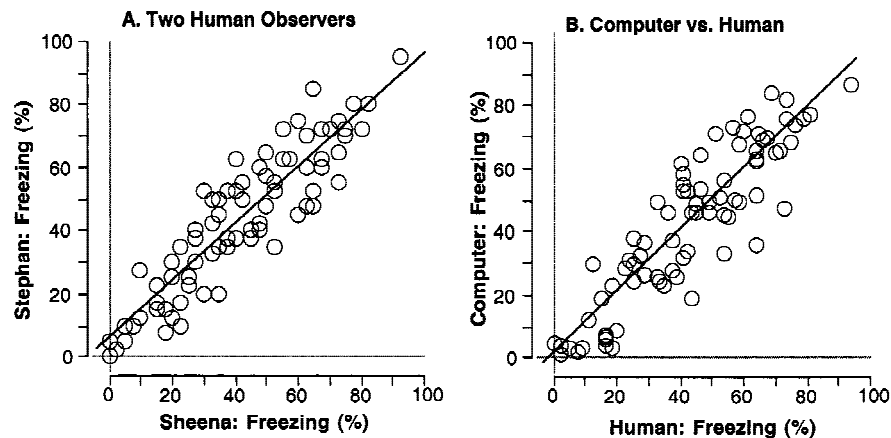
Figure 2: Experiment 1. Comparison of human and computer scoring: correlation and linear fit. (*A*) Two human observers scored the same 40 mice freezing during 5-min context and tone tests (80 observations for each observer; % time freezing). The observations between the two observers exhibited a high correlation and excellent linear fit. (*B*) Computer and human (average of the two observers) scores were compared. The observations exhibited a high correlation and excellent linear fit, indicating that the computerized scoring system yielded accurate scores across a wide range of freezing.

(SA = 6.0 + 0.90SJ) (Fig. 2A). To ensure that this correlation was not driven by variance attributable to the independent variables, we also examined within-cell correlations at each shock intensity and for context and tone separately. That is, we wanted to ensure that the correlation was not caused simply because the observations covaried as a result of the independent variables. In all cases (at each of four shock intensities and for context or tone), the correlations remained high ($r = 0.83$–0.92, $P < 0.0001$), indicating that the correspondence was robust under any of these conditions.

The two observers were then averaged and compared with computer-scored freezing. The correspondence between human and computer-scored freezing is similar to that between the two human observers (Fig. 2B). A high correlation ($r = 0.90$, $P < 0.0001$) and excellent fit were ob-

served (Computer = 1.2 + 0.98Human). The near-0 intercept and slope near 1 indicate that the computer does not systematically overestimate or underestimate freezing. We also examined within-cell correlations at each shock intensity and modality. In all cases, correlations remained high ($r = 0.78$–0.93, $P < 0.0001$) indicating a high degree of correspondence under any shock intensity and in either modality.

SHOCK INTENSITY AND MODALITY EFFECTS

We also examined whether the automated measurements could accurately reflect the effects of two independent variables, shock intensity (0.25, 0.50, 0.75, or 1.0 mA) and test modality (context or tone). Figure 3A depicts the effect of shock intensity on human (average of the

Figure 3: Comparison of human and computer scoring: impact of shock intensity and modality. (*A*) Shock intensity. Different groups of mice were given training under four shock intensities, which impacted freezing (context and tone, mean % time ± S.E.M.) measured on separate days. The computer accurately tracked the effect of shock intensity measured by human observers. (*B*) Modality. Mice were given both context and tone freezing tests (on separate days). The computer accurately mimicked the difference in conditioning to the tone and context detected by human observers.
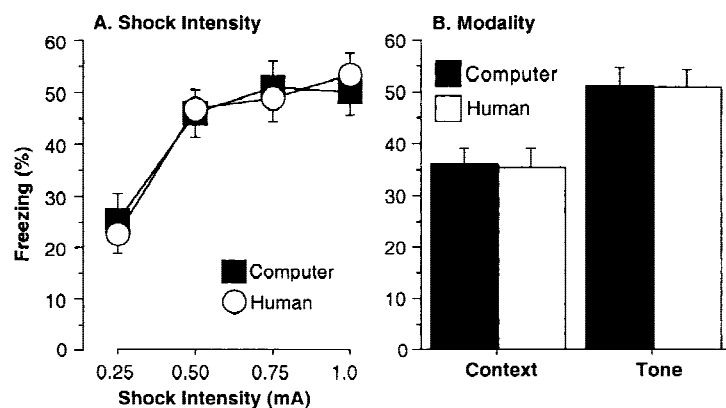
two observers) and computer-scored freezing, whereas Figure 3B depicts the effect of modality. This figure shows that the computer produced mean scores equivalent to the human observers. For shock intensity, we performed a 4 (shock intensity) × 2 (observer: computer or human) ANOVA. There was a significant main effect of shock [$F(3,152) = 16$, $P < 0.0001$; power = 1.0], no effect of observer type [$F(1,152) = 0.003$, $P > 0.9$], and no shock × observer interaction [$F(3,152) = 0.17$, $P > 0.9$], indicating that the mean scores the computer produced were equivalent to those produced by human observers at every shock intensity. For modality, we performed a 2 (modality) × 2 (observer: computer or human) ANOVA. There was a significant main effect of modality [$F(1,156) = 20$, $P < 0.0001$; power = 1.0], no effect of observer type [$F(1,156) = 0.002$, $P > 0.9$], and no modality × observer interaction [$F(1,156) = 0.0002$, $P > 0.9$], indicating that the mean scores the computer produced were equivalent to human scores under either modality.

Combined with the correlative analyses, these data show that our computerized system produces freezing scores that are quantitatively and qualitatively equivalent to those produced by human observers.

ACTIVITY AND ACTIVITY SUPPRESSION

One additional benefit of the automated system is that activity scores are collected in parallel with freezing. These scores can be useful in measuring baseline activity before the first tone–shock pairing (on the conditioning day) as this may be an easy way to measure generalized activity. For example, rats with hippocampal lesions and GABA-A knockout mice are hyperactive during this period

of time (DeLorey et al. 1998; Maren et al. 1998). Second, we wanted to examine whether activity during the context test might be a useful secondary index of fear. This could be useful because, on occasion, we have observed mice that do not freeze but appear to be afraid because they exhibit little activity during testing. Activity suppression is known to be an index of fear and anxiety in many organisms (Estes and Skinner 1941).

Mean raw activity scores [of $\sigma(\sigma_1, \sigma_2, \sigma_3)$; the dependent activity measure throughout these experiments; see Fig. 1D] were computed for the 120-sec baseline period prior to the tone-shock pairing (Fig. 4A). It was not surprising that there was no significant effect of shock intensity on baseline activity [$F(3,36) = 0.2$, $P = 0.9$], because this period occurred prior to the delivery of any stimuli. However, that the four randomly assigned groups were equivalent in baseline activity suggests this index is reliable. Although baseline activity was not expected to differ in the present study this measure has been sensitive to other manipulations in our laboratory. For example, we have found that this measure exhibits the characteristic decline known as habituation of exploratory activity observed in novel environments (data not shown; see, e.g., Anagnostaras et al. 1999).

We also examined if activity could be a useful index of fear during the context test. We computed average activity during the first 120 sec of the context test and found a significant effect of shock intensity [$F(3,36) = 3.1$, $P = 0.04$; power = 0.68], as context test activity was lower than baseline activity for every shock intensity group [separate paired comparisons, $F(1,9) > 28$, $P < 0.001$], indicating that this measure is an index of fear (Fig. 4A, context). However, the effect of shock intensity on activity during the context test was much weaker
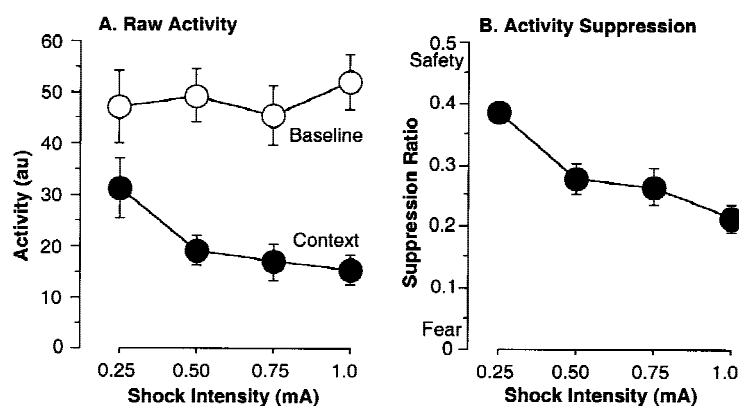

Figure 4: Activity and fear conditioning. (A) Raw activity was measured by the computer during the first 2 min of the training day and the first 2 min of the context test (mean arbitrary units ± s.e.m.). Raw activity was an indicator of fear, but it was not as powerful as freezing, because of a high degree of variability in the activity baseline. (B) Activity suppression was computed as suppression ratio = (activity during testing)/(activity during baseline + activity during testing) × (mean suppression ratio ± s.e.m.). Activity suppression was a powerful indicator of fear comparable to freezing.

than that on freezing (see above), indicating that raw activity is not a particularly powerful index of fear. This appeared to be due to a high degree of variability in baseline activity. This variability is not due to genetic heterogeneity because the subjects were inbred C57BL6 mice. To determine whether natural tendencies in baseline activity were contributing to the noisiness of context test activity as an index of fear, we examined the correlations between baseline activity, context test activity, and freezing. Context test activity was correlated with context freezing ($r = -0.65$, $P < 0.0001$), but it was also strongly correlated with baseline activity ($r = 0.64$, $P < 0.0001$), whereas freezing was not correlated with baseline activity ($r = -0.03$, $P > 0.8$). Thus, it appears that activity during the context test is heavily contaminated by the mouse's own baseline activity even before it is conditioned, whereas freezing is not. For this reason raw activity is probably not a useful index of fear in mice, because it would be heavily confounded with any manipulation that affects baseline activity; these effects are quite common in studies of mutant, drugged, or lesioned animals.

To control for this we generated a suppression ratio for activity. This standard technique has been used in many paradigms where baseline is variable. For example it has been used to score conditioned suppression of bar-pressing behavior during fear conditioning (e.g., Rescorla 1968). The suppression ratio (SR) was defined as (activity during test)/(activity during test + activity during baseline). Very low values indicate a high level of fear, 0.5 indicates no fear, and values greater than 0.5 can indicate conditioned safety (Annau and Kamin 1961). Activity suppression proved to be a powerful measure of context fear (Fig. 4B), yielding a large main effect of shock intensity [$F(3,36) = 9.3$, $P = 0.0001$; power = 1.0]. It was also well correlated with freezing ($r = -0.85$, $P < 0.0001$) and activity during the context test ($r = 0.73$, $P < 0.0001$), but importantly not with baseline activity ($r = -0.03$, $P > 0.8$). Measurements of activity suppression should compare a baseline and test period of equal lengths because there are changes in activity over time. Moreover, as with contextual freezing, it has been found previously that activity suppression during a contextual fear test is impaired by hippocampal lesions (Maren et al. 1998; Maren 1999). There is also evidence that suppression behavior may be dissociable from freezing and represent a separate response (Amorapanth et al. 1999).

UNCONDITIONAL RESPONSE

Another additional advantage of computerized scoring is the ability to score the UR to shock. This can be useful in examining whether the mouse's ability to feel the painful foot-shock stimulus is changed. Importantly, this measure of pain sensitivity does not require any additional animals or procedures and actually measures the response to the actual shock the mouse received during conditioning. The 2-sec activity burst UR and a 2-sec baseline period immediately prior to the shock were digitized from videotape at 10 Hz, and velocity in cm/sec was computed for these periods for the four shock intensity groups (Fig. 5). As expected, there was no effect on baseline activity, which occurred before the shock [$F(3,36) = 0.5$, $P > 0.6$], but a large effect on the activity burst [$F(3,36) = 32$, $P < 0.0001$; power = 1.0] indicating this is a powerful measure of shock reactivity in mice. Moreover, even for the lowest shock intensities used, the measurement of activity during the UR was greater than during the baseline [paired comparisons for each shock, UR vs. BL; $F(1,9) > 16$, $P < 0.01$]. Activity burst velocity was also well correlated with the set shock intensity ($r = 0.80$, $P < 0.0001$). Additionally, as will described in a forthcoming paper (B.P. Godsil, J.R. Spooner, S.G. Anagnostaras, G.D. Gale, and M.S. Fanselow, unpubl.), activity burst velocity was a good predictor of later freezing ($r = 0.57$, $P < 0.0001$) and of activity suppression during the
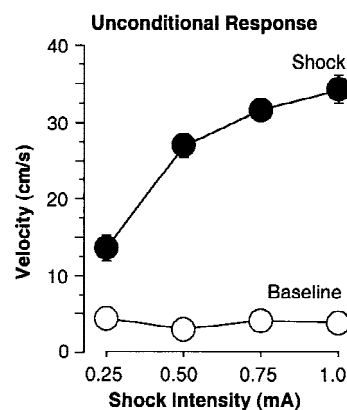

Figure 5: Unconditional response. Velocity in cm/s (mean ± s.e.m.) was computed for the 2 sec prior to (baseline), and during the first 2-sec shock on the training day. Activity burst velocity was reliably predicted by foot-shock intensity and may be a useful indicator of pain reactivity.

context test ($r = -0.63$, $P < 0.0001$). These correlations indicate that this measure of shock reactivity is well associated with the level of fear conditioning. It was also not correlated with baseline measures of activity (activity during 2-min baseline period, $r = 0.08$, $P > 0.63$; or the 2-sec baseline for the activity burst, $r = -0.17$, $P > 0.3$).

Experiment 2: Retrograde Amnesia of Contextual Fear After Electrolytic Dorsal Hippocampal Lesions

To verify the robust retrograde amnesia of contextual fear observed by human observers after electrolytic dorsal hippocampal (DH) lesions in rats and mice (Kim and Fanselow 1992; Frankland et al. 1998; Anagnostaras et al. 1999), Experiment 2 was performed. Freezing was scored by the computer, as described in Experiment 1.

TRAINING

Prior to receiving any lesion, 20 mice were given three unsignaled shocks (2-sec, 0.5 mA, each shock 1 min apart) 2 min after placement in the chambers.

SURGERY

One day after training, mice received electrolytic DH ($n = 11$) or sham lesions ($n = 9$), as described in Frankland et al. (1998). Briefly, mice were given anesthesia (chloral hydrate, 400 mg/kg, i.p.), mounted in a stereotaxic apparatus and then an insulated (except for the tip) electrode was lowered into the dorsal hippocampus bilaterally (from bregma, -1.8 posterior, ±1.9–2.0 lateral, -2.1–2.2 ventral). Anodal constant current (3.0 mA, 5 sec) was passed, the wound was closed and mice were allowed to recover for 1 week. Sham lesions were similar but no current was passed.

TESTING

One week after surgery, mice were returned to the original training context for a 3-min contextual freezing test.

HISTOLOGY

Post-mortem histological verification with thionin stain was performed as described previously (Frankland et al. 1998). Three mice were excluded from the DH group; one had only a cortical lesion placed too dorsally, and two others had only unilateral DH lesions. Exclusion of these mice did not affect the statistical conclusions below.

Results

Figure 6A depicts computer-scored post-shock freezing during the 5-min prelesion training period. Note the low baselines prior to the first shock (first 2 min) further demonstrating the ability of our system to score low freezing prior to training. Mice also exhibited post-shock freezing as shocks were delivered at the beginning of minutes 2, 3, and 4. There was significant development of post-shock freezing [MANOVA, main effect of measure, $F(1,4) = 35$, $P < 0.0001$], and no prelesion group differences [$F(1,14) = 0.3$, $P > 0.5$]. Figure 6B depicts computer-scored freezing during the 3-min contextual freezing test, performed 1 week after the lesion. DH lesions produced a robust retrograde amnesia of contextual fear, reducing freezing roughly in half [main effect of group, $F(1,14) = 17$, $P = 0.001$]. Thus, we verified the ability of our system to detect DH lesion-induced retrograde amnesia of contextual fear.
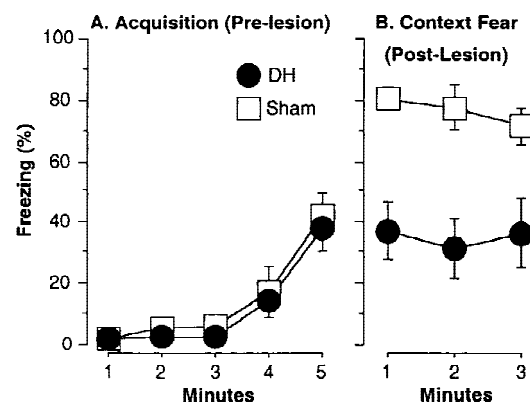

Figure 6: Experiment 2. Retrograde amnesia of contextual fear after electrolytic DH lesions. Mice were given three unsignaled shocks, followed on the next day by electrolytic DH lesions and a contextual fear test 1 week later. Freezing was scored by the computer. (*A*) Prelesion acquisition of post-shock freezing (mean ± S.E.M.). Naive mice exhibited characteristically low baseline freezing (first two min) and the development of post-shock freezing. (*B*) Post-lesion contextual fear. Mice that received DH lesions 1 day after training exhibited a robust retrograde amnesia of contextual fear.

Anagnostaras et al.

Discussion

Pavlovian fear conditioning provides a high-efficiency and well-controlled assay for the assessment of learning deficits in mice. The automation of the scoring of freezing behavior promises a substantial improvement over observer-scored freezing primarily because of a reduction in workload and possible subjective bias. We were able to accurately measure freezing behavior, as well as measure a number of other aspects of fear conditioning that would be difficult to measure without a computer. These included (1) the assessment of baseline activity, which could be a useful index of generalized activity and exploration of the contextual CS, (2) the measurement of activity suppression as a secondary index of fear, which could be useful in animals for which the freezing response is suspected to be impaired, and (3) the measurement of the foot-shock-elicited activity burst, which can be used to ensure that differences in shock reactivity do not account for apparent impairments in conditioning. We believe the measurement of these behaviors should be the essential parts of the fear conditioning assay.

Our examination of digital video-based scoring of mice freezing can be viewed as an indicator of potential problems in the automated measurement of freezing. It was difficult, for example, to accurately score the near-zero baseline freezing because of the relatively weak signal (compared with noise) when mice were only slightly active. Although we have not systematically examined it, we suspect this is a more severe problem in video-scoring of mice, rather than rats, because mice provide a poorer signal because of both their color and size (S.G. Anagnostaras, A.J. Silva, and M.S. Fanselow, unpubl.). For example, Maren (1998) and others have successfully used startle platforms to index activity and freezing in rats, but this method may not work nearly as well for mice because of the relatively smaller movement signals (S. Maren, pers. comm.).

Moreover, our examination of computer-scored automated freezing is supported not only by a correlative analysis but also by evidence of an excellent linear fit, robust noise rejection, and accurate baseline freezing. These latter three criteria have not, to our knowledge, been systematically examined in other studies (e.g., Richmond et al. 1998; Valentinuzzi et al. 1998). Although any movement index can potentially be used to index freezing, we believe any new measure should exhibit all four of these properties. Moreover, the measurement of an activity suppression ratio in addition to freezing is likely to increase the sensitivity of this assay in detecting interesting phenotypes. The addition of baseline activity measurement and reactivity to shock as standard components of a protocol for fear conditioning also enhance the interpretability of the results.

The development of high-efficiency automated protocols for assaying learning and memory deficits has become imperative because of several recent developments in genetics. First, investigators using traditional "reverse" genetics to make mutants have now generated several thousand mutants, most of which have not been screened for learning/memory phenotypes. Just the addition of these to our knowledge base would be a substantial contribution. However, it is likely that many of the genes that influence memory are still not known. For this reason, techniques using "forward" genetics, identifying genes from the phenotypes they confer, will become particularly important. One such method is to use a chemical agent such as *N*-ethyl-*N*-nitrosourea (ENU) to mutagenize mice and then screen the progeny for the altered phenotype of interest. This endeavor will require high-throughput automated systems that can screen thousands of animals and generate precise data.

Given the large growth of the field of behavioral genetics, the importance of standardized data and protocols must be emphasized. Observers with only limited behavioral experience may produce less reliable behavioral scores and any observer can be subject to bias. Therefore, an automated system should enhance reliability and reproducibility of findings. The taking of two parallel measures of fear, freezing and activity suppression, will further improve the interpretability of the findings.

However, some caution should be exercised when converting to an automated system for the measurement of behavior. Automated systems detach the experimenter from the actual behavior of the animal. These systems can be subject to artifacts generated by the algorithm that is used to give behavioral scores. For example, in automating freezing we found that any number of algorithms that might have a high correlation with human-scored freezing could still generate artifacts such as high baselines, low ceilings, and non-linear fit between computed and human scores. As such, considerable vigilance should be taken in ensuring that

an automated system produces accurate scores, because once committed to a particular system, problems will not unlikely be noticed later when observers are no longer actually watching the mice.

Nonetheless, we found that an ordinary Macintosh computer using an regular color camera could simultaneously score four freezing chambers accurately and quickly, using our adaptation to NIH's free Image software. Aside from a reduction in workload and experimenter bias, the use of digital video also allowed us to easily use activity suppression as a secondary index of fear, as well as score both baseline activity and the unconditional response to footshock. With some further improvements we believe this system for measuring fear conditioning deficits in mice can be used as a high-efficiency screen for learning and memory problems in mutant mice.

Acknowledgments

References

Amorapanth, P., K. Nader, and J.E. LeDoux. 1999. Lesions of periaquaductal gray dissociate conditioned freezing from conditioned suppression behavior in rats. *Learn. & Mem.* **6:** 491–499.

Anagnostaras, S.G., S. Maren, and M.S. Fanselow. 1999. Temporally graded retrograde amnesia of contextual fear after hippocampal damage in rats: Within-subjects examination. *J. Neurosci.* **19:** 1106–1114.

Annau, Z. and L.J. Kamin. 1961. The conditioned emotional response as a function of intensity of the US. *J. Comp. Physiol. Psychol.* **54:** 428–432.

Bolles, R.C. 1970. Species-specific defense reactions and avoidance learning. *Psych. Rev.* **72:** 32–48.

Bourtchuladze, R., B. Frenguelli, J. Blendy, D. Ciofi, G. Schutz, and A.J. Silva. 1994. Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. *Cell* **79:** 59–68.

Campeau, S., M.J. Miserendino, and M. Davis. 1992. Intra-amygdala infusion of the N-methyl-D-aspartate receptor antagonist AP5 blocks acquisition but not expression of fear-potentiated startle to an auditory conditioned stimulus. *Behav. Neurosci.* **106:** 569–574.

Chen, C., J.J. Kim, R.F. Thompson, and S. Tonegawa. 1996. Hippocampal lesions impair contextual fear conditioning in two strains of mice. *Behav. Neurosci.* **110:** 1177–1180.

Cho, Y.H., E. Friedman, and A.J. Silva. 1999. Ibotenate lesions of the hippocampus impair spatial learning but not contextual fear conditioning in mice. *Behav. Brain Res.* **98:** 77–87.

Collingridge, G.L., S.J. Kehl, and H. McLennan. 1983. Excitatory amino acids in synaptic transmission in the Schaffer collateral-commissural pathway of the rat hippocampus. *J. Physiol.* **334:** 33–46.

DeLorey, T.M., A. Handforth, S.G. Anagnostaras, G.E. Homanics, B.A. Minassian, A. Astourian, M.S. Fanselow, A. Delgado-Escueta, G.D. Ellison, and R.W. Olsen. 1998. Mice lacking the beta3 subunit of the GABAA receptor have the epilepsy phenotype and many of the behavioral characteristics of Angelman syndrome. *J. Neurosci.* **18:** 8505–8514.

Estes, W.K. and B.F. Skinner. 1941. Some quantitative properties of anxiety. *J. Exp. Psychol.* **29:** 390–400.

Fanselow, M.S. 1990. Factors governing one-trial contextual conditioning. *Anim. Learn. Behav.* **18:** 264–270.

Fanselow, M.S. and R.C. Bolles. 1979. Naloxone and shock-elicited freezing. *J. Comp. Physiol. Psychol.* **93:** 736–744.

Frankland, P.W., V. Cestari, R.K. Filipkowski, R.J. McDonald, and A.J. Silva. 1998. The dorsal hippocampus is essential for context discrimination but not for contextual conditioning. *Behav. Neurosci.* **112(4):** 863–874.

Gerlai, R. 1998. Contextual learning and cue association in fear conditioning in mice: A strain comparison and a lesion study. *Behav. Brain Res.* **95:** 191–203.

Good, M. and R.C. Honey. 1997. Dissociable effects of selective lesions to hippocampal subsystems on exploratory behavior, contextual learning, and spatial learning. *Behav. Neurosci.* **111:** 487–493.

Kim, J.J. and M.S. Fanselow. 1992. Modality-specific retrograde amnesia of fear. *Science* **256:** 675–677.

Kim, J.J., J.P. DeCola, J. Landeira-Fernandez, and M.S. Fanselow. 1991. N-methyl-D-aspartate receptor antagonist APV blocks acquisition but not expression of fear conditioning. *Behav. Neurosci.* **105:** 126–133.

Kim, J.K., R.A. Rison, and M.S. Fanselow. 1993. Effects of amygdala, hippocampus, and periaqueductal gray lesions on short- and long-term contextual fear. *Behav. Neurosci.* **107:** 1093–1098.

Lee, H. and J.J. Kim. 1998. Amygdalar NMDA receptors are critical for new fear learning in previously fear-conditioned rats. *J. Neurosci.* **18:** 8444–8454.

Lee, Y., D. Walker, and M. Davis. 1996. Lack of a temporal gradient of retrograde amnesia following NMDA-induced lesions of the basolateral amygdala assessed with the fear-potentiated startle paradigm. *Behav. Neurosci.* **110:** 836–839.

Logue, S.F., R. Paylor, and J.M. Wehner. 1997. Hippocampal lesions cause learning deficits in inbred mice in the Morris water maze and conditioned-fear task. *Behav. Neurosci.* **111:** 104–113.

Maren, S. 1998. Overtraining does not mitigate contextual fear conditioning deficits produced by neurotoxic lesions of the basolateral amygdala. *J. Neurosci.* **18:** 3088–3097.

———. 1999. Neurotoxic or electrolytic lesions of the ventral subiculum produce deficits in the acquisition and expression of Pavlovian fear conditioning in rats. *Behav. Neurosci.* **113:** 283–290.

Maren, S. and M.S. Fanselow. 1996. The amygdala and fear conditioning: Has the nut been cracked? *Neuron* **16:** 237–240.

———. 1997. Electrolytic lesions of the fimbria/fornix, dorsal hippocampus, or entorhinal cortex produce anterograde deficits in contextual fear conditioning in rats. *Neurobiol. Learn. Mem.* **67(2):** 142–149.

Maren, S., G. Aharonov, D.L. Stote, and M.S. Fanselow. 1996a. N-methyl-D-aspartate receptors in the basolateral amygdala are required for both acquisition and expression of conditional fear in rats. *Behav. Neurosci.* **110(6):** 1365–1374.

Maren, S., G. Aharonov, and M.S. Fanselow. 1996b. Retrograde abolition of conditional fear after excitotoxic lesions in the basolateral amygdala of rats: Absence of a temporal gradient. *Behav. Neurosci.* **110:** 718–726.

Maren, S., G. Aharonov, and M.S. Fanselow. 1997. Neurotoxic lesions of the dorsal hippocampus and Pavlovian fear conditioning in rats. *Behav. Brain Res.* **88:** 261–274.

Maren, S., S.G. Anagnostaras, and M.S. Fanselow. 1998. The startled seahorse: Is the hippocampus necessary for contextual fear conditioning? *Trends Cog. Sci.* **2:** 39–42.

McNish, K.A., J.C. Gewirtz, and M. Davis. 1997. Evidence of contextual fear after lesions of the hippocampus: A disruption of freezing but not fear-potentiated startle. *J. Neurosci.* **17:** 9353–9360.

Nadel, L. and J. Wilner. 1980. Context and conditioning: a place for space. *Physiol. Psychol.* **8:** 218–228.

Paylor, R., R. Tracy, J. Wehner, and R. Rudy. 1994. DBA/2 and C57BL/6 mice differ in contextual fear but not auditory fear conditioning. *Behav. Neurosci.* **108(4):** 810–817.

Phillips, R.G. and J.E. LeDoux. 1992. Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. *Behav. Neurosci.* **106:** 274–285.

Rescorla, R.A. 1968. Probability of shock in the presence and absence of CS in fear conditioning. *J. Comp. Physiol. Psychol.* **66:** 1–5.

Richmond, M.A., C.A. Murphy, B. Pouzet, P. Schmid, J.N. Rawlins, and J. Feldon. 1998. A computer controlled analysis of freezing behaviour. *J. Neurosci. Methods* **86:** 91–99.

Rogan, M.T., U.V. Stäubli, and J.E. LeDoux. 1997. Fear conditioning induces associative long-term potentiation in the amygdal. *Nature* **390:** 604–607.

Scoville, W.B. and B. Milner 1957. Loss of recent memory after hippocampal bilateral lesions. *J. Neurol. Neurosurg. Psych.* **20:** 11–12.

Silva, A.J., K.P. Giese, N.B. Fedorov, P.W. Frankland, and J.H. Kogan. 1998. Molecular, cellular, and neuroanatomical substrates of place learning. *Neurobiol. Learn. Mem.* **70:** 44–61.

Squire, L.R. 1992. Memory and the hippocampus; A synthesis from findings with rat, monkeys, and humans. *Psychol. Rev.* **99:** 195–231.

Valentinuzzi, V.S., D.E. Kolker, M.H. Vitaterna, S. Kazuhiro, A. Whitely, S. Low-Zeddies, F.W. Turek, E.A.M. Ferrari, R. Paylor, and J.S. Takahashi. 1998. Automated measurement of mouse freezing behavior and its use for quantitative trait locus analysis of contextual fear conditioning in (BALB/cJ x C57BL/6J)F2 mice. *Learn. & Mem.* **5:** 391–403.

Computer-Assisted Behavioral Assessment of Pavlovian Fear Conditioning in Mice

Stephan G. Anagnostaras, Sheena A. Josselyn, Paul W. Frankland, et al.

References	This article cites 39 articles, 8 of which can be accessed free at: **http://learnmem.cshlp.org/content/7/1/58.full.html#ref-list-1**	
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**	