Published in final edited form as:

Eur J Neurosci. 2012 February ; 35(3-4): 598–613. doi:10.1111/j.1460-9568.2011.07974.x.

Extinction reveals that primary sensory cortex predicts reinforcement outcome

Kasia M. Bieszczad and Norman M. Weinberger

Center for the Neurobiology of Learning and Memory, Department of Neurobiology and Behavior, University of California, Irvine, CA 92697-3800

Abstract

Primary sensory cortices are traditionally regarded as stimulus analyzers. However, studies of associative learning-induced plasticity in the primary auditory cortex (A1) indicate involvement in learning, memory and other cognitive processes. For example, the area of representation of a tone becomes larger for stronger auditory memories and the magnitude of area gain is proportional to the degree that a tone becomes behaviorally important. Here, we used extinction to investigate whether “behavioral importance” specifically reflects a sound’s ability to predict reinforcement (reward or punishment) vs. to predict any significant change in the meaning of a sound. If the former, then extinction should *reverse* area gains as the signal no longer predicts reinforcement. Rats ($n = 11$) were trained to bar-press to a signal tone (5.0 kHz) for water-rewards, to induce signal-specific area gains in A1. After subsequent withdrawal of reward, A1 was mapped to determine representational areas. Signal-specific area gains — estimated from a previously established brain–behavior quantitative function — were reversed, supporting the “reinforcement prediction” hypothesis. Area loss was specific to the signal tone vs. test tones, further indicating that withdrawal of reinforcement, rather than unreinforced tone presentation *per se*, was responsible for area loss. Importantly, the amount of area loss was correlated with the amount of extinction ($r = 0.82$, $p < 0.01$). These findings show that primary sensory cortical representation can encode behavioral importance as a signal’s value to predict reinforcement, and that the number of cells tuned to a stimulus can dictate its ability to command behavior.

Keywords

auditory cortex; instrumental conditioning; learning-induced plasticity; memory strength; rat

Introduction

The function of the primary auditory, somatosensory and visual cortices is traditionally assumed to be modality-specific stimulus analysis, based in part on their having tonotopic, somatotopic and retinotopic “maps”, respectively. However a broader conceptualization appears necessary because they also exhibit multimodal characteristics (Pascual-Leone & Hamilton, 2001) and cognitive functions (Rauschecker, 1999). Most extensively studied in primary auditory cortex (A1), the latter include selective attention, concept formation, expectancy, learning and memory, motor planning and goal-directed behaviors (e.g., Gonzalez-Lima & Scheich, 1986; Wetzel *et al.*, 1998; Villa *et al.*, 1999; Fritz *et al.*, 2003, 2007; Ghose & Bearl, 2010; Jaramillo & Zador, 2011).

Indeed, findings have led to alternative conceptualizations of A1 as a “semantic processor” (Scheich *et al.*, 2011) and “auditory problem-solver” (Weinberger, 2011). Both formulations have in common the idea that A1 does not simply analyze the physical parameters of sound, but also is deeply involved in the *interpretation of its meaning*.

Associative re-tuning of frequency representations is a major feature of learning-related plasticity in A1 throughout a lifetime. The best response of many cells shifts toward the frequency of an acoustic signal tone as that signal gains behavioral significance (Bakin & Weinberger, 1990; Edeline *et al.*, 1993; Gao & Suga, 2000; Kisley & Gerstein, 2001; Carpenter-Hyland *et al.*, 2010). Such frequency-tuning shifts have been linked to aspects of memory storage because they are associative, highly specific, consolidate (strengthen over time), may last indefinitely, develop in various tasks of both positive and negative valence and occur across species (reviewed in Weinberger, 2007). Determination of tonotopic maps after training reveals cortical reorganization that gains representational area for a signal tone both in animals (Recanzone *et al.*, 1993; Rutkowski & Weinberger, 2005; Polley *et al.*, 2006; Zhou & Merzenich, 2007; Hui *et al.*, 2009; Bieszczad & Weinberger, 2010b, 2010c; Takahashi *et al.*, 2010; Zhou *et al.*, 2010) and humans (Molchan *et al.*, 1994; Morris *et al.*, 1998; Menning *et al.*, 2000).

Nonetheless, actual function(s) of learning-induced tuning shifts and expansion of specific cortical representations are largely unknown. They may encode level of behavioral importance because the greater the importance, the greater the cortical expansion (Rutkowski & Weinberger, 2005). However, explication of “importance” is needed because stimuli can become significant for many reasons. For example, stimuli may gain salience when they *predict reinforcement* (i.e., any reward or any punishment). Otherwise, they could gain behavioral significance if they predict a *relevant change* in the environment, regardless of whether it is reinforcement or the withdrawal of expected reinforcement. These alternatives to explicate stimulus-significance can be tested by the use of experimental extinction (withdrawal of reinforcement) following acquisition of the tone-cued reward of an instrumental response. If expanded representation serves to predict reinforcement, then it should be reversed by extinction. However, if it is a sign of any relevant change in auditory predictions of the state of affairs, then extinction should *not* eliminate cortical expansion because the absence of expected reward following response to an acoustic signal is itself a behaviorally significant event.

Materials and Methods

Subjects

The subjects were eleven male Sprague–Dawley rats (300–325 g, Charles River Laboratories; Wilmington, MA), housed in individual cages in a temperature controlled (22 °C) vivarium and maintained on a 12/12 h light/dark cycle (lights on at 6:00 AM) with free access to food and water before training. They were handled daily and retained in the vivarium for a minimum of one week prior to any treatments. Water was restricted immediately before training began with water supplements to maintain body weight to ~15% below non-restricted littermates (mean experimental group body weight throughout training was 85.5% (± 1.5) of *ad libitum* littermates). All experiments were carried out with measures to minimize animal pain and discomfort with procedures approved by the University of California Irvine’s Institutional Animal Care and Use Committee (IACUC) and in accordance with the National Institutes of Health’s (NIH) Office of Laboratory Animal Welfare guidelines.

Training apparatus and stimuli

The training apparatus was the same as that used previously (Bieszczad & Weinberger, 2010b). Briefly, training was conducted in a sound-attenuated instrumental conditioning chamber (H10-11R, Coulbourn Instruments, Whitehall, PA) fitted on one wall with a bar manipulandum, a water cup attached to a lever (H14-05R, Coulbourn) that could deliver 0.02 ml of water to an opening 9 cm to the left of the bar (H21-03R, Coulbourn), a speaker (H12-01R, Coulbourn) 13 cm above this opening, and an overhead house light (H11-01R, Coulbourn). The chamber was enclosed in a sound-attenuating chamber (H10-24A Coulbourn).

All tone stimuli were generated using a Tucker–Davis Technologies (TDT, Alachua, FL) System 3 (RP2.1 Enhanced Real-Time Processor) and RpvdsEx software. Tones were always 10 s in duration (70 dB SPL) and cosine-squared gated with rise/fall time [10–90%] of 10 ms. Tone levels for all frequencies used in training were calibrated for three locations in the training chamber at animal head height and set to average 70 (± 2) dB SPL.

Experimental approach

The experiment consisted of three behavioral phases: (1) *acquisition* of a tone-cued bar-pressing task to obtain water reward over a period of up to 17 days; (2) *extinction* of acquisition by omission of rewards in the same tone-cued bar-pressing task for 3 days; and (3) a *spontaneous recovery* test 48 h after the end of extinction (Fig. 1).

However, determination of the effects of extinction on the expanded representation of a signal frequency in A1 entails unique problems. For example, to determine the extent of acquisition-induced change in representational area, ideally, one would obtain frequency maps before and after acquisition, and to determine the extinction-induced change, a third time after extinction.

Unfortunately, complete and fine-grain tonotopic maps can be obtained *only once in a subject*. This limitation is due to the fact that only comprehensive maps of the primary auditory cortex are acceptable for determining learning-induced changes in cortical representational area. These require a very large number (e.g., 60–80) of microelectrode penetrations of the cortex (e.g., Rose & Woolsey, 1949; Downman *et al.*, 1960; Merzenich *et al.*, 1975; Knight, 1977; Reale & Imig, 1980; Sally & Kelly, 1988; Thomas *et al.*, 1993; Harrison *et al.*, 1996; Stiebler *et al.*, 1997; Heil & Irvine, 1998; Linden *et al.*, 2003; Polley *et al.*, 2007; Hui *et al.*, 2009; Hackett *et al.*, 2011), which might interfere with subsequent normal cortical function.

Nonetheless, to investigate the effect of extinction on cortical area of frequency representation in A1 requires two elements: (a) determination of area changes after acquisition and (b) after extinction. A typical approach to the former is the use of between-groups designs in which maps obtained in the trained group after acquisition are compared to a suitable untrained control group. Indeed, this approach has revealed signal-specific expanded representations in A1, starting with Recanzone *et al.* (1993), in which maps in trained primates were compared to those from naïve controls.

The second is a more difficult problem, because determination of the effects of extinction requires that post-extinction maps be compared to pre-extinction (i.e., post-acquisition) maps. However, one cannot obtain the latter directly from the same animals because of the constraint of being able to obtain the map only once per subject. A solution to this problem could be provided if one could obtain a measure of the amount of tonotopic area gain in the post-acquisition/pre-extinction map without the need for the direct electrophysiological mapping of A1. This could be attained if a quantitative relationship between acquisition

performance and the amount of A1 area-gain were known. Such a correlation could be applied to make it possible to use the acquisition performance of the current subjects to *interpolate* an estimate of the amount of area expansion post-acquisition. These interpolated values could later be compared with the A1 area of representation directly obtained by electrophysiological measurement following extinction training in the same subject to determine area changes after extinction. Fortunately, this type of behavior–brain relationship has been established from previous research for the task used in the current study. It is

$$y=0.126x+0.083$$

EQ.
1

where y is the percent of A1 area representing the signal frequency, and x is the TOTE Learning Index (TLI) [see below, *TOTE Learning Index (TLI)*] (Fig. 2).

The TLI quantifies the extent to which rats use a “tone-onset-to-error” (TOTE) learning strategy to solve the problem: animals start to bar-press at tone onset and continue responding, ignoring tone-offset, until they receive an error signal/time-out for a bar-press during the silent inter-trial interval (ITI). The basis for the critical role of TOTE in the formation of specific representational gain in A1 is not yet known. It may reflect the fact that the TOTE strategy emphasizes tone onsets while ignoring tone offsets, and cells in A1 are particularly sensitive to onset transients (Heil, 1997a, 1997b; Phillips *et al.*, 2002; see also Wang *et al.*, 2005; Hoshino, 2007). In any event, the relationship does provide the basis for estimating the amount of area gain during acquisition given the level of TOTE behavioral performance. Therefore, the estimated area gain of the signal frequency in the current subjects could be determined by interpolation from this function, using the TLI for each subject’s behavior (x) to calculate the presumptive individual amounts of signal-specific A1 area (y) after the acquisition of this particular tone-cued instrumental task.

Behavioral Training and Testing

Acquisition—The experimental protocol is illustrated in Fig. 1. Acquisition training was identical to that used previously to induce signal-specific expansion (Experiment 1, Fig. 7 in Bieszczad & Weinberger, 2010b). Prior to any training, animals were shaped to bar-press for water reward on a free operant schedule (1:1 reward ratio) for 5 days. Next, they were trained to bar-press in the presence of a 5.0 kHz tone (10 s, 70 dB SPL) for water reward in four sessions using a short inter-trial interval (ITI, mean = 8 s; range = 4–12 s). The ITI was increased to a mean of 15 s (range = 5–25 s) for all sessions thereafter. Sessions were 45 min long, which did not permit subjects to become satiated and thus reduce responding (Fig. 1, “Acquisition”).

Prior studies have revealed that the development of specific representational plasticity in A1 depends upon the use of a learning strategy based on attending to tone onset and ignoring tone offset called “tone-onset-to-error” or TOTE (Berlau & Weinberger, 2008; Bieszczad & Weinberger, 2010a; 2010b). Use of the TOTE learning strategy was promoted by continuing the rewarded period beyond the tone’s offset. That is, the first bar-press to occur within a “Free Period” that could last up to 7 s after tone-offset was also rewarded. This additional rewarded bar-press occurred provided that an animal had bar-pressed to receive at least one reward during the preceding tone. Then, the first bar-press following the maximum 17 s reward period (10 s tone + ≤ 7 s Free Period) was punished by a flashing light and a time-out period. Time-out periods increased the time until the next trial by 3 s during the first four days of training or by 7 s for all subsequent training sessions. Bar-presses after the first time-out period during the remaining inter-trial periods, were punished 50% of the time (randomly programmed).

Successful performance was calculated as the proportion of rewarded bar-presses (BP) made during the tone, relative to the sum of all bar-presses during the tone and during inter-trial intervals. Rewarded bar-presses during the Free Periods were not included in the calculation of performance:

$$\text{Performance} = \frac{\# \text{ Tone BPs}}{\# \text{ Tone BPs} + \# \text{ ITI BPs}} \times 100\% \quad \text{EQ. 2}$$

Subjects were trained to asymptote, defined as a coefficient of variation (CV = standard deviation/mean) in performance across 3 consecutive days of ≤ 0.10 .

Extinction—The purpose of extinction training was to determine if expanded representation of a signal frequency was behaviorally important, either because it predicted reinforcement or because it predicted any relevant change in acoustic signaling, e.g., the absence of expected reward. If expanded representation encodes reinforcement prediction, then it should be reversed by extinction. But if it encodes any relevant change in the meaning of a sound, then extinction should *not* eliminate cortical expansions because the withdrawal of expected reward is a significant event, *viz.*, the training tone would take on a new meaning of a changed outcome and thus remain behaviorally significant.

However, to interpret plasticity in A1 frequency maps obtained after extinction, it is necessary to know if subjects actually behaviorally extinguished responses to the signal tone. If behavioral extinction were never achieved, then the effect on A1 tonotopy would not be determinable; as a result, the two alternative hypotheses would not have been tested. Additionally, behavioral extinction needed to be evaluated for its *specificity* to the signal frequency, because behavioral extinction might actually simply reflect an equal loss of response to all frequencies (i.e., behavioral extinction to tones in general). Without *signal-specific* extinction, the absence of an effect on *signal-specific* A1 representation would likewise fail conclusive interpretation. Therefore, extinction training included unreinforced presentation of additional generalization-test frequencies as well as the signal frequency, to address both points by determining (1) the *degree* and (2) the *specificity* of behavioral extinction.

Extinction training began the day after a subject reached asymptote of acquisition. To insure that performance was still at a high level before the start of extinction, the first session began with 10 standard trials on which reward was given for bar-presses during the signal tone (Fig. 1, “TOTE Test”). This was followed without break by pseudo-random ordered presentations of the signal tone (5.0 kHz) and four generalization test tones (1.2, 2.4, 10.3 and 21.2 kHz). Test tone frequencies were spectrally distant from the signal tone and their neighbors and were separated by more than an octave (i.e., by 1,300 cents) to avoid octave generalization. Each of the five tones (i.e., four generalization test tones and the signal tone) was presented without reinforcement 20 times for a total of 100 pseudo-randomly intermixed trials in an extinction session. Two additional identical extinction sessions also of 100 trials each were presented on the following days to promote behavioral extinction (Fig. 1, “Extinction”).

Extinction is not erasure of memory, but rather inhibitory learning that opposes prior excitatory learning (Pavlov, 1927; Konorski, 1948). Therefore, a reduction in bar-presses during a tone is an indication of (1) the *degree* to which inhibitory learning successfully opposes prior excitatory learning, i.e., that the *prediction of reinforcement* is weakened. An index of the strength of extinction has at its foundation the *number* of bar-presses to the unrewarded signal: the fewer the responses, the stronger the extinction. However, this alone

cannot quantify *signal-specific extinction* because subjects might also behave the same way to generalization test tones. That is, animals that respond in equal numbers to the signal and generalization test tones do not have behavioral specificity for the signal-frequency. Therefore, to provide a valid estimate of (2) the frequency *specificity* of signal extinction, it was necessary to determine the extent to which bar-presses to the signal tone (Signal BPs) differed from a flat (equal) distribution of the bar-presses to all five tones (Total BPs divided by 5). Thus all bar-presses to both the signal and generalization test tones were taken into account to determine:

$$Specificity = \frac{\left[\# \text{Signal BPs} - \left(\frac{\# \text{Total BPs}}{5} \right) \right]}{\left(\frac{\# \text{Total BPs}}{5} \right)} \quad \text{EQ. 3}$$

For example, if an animal bar-pressed 10 times to each of the tones, then $Specificity = [10 - (50/5)]/(50/5) = 0$. Thus, a value = 0.0 would indicate a complete lack of specificity. In contrast, if an animal bar-pressed 10 times *only to the signal tone*, then $Specificity = [10 - (10/5)]/(10/5) = 4$. Thus, a value = 4.0 indicates complete selectivity for the signal.

An index for behavioral “signal-specific responses in extinction” (SSRE), combined the degree (1) and specificity (2) of bar-presses in extinction. SSRE was the *number of bar-presses* to the signal tone, adjusted by the *specificity of bar-presses* for the signal frequency:

$$SSRE = \# \text{Signal BPs} \times Specificity \quad \text{(EQ. 3)} \quad \text{EQ. 4}$$

For example, if an animal bar-pressed 10 times to each of the five test tones, then $Specificity = 0.0$ and $SSRE = 10 \times 0.0 = 0$. If an animal bar-pressed 10 times only to the signal tone, then $Specificity = 4.0$ and $SSRE = 10 \times 4.0 = 40$. Likewise, this index could also distinguish animals that behaved with the same frequency-specificity but with different degrees in the absolute number of signal bar-presses. For example, if an animal bar-pressed as many as 30 times and only to the signal tone, then $Specificity = 4.0$ and $SSRE = 30 \times 4.0 = 120$.

Spontaneous recovery and final strength of behavioral extinction—Extinction can be followed by spontaneous recovery (Pavlov, 1927). Therefore, quantification of the *final* strength of extinction must, in addition to degree and specificity, account for the amount of *spontaneous recovery*. It is possible that there be partial recovery of response to the signal, no recovery, or more pronounced extinction. Animals were tested for spontaneous recovery after a 2-day delay following the last (3rd) extinction session. The protocol was identical to extinction sessions, i.e., signal and generalization test tones without rewards (Fig. 1, “Spontaneous Recovery”). The degree and specificity of spontaneous recovery was determined by the SSRE-index as it was for extinction.

The “Final Strength of Extinction” was determined by the magnitude of response decrement during extinction adjusted by the magnitude of spontaneous recovery. The magnitude of response decrement during extinction training was calculated as the difference in SSRE between the 1st and 3rd extinction sessions (i.e., Ext Day 3 – Ext Day 1), because these sessions provided the first and last sessions of extinction training data. The magnitude of spontaneous recovery was calculated to be the value of SSRE on the 3rd extinction session subtracted from the SSRE value for the spontaneous recovery session, because any spontaneous recovery occurs from a baseline of the behavioral response during the immediately preceding session, i.e., Extinction Day 3. Thus:

$$\text{Final Strength of Extinction} = \left[\frac{(\text{Spontaneous Recovery}_{\text{SSRE}} - \text{Extinction Day } 3_{\text{SSRE}})}{(\text{Extinction Day } 3_{\text{SSRE}} - \text{Extinction Day } 1_{\text{SSRE}})} \right] \quad \text{EQ. 5}$$

Acoustic stimulation and electrophysiological recording

A1 was mapped in a terminal session 24–48 h after the test for spontaneous recovery. An additional group of untrained naïve animals ($n = 9$) had been mapped as a comparison group. Both the mapping methods and the results for the naïve group are standard, identical to the current experimental group, and were reported previously (Bieszczad & Weinberger, 2010a; 2010b). Briefly, rats were anesthetized with sodium pentobarbital to minimize pain and discomfort (50 mg/kg, i.p.). After a craniotomy and removal of dura mater over the temporal cortex, photographs of the cortical surface were taken to record the position of each microelectrode recording site. Thereafter, calibrated acoustic stimuli were delivered to the contralateral ear with the speaker placed 2–3 cm from the ear canal. The stimuli consisted of broadband noise (bandwidth = 1 kHz–50 kHz, 0–80 dB SPL in 10 dB increments, 20 repetitions) and pure tone bursts (50 ms, cosine-squared gate with rise/fall time [10–90%] of 7 ms, 0.5–54.0 kHz in quarter-octave steps, 0–80 dB SPL in 10 dB increments). Stimuli were presented once every 700 ms with noise- or frequency-level combinations pseudo-randomized by TDT System 3 software. Frequency response areas (FRA) were obtained at each cortical recording site using 10 repetitions of the frequency/level-stimulus set (252 stimuli in total).

All maps were determined identically by extracellular recordings of multi-unit clusters made with a linear array of 4 parylene-coated microelectrodes (0.2–2.0 MΩ, FHC, Bowdoin, ME) that were lowered to layers III–IV, perpendicular to the surface of the cortex (400–600 μm deep) via a stepping microdrive (Burleigh Inchworm). The density of recordings across A1 was kept constant in all animals (250–500 μm apart while avoiding blood vessels). Sampling was in all areas of the exposed cortex that reliably elicited acoustic tone-driven neural responses until the physiologically-defined borders of A1 were reached (see below).

Neural activity was amplified (1000×), band-pass filtered (0.3–3.0 kHz, TDT RA16 Medusa Base Station) and monitored on a computer screen and audio-monitor system (Grass AM8). Only discharges having > 2:1 ratio were included in analyses. Responses to noise bursts were recorded before tone-stimuli were presented. The neural responses to noise were later compared with responses to tones at each site as evidence for the borders of A1. Standard physiological criteria used to determine the borders and total size of A1 were to include those sites with lower response thresholds for tones compared to noise and having a caudal-to-rostral, low-to-high frequency tonotopic organization (e.g., Rutkowski *et al.*, 2003; Polley *et al.*, 2007). Complete mapping of A1 generally required 60–80 penetrations over a period of 8–12 h (Fig. 1, “Electrophysiology”).

Determination of representational frequency area

Frequency response areas (FRAs, noted above) were constructed offline for evoked spike-timing data using custom Matlab R2009a software. Tone-evoked discharges during a response time window that would capture A1 onset responses (6–40 ms after tone onset) were determined for each stimulus presentation. The evoked spike rate was defined as a response in the onset time window that was greater than $2 \times \text{s.e.m.}$ above the mean spontaneous rate determined from a 50 ms period immediately preceding the presentation of a tone.

The characteristic frequency (CF) of a responsive site was defined as the stimulus frequency having the lowest threshold (CF threshold) for an evoked response (i.e., highest neural sensitivity). CFs were determined from FRAs constructed for each recording site. All recording sites were identified according to their Cartesian coordinates across the surface of the cortex. Each coordinate point (i.e., that references each recording site) was labeled with the determined CF and entered into a Voronoi tessellation algorithm (in Matlab R2009a) to construct areal polygons around each point. The sizes of polygons were summed within CF half-octaves (1.1–1.5, 1.6–2.0, 2.1–3.0, 3.1–4.0, 4.1–6.0 (tone signal-bin), 6.1–8.0, 8.1–12.0, 12.1–16.0, 16.1–24.0 and 24.0–32.0 kHz) to represent the areal distribution of frequency. To account for natural variability in the absolute size and orientation of topographic maps (Merzenich *et al.*, 1975, 1987), all half-octave map areas were expressed in relative terms as percentages with respect to the total area (100%) of A1. The amounts of A1 area occupied by each half-octave band within individual animals were then averaged ($\pm$ s.e.m.) to determine the distribution of areal representation of frequency in A1 for the group.

Estimation of area gained after acquisition training

As this study had to obtain electrophysiological cortical maps after extinction training, the limitation of being able to obtain only one map within a subject required an estimate of the area of gain after acquisition training. This was obtained using a quantitative relationship between behavior and signal-specific area of representation in A1 from a previous study (Bieszczad & Weinberger, 2010b). Specifically, area varies with the degree to which subjects use a particular learning strategy to perform the task (Fig. 2). The incidence of this “tone-onset-to-error” (TOTE) learning strategy is significantly positively and linearly correlated with the amount of signal-specific representational area gain in A1 for the frequency of the tone ($5.0 \text{ kHz} \pm 0.25$ octaves) after acquisition.

The previously established brain–behavior correlation is computed with a “TOTE Learning Index” (TLI) and described by the relationship in (EQ. 1), where y is the percent of A1 area representing the signal frequency of $5.0 \text{ kHz} (\pm 0.25 \text{ octaves})$, and x is the TOTE Learning Index, i.e., TLI (from Bieszczad & Weinberger, 2010b) (Fig. 2). That the y -intercept of this function is at the level of the naïve area of signal-representation also suggests its validity to predict gains in area, i.e., as enlarged from the naïve state, induced by learning. The estimated area gain of the signal frequency in the current subjects could be determined by interpolation from this function, using the TLI per each subject’s behavior (x) to calculate the presumptive individual gains in signal-specific area (y) accrued after the acquisition of this particular tone-cued appetitive instrumental task.

Each subject’s estimated area was later used as the “baseline” or “pre”-extinction condition to determine the change in A1 representation induced by subsequent extinction, i.e., revealed by comparing this to the “post”-extinction condition determined electrophysiologically. There was no significant difference in the *total size* of A1 between groups of animals that had learned to bar-press to tones for rewards before extinction (prior study without extinction training: $5.58 \text{ mm}^2 \pm 1.62$; Bieszczad & Weinberger, 2010b) and after extinction (current group with extinction training: $6.26 \text{ mm}^2 \pm 1.35$; $t_{16} = 0.97$, $p > 0.05$). Therefore, any change in the relative area of signal-representation after extinction is likely due to the reorganization of frequency representation within A1 and not an absolute enlargement or shrinkage of the size of A1. Thus, the difference between the estimated area gain after association between signal and reward and the area determined neurophysiologically after extinction revealed *representational plasticity induced by extinction-learning*. This analysis identified the magnitude of extinction-induced reversal of the acquisition-induced area gain.

TOTE Learning Index (TLI)

The incidence of TOTE has been reliably quantified by computing a stringent TOTE-learning index (TLI) from the pattern of responses during a single test session administered after acquisition (Fig. 2). Tone trials in the TOTE test session included rewards only for those bar-presses that were made during the 10 s tone, and delivered errors for all bar-presses made during silences between tone presentations (Fig. 1, “TOTE Test”). That is, trials did not include “Free” periods. TLI test trials were such so they could distinguish a true TOTE pattern of behavior, i.e., instrumental responses that begin at tone onset and continue only until a response generates an error signal during the inter-trial period, from patterns of behavior that would reflect alternative strategies. Alternative strategies would be evident as patterns in which animals might begin responding at tone-onset but stop *before* a response during silence would elicit the error signal. This indicates a learning strategy that uses the tone’s offset to cue the end of the reward period and to stop responding, or one in which animals effectively “count” the expected maximum number of rewards per tone-trial. Thus for each animal, the incidence of TOTE in the pattern of responding to tones for rewards was calculated by TLI identically as in a previous report (Bieszczad & Weinberger, 2010b). TLI is defined as the number of trials on which subjects bar-pressed during the tone and stopped after a single bar-press within 5 s of the tone’s offset resulted in an error signal. A 5 s window beginning at tone offset was used in this assessment of TOTE strategy because the typical latency to bar-press after tone offset was < 5 s throughout training (i.e., shorter than the duration of the Free Period) (Bieszczad & Weinberger, 2010b). It is calculated as: $TLI = \# \text{ trials with TOTE behavioral pattern} / 10 \text{ test trials}$. Values range from 0.0 to 1.0, with 1.0 being the strongest use of the “tone-onset-to-error” strategy.

Statistics

All performance parameters were analyzed using repeated measures ANOVA ($\alpha = 0.05$) for analyses across training sessions with *post hoc* Holm–Sidak tests. Correlations were computed using Pearson’s *r* statistic. These analyses and all descriptive statistics were executed using Matlab R2010a statistical packages.

Results

Behavioral acquisition and gains of signal-specific representation in A1

Rats learned to bar-press to the 5.0 kHz tone to obtain water reward (Fig. 1 “Acquisition”). Group performance increased to an asymptotic level of $56.9 \pm 3.2\%$ ($F_{16,160} = 39.25$, $p < 0.01$) achieved in 15.6 ± 1.2 days (Fig. 3A). This asymptotic level is consistent with the marked difficulty of the task, caused by the ambiguity of errors, as only 50% of ITI errors received an error-signaled time-out period.

Estimates of cortical area gain were obtained by interpolation from a previously established correlate with behavior. The rationale for this approach was based on the facts that (a) the animals were trained in the exact same protocol in which this relationship had been discovered (Bieszczad & Weinberger, 2010b), and (b) the acquisition functions of these groups were virtually identical ($F_{16,256} = 1.60$, $p > 0.05$; Fig. 3B, inset) and the number of training sessions to reach asymptote were not significantly different (15.6 ± 1.2 in the current study and 25.9 ± 7.3 in the previous; $t_{(16)} = 1.74$, $p > 0.10$). As in the previous study (Bieszczad & Weinberger, 2010b), the TOTE-Learning Index (TLI; Methods, *Quantification of the TOTE Learning Index*) was determined during a test session after the acquisition of the task (Fig. 1, “TOTE Test”). Thus, TLI values obtained for each animal were used with (EQ. 1) to estimate the amount of area gain after acquisition training (Fig. 3B, closed diamonds).

Next, to infer the amount of cortical expansion of the 5.0 kHz signal frequency (± 0.25 octaves) in A1, we compared the interpolated area value for the current group with the area for the same frequency range in naïve animals. The mean naïve representation of the signal frequency constitutes 7.7% (± 1.6) of the tonotopic map in A1 (reported in Bieszczad & Weinberger, 2010b). The mean estimated area of signal-representation for the current acquisition group was 12.1% (± 3.2), which indicates a significant cortical expansion ($t_{18} = 3.58$, $p < 0.001$) (Fig. 3B; see also Fig. 6A). As training was sufficient to produce the expected expansion of area for the signal frequency (5.0 kHz), the next step was to determine the effects of subsequent *extinction* on these putative area gains.

Specificity of behavioral extinction

Standard investigations of behavioral extinction simply eliminate reinforcement. However, this is insufficient to also determine the *frequency-specificity* of extinction. Thus, we also included four novel frequencies (Fig. 1, “Extinction”). Therefore, extinction training consisted of unrewarded presentations of the signal tone (5.0 kHz) and four randomly intermixed flanking generalization test tones (1.2, 2.4, 10.3 and 21.2 kHz).

All subjects responded at least once to each tone frequency during the first extinction session, which enabled tracking of the magnitude of change in behavioral responses for different frequencies throughout the further two days of extinction training and also during the spontaneous recovery test. The number of responses across extinction sessions was greatest for the signal-frequency, with overall fewer responses to tone frequencies that were more distant from the signal, ($F_{4,98} = 5.97$, $p < 0.001$; Fig. 4). Therefore the trained group showed evidence of specificity that the frequency of the signal-tone *predicted the availability of reinforcement*.

A value for the “signal-specificity of response in extinction” (SSRE), was used to quantify the signal-specificity of behavioral responses for each animal throughout extinction and spontaneous recovery. A greater degree and higher specificity of responses to the signal was indicated by greater SSRE values. Thus, greater SSRE indicated whether animals behaved as though the frequency of the signal had a greater strength of prediction for the availability of reinforcement than other frequencies.

Fig. 5 provides examples of SSRE values for each of the three days of extinction training and also for the spontaneous recovery test. Individual patterns could be discerned. For example, on the first day of extinction, the strength of prediction could be greatest to the signal tone (Fig. 5B) or for a novel generalization test tone (Fig. 5A, 5C). However, there was a common pattern across individuals for SSRE values to decline over the three days of extinction training, i.e., the SSRE values were lower on the last day of extinction than on the first. Next, during the spontaneous recovery test, this trend could reverse, i.e., the SSRE values could be greater in spontaneous recovery (Fig. 1, “Spontaneous Recovery”) than on the last day of extinction. This general pattern was observed over a wide range of SSRE values, e.g., on extinction day 1 ranging from 0.9 (the signal’s lower reinforcement prediction, Fig. 5C) to 35.6 (the signal’s higher reinforcement prediction, Fig. 5B).

To investigate the relationship between the strength of final signal-specific extinction and cortical area required accounting for spontaneous recovery. This was quantified by calculating the change in SSRE values after the 2-day recovery period, relative to the change in SSRE values over the extinction phase of training (Methods, *Spontaneous recovery and overall strength of extinction*; Fig. 1, “Spontaneous Recovery”). Thus, we determined “Final Strength of Extinction”, defined as the amount of behavioral extinction to the signal, adjusted for spontaneous recovery.

Table 1 presents individual SSRE values across extinction training (1st Day, Column A; 3rd Day, Column B), spontaneous recovery (Column C) and the resultant Final Strength of Extinction (Column D). Overall, this table illustrates individual variability in extinction learning, as is well known for acquisition learning. Some animals developed very strong extinction (e.g., subject Amb10: Final Strength of Extinction was – 113%, i.e., the response to the tone during the spontaneous recovery test was *further reduced* by 113% of the magnitude of change in response across the sessions of extinction training); while others developed no extinction (e.g., Amb15: Final Strength of Extinction was +297%, i.e., the response to the tone during the spontaneous recovery test was actually *increased* by 297%). Actually, there were four cases of robust spontaneous recovery in which extinction training failed to significantly reduce behavioral responses to the signal (positive values for Final Strength of Extinction in Table 1, Column D).

Given quantification of the strength of extinction, it was possible to determine the relationship of extinction strength to the magnitude of representational plasticity of area in A1 following extinction training. This was achieved by comparing individual magnitudes of change in the representational area of the signal-frequency in A1 with respective individual indices of extinction strength.

Extinction-induced effects on representation of frequency area in A1

To determine whether extinction had induced representational plasticity in A1, we compared tonotopic maps obtained after extinction with the estimated size of signal-area after acquisition (e.g., Fig. 3B). If extinction had induced representational plasticity that was specific to the extinguished signal tone, then the previously established cortical expansion of the signal-frequency should be reduced to the naïve level. Indeed, in the group, extinction induced a reversal to the naïve level for the signal frequency band ($F_{1,19} = 1.36$, $p > 0.05$) (Fig. 6A). Signal-specific area was reduced from the interpolated group mean of 12.1% (± 3.2) of signal-area in A1 after acquisition to 7.5% (± 3.7) of signal-area in A1 after extinction.

In contrast to the loss of area gain for the *signal* frequency, there was no reduction in area of any of the generalization test frequencies: e.g., areas for characteristic frequencies (CF) ranging from 1.1 to 32.0 kHz were not significantly reduced (all $F_{1,19}$ ratios, $p > 0.05$) (Fig. 6A). Indeed, the extinguished group did not differ from the naïve group in any frequency-band (CF Band $\times$ Group: $F_{9,199} = 1.52$, $p > 0.05$). Together, these findings indicate that a reduction in frequency-area was specifically induced *only* for the frequency that had been previously associated with reward.

In most subjects (7/11), the signal frequency's percentage of the total area of A1 was not different from the naïve level of 7.7% (± 1.6) (Z-score relative to mean of naïve group: $z < 1.80$, $p > 0.05$). However, in one animal, signal area was not reduced by extinction training, i.e., the estimated area gain was actually larger (1/11; $z = 4.87$, $p < 0.05$). In the remaining three subjects, signal-area of representation was remarkably reduced to significantly *less* than that of naïves (3/11; $z > 1.98$, $p < 0.05$).

Loss of cortical expansions is correlated with strength of extinction

Extinction training provides a method to explicate the meaning of “behavioral importance” for area expansion. In addition, the data provide some insight into the process of extinction itself. In particular, might area loss be a mechanism to help explain the strength of behavioral extinction? To address this issue, we compared the strength of extinction with the amount of area loss in A1 within subjects. To determine the *amount of signal-area loss* in each subject, the post-extinction mapped area was subtracted from the estimated area of gain

(Fig. 3B). In 10 of 11 subjects, extinction training produced an estimated specific decrease in the signal's representation (Table 2). Importantly, the loss in area was likely due to an active process of extinction and not merely a passive dissipation of areagains over time. If area losses had occurred simply with the passage of time and independently of extinction training, then the signal-specific behavior four days after the last acquisition session, i.e., during the spontaneous recovery session, might correlate with the specific loss in A1 area. However, this was not the case; the signal-specific response in extinction (SSRE) during the spontaneous recovery session was not significantly correlated with the amount of area loss ($r = 0.26$; $p > 0.05$). In contrast, the area loss was significantly correlated with the "Final Strength of Extinction", which takes into account the extinction behavior [EQ. 5]: the stronger the extinction, the greater the loss ($r = 0.82$, $p < 0.01$). This correlation was not driven by extreme values because elimination of either one or both of the lowest and highest values left the correlations still statistically significant (lowest point removed: $r = 0.71$, $p < 0.01$; highest point removed: $r = 0.80$, $p < 0.01$; both highest and lowest points removed: $r = 0.68$, $p < 0.05$) (Fig. 7). In contrast to the signal-tone, there were no significant correlations for any of the generalization test frequencies (1.2 kHz: $r = 0.64$; 2.4 kHz: $r = 0.24$; 10.3 kHz: $r = 0.29$; and 21.2 kHz: $r = 0.18$; all $p > 0.05$).

Therefore, the relationship between a frequency-specific cortical change and a frequency-specific behavioral change holds only for the signal frequency that had previously acquired behavioral significance.

Discussion

Summary of the Findings

The goal of this study is to gain a better understanding of expanded representations of an acoustic signal in the primary auditory cortex during learning of an instrumental auditory task. Prior studies supported the hypothesis that specific representational cortical expansion is an increasing function of the behavioral importance of a signal tone. The present study sought to distinguish between two hypotheses to better explicate the meaning of "importance". Signal representation might expand because a sensory stimulus (a) predicts reinforcement (reward or punishment) or (b) predicts any behaviorally relevant state of affairs, regardless of whether it is a reinforcement or the withdrawal of an expected reinforcement. These alternatives were tested by the use of experimental extinction. If expanded representations predict reinforcement, then they should be reversed by extinction. However, if they indicate any important event, then extinction should not eliminate cortical expansions because the absence of expected reward itself is a behaviorally significant event.

Use of extinction training after acquisition of an auditory-cued task revealed no significant group difference in the area of representation of the signal frequency between the trained group and a naïve control group (Fig. 6A). This finding indicates a reversal of prior acquisition-based area gain, dependent upon the validity of estimates of such gains. As such, this result supports the "reinforcement prediction" hypothesis of expanded representation rather than the "significant change" account.

Additionally, the results are informative about the processes of extinction itself, based on three further findings. First, the loss of area in A1 was specific to the frequency of the signal (± 0.25 octaves); generalization test tones suffered no loss of area. Second, the behavioral strength of extinction was significantly correlated with the amount of signal-specific loss of estimated prior gain in cortical area: the stronger the extinction, the more the area was reduced (Fig. 7). (This relationship extended even to the area *gain* of one animal (Table 2, *Amb15*; Fig. 7, point "a").) Third, extinction could have greater effects than reversing

acquisition-based area gains: three animals had signal representational areas that were significantly *smaller* than naïves (Table 2, *Amb16*, *18* & *19*).

Validity of the Findings

Before further consideration of the findings and their potential significance, it is necessary to consider their validity. There are three major issues: (a) the use of estimated areas after acquisition, (b) the use of tonotopic maps and (c) determination of the strength of extinction.

Estimated Areas of Representational Gain—Current technology does not permit repeated within-subject comprehensive and fine-grain mapping of cortical representations of acoustic frequency in the auditory cortex. While optical and other alternatives to extensive sampling using acute (terminal) microelectrode recording have been attempted, they have yielded insufficiently precise mapping information because of the inability to obtain complete intensity (including threshold level) information and poor spatial (characteristic frequency) resolution (e.g., Bakin *et al.*, 1996; Tsytarev *et al.*, 2004). Therefore, while it was essential to obtain maps after extinction training to study its effects, it was impossible to obtain maps after acquisition in the same subjects. Rather, it was necessary to *estimate* the degree of signal-specific expansion of area following acquisition. The relationship between behavior and area gain had been determined previously (Bieszczad & Weinberger, 2010b) (Fig. 2), and so was used in the present to obtain by interpolation the estimated area of signal ($5.0 \text{ kHz} \pm 0.25 \text{ octaves}$) representation in each subject (Fig. 3B). The validity of the present findings of extinction-based reduction of acquisition-based expansion depends critically upon the estimated gain in area following acquisition.

There may be some arguments for rejecting estimates of area based on prior findings. It might be that given the complexity of the training protocol, plus individual animal variability, the prior study (Bieszczad & Weinberger, 2010b) and the current experiment are too different to use data from the first to estimate area gains in the second. However, the protocols were identical and they produced the same acquisition functions (Fig. 3B). A related concern may be that one cannot conclude there was a change in area unless direct measures are obtained before and after extinction within the *same* animal. While such before/after measures are desirable, within-subjects comparisons are often not possible and between-groups designs also are valid. The latter include comparisons between past and present findings, the basis for cumulative knowledge. Unusual is that the prior acquisition study, which is the basis for the current experiment, provides a quantitative, rather than qualitative, relationship between brain and behavior in a learning/memory context. As such quantitative relationships become more frequent, their acceptance is likely to grow as in other domains of science.

A cursory consideration of the findings might suggest that as the training group did not differ from the naïve group in representational area of the signal frequency (or any other frequency), then it is valid to conclude that “nothing happened”, i.e. there had not been any acquisition-based signal-specific expansion and subsequent extinction-based reversal of the area gains. This position would deny that two groups (Bieszczad & Weinberger, 2010b and the current group) that were trained identically, using the same protocol that is known to promote specific expansions, and that exhibited the same acquisition functions, indeed had completely different outcomes: area gain in the former and no area gain in the latter. While such a position would contravene much of cumulative science, it cannot be discounted. A related issue might be that while acquisition training produced area gains, they passively dissipated after the termination of such training. However, there is other evidence of expansion and extinction-induced reversal in the present findings. The behavioral strength of extinction was *significantly correlated* with the amount of signal-specific loss of estimated

prior gain in cortical area. Furthermore, this relationship was dependent upon signal-specific behavior during extinction training [Methods, EQ. 5]. Therefore, the significant relationship between reversal of area gain and the final strength of extinction cannot be explained by the position that there was neither gain nor an extinction-induced loss of representational area. In summary, the major finding is support for the “reinforcement prediction” hypothesis, rather than the “significant change” hypothesis.

Maps of Frequency Representation—Initial studies of learning-related frequency representations determined modifications of tuning by obtaining receptive fields (RF) at the training intensity (e.g., 70 dB) (Diamond & Weinberger, 1986; Bakin & Weinberger, 1990; Gao & Suga, 1998; Kisley & Gerstein, 2001; Blake *et al.*, 2002). Cells generally increased their discharges to the training frequency relative to other frequencies, often producing a shift in tuning toward or to the signal tone. Signal-specific increased rate of discharge indicated that the brain may make use of a rate code to index sounds that become behaviorally salient. Gain in representational area was considered a possible complementary mode of encoding stimulus significance because the recruitment of more cells to a signal frequency would increase its influence, even in the absence of actual changes in their rate of discharge. Evidence for such a “spatial code” of stimulus importance was forthcoming (Rutkowski & Weinberger, 2005). Inquiry into such spatial modifications of frequency representation in A1 requires analysis of its tonotopic map.

This study combines a standard approach in auditory neurophysiology (i.e., tonotopic mapping) with a normative task in learning and memory (i.e., appetitive instrumental conditioning). Unfamiliarity with the former might put in question the utility of tonotopic maps to investigate learning-related plasticity in the brain. For example, it has been argued that such maps are arbitrary because they depend upon the sampling density of recordings as well as the areas of auditory cortex from which recordings are obtained.

However, tonotopic maps are not arbitrary. First, primary auditory cortex has an accepted physiological definition that requires neurons to respond at short latency to pure tones, with lower thresholds than for noise stimuli, and have a systematic arrangement of increasing CF across the cortical surface (e.g., Rutkowski *et al.*, 2003; Polley *et al.*, 2007). Thus, mapping determines the boundaries of A1 by these physiological criteria. Second, all mapping methods, including the density of recording from about 6–8 sites within a square millimeter of cortex (i.e., electrode placements 250–500 μm apart), were identical for all subjects regardless of group treatment.

Even so, the use of maps might seem problematic because they represent cellular responses at threshold, while training stimuli are well above threshold. However, frequency representation is most specific at threshold where maps of characteristic frequency have revealed that A1 has a tonotopic organization (Merzenich & Brugge, 1973; Merzenich *et al.*, 1973). The underlying frequency organization is not manifest at high stimulus levels because response bandwidth rapidly increases with increasing sound level and the spatial extent of cortical activation to a pure tone can be extensive (Schreiner, 1998).

Numerous prior studies of A1 plasticity to signal stimuli during training trials have reported increased responses to acoustic signals in A1 (reviewed in Weinberger & Diamond, 1987) but could not provide information about the degree of specificity of effects as they predated the use of multiple test frequencies (but see Ohl & Scheich, 1996, 1997). Studies of frequency receptive fields in A1 at high training stimulus levels have consistently reported signal-specific response facilitation (Bakin & Weinberger, 1990; Edeline *et al.*, 1993; Gao & Suga, 1998, 2000; Ji *et al.*, 2001; Kisley & Gerstein, 2001; Blake *et al.*, 2002; Fritz *et al.*, 2003, 2005). However, they could not provide information on representational area as they

did not include determination of threshold. Synthesis will ultimately be needed of findings from studies of responses during training trials, receptive fields and representational maps. The current study provides needed information on the latter.

Determination of the Strength of Extinction—Determination of the strength of extinction differed from and was more complex than standard approaches to extinction. Thus, it might be argued that the current approach is arbitrary. For example, usual measures of extinction may be limited to the reduced number of behavioral responses during an extinction session (Skinner, 1938; Williams, 1938). However, such measures would have been inadequate for the goals of this study, which also had to (a) determine the *frequency-specificity* of extinction, and (b) obtain post-extinction maps of the frequency organization of the primary auditory cortex. Given the complexity and novelty of the current approach, its rationale is summarized below.

The extinction/spontaneous recovery experimental design was dictated by both of these requirements. While the strength of extinction can be measured by the number of bar-presses performed during a session in which reward has been withdrawn, *specificity* required presentation of probe frequencies as well as the signal tone. This could determine the specificity of responses to the signal tone (EQ. 3), to ultimately yield an individual's value for *signal-specific response in extinction* (SSRE) (EQ. 4).

Furthermore, extinction is subject to spontaneous recovery. Indeed, despite a change from 100% to 0% reward and three days of extinction training, many animals still exhibited behavioral responses to tones in partial or full recovery. (Achieving maximum effects probably will require training of extinction to asymptote in future studies.) Thus, quantification of effective “Final Strength of Extinction” required a correction for the amount of spontaneous recovery (Methods, (EQ. 5). By these methods, the extinction strength to the signal-frequency was revealed to significantly correlate with the magnitude of change (loss) in area between the estimated amount of gain after acquisition training and the measured amount of representational area for the signal tone following extinction/spontaneous recovery ($r = 0.82$, $p < 0.01$) (Fig. 6).

Relevance for the Neural Bases of Extinction and Related Processes

We consider here three findings relevant to understanding the processes of the neural bases of experimental extinction and related processes: (a) specificity to the signal frequency, (b) correlation of extinction strength to area loss, (c) reduction of area below a naïve level.

Specificity to the signal frequency—The estimated extinction-based reduction of representational area in A1 occurred only for the signal frequency. The four generalization test tones, each presented on as many extinction trials as the signal tone, suffered no loss of area. Thus, reduction of representational area is not due simply to the presentation of non-reinforced stimuli, but rather is confined to stimuli that had previously been reinforced.

A possible mechanism for the specific reduction in area might be a mismatch between the predicted (i.e., expected) and actual reinforcement (i.e., outcome) of a signal-stimulus event (Pedreira *et al.*, 2004). While processes for expectation/outcome violation have been shown to especially depend on activity in the midbrain dopamine areas, e.g., ventral tegmental area (VTA) (Schultz, 2001; Zellner & Ranaldi, 2010) and basal ganglia (Schultz *et al.*, 2000; Lauwereyns *et al.*, 2002), it appears that the area of stimulus representation in primary auditory cortex also lies within this domain. In line with this hypothesis, primary auditory cortical cells do exhibit activity related to reward expectancy and also to the discrepancy between the expected and actual reward outcome, for example, in monkeys performing an acoustic categorization task (Brosch *et al.*, 2011). These responses are likely linked to

activity of the orbital, prefrontal and infralimbic cortices which are sensitive to expected reward value and “surprise” operationally defined as prediction error (Watanabe, 1990; Rolls, 2000; Gottfried *et al.*, 2003; Kringelbach, 2005). Indeed, such “top down” influences on the primary auditory cortex are known (Polley *et al.*, 2006; Fritz *et al.*, 2010). Prefrontal cortical stimulation can even *induce* plasticity in the auditory cortex (Yang *et al.*, 2007). In addition, studies of the extinction of tone-conditioned fear have shown that increased tone-induced activity in the infralimbic cortices during tone-extinction predicts stronger extinction memory (Milad & Quirk, 2002; Thompson *et al.*, 2010). This evidence permits control of the strength of extinction by the modification of signal representation in A1 via a frontal cortical influence. Such prefrontal modulation of A1 may depend on its activation of the cholinergic basal forebrain (Rasmusson *et al.*, 2007), especially by the modulation of muscarinic cholinergic action through the nucleus basalis (Metherate & Weinberger, 1990; Bakin & Weinberger, 1996; Kilgard & Merzenich, 1998; Dimyan & Weinberger, 1999; Weinberger *et al.*, 2006). The potential role of the cholinergic system in extinction-based reversal of area gain in A1 remains to be investigated.

Correlation of extinction strength to area loss—The loss of area was correlated with the strength of extinction: the greater the loss, the stronger the extinction (Fig. 7). Prior studies of extinction apparently have not investigated the relationship between extinction strength and neurophysiological plasticity. However, the current findings are generally concordant with previous reports of reversal of cortical plasticity due to extinction. Thus, early studies of the electroencephalogram (EEG) during classical conditioning found that as conditioned stimuli became behaviorally important, i.e., reflected new associations, they began to elicit sensory (and other) cortical activation, which was reversed by extinction training (reviewed in Morrell, 1961). Such extinction-based reversal has also been found for acquisition-induced increased amplitude of signal-evoked potentials in the cat (Galambos *et al.*, 1956) and for signal-enhanced unit discharges in the rat (Disterhoft & Stuart, 1976). More recently, signal-specific receptive field tuning changes in A1 have been found in secondary and ventral-ectosylvian auditory fields in the cat during acquisition and this RF plasticity is reversed during extinction (Diamond & Weinberger, 1986). Consistent with the involvement of the auditory cortex in extinction, Teich *et al.* (1989) reported in the rabbit and Song *et al.* (2010) in the rat, that lesions of the auditory cortex prevented extinction in discriminative auditory training. The present study now extends extinction-induced reversal of acquisition-based effects to increased representational area.

Increases and decreases of the size of representational area itself may have considerable behavioral and physiological sequelae. For example, the same cells that participate in encoding behavioral significance in memory are likely also to be active in working memory tasks (Sakurai, 1994). Thus, memory-induced gains would seem to increase the probability that cells tuned to important sounds would be engaged by stimuli in an ongoing task, including in noisy situations or during “off-modality” attention when auditory cortical resources might be reduced. Additionally, when a stimulus has gained neuronal resources, it should have stronger downstream effects because a greater number of responsive cells should engage more targets and become involved in more neural networks. Among the many possible effects might be that a stimulus that commands more cells would have an advantage in winner-take-all systems. Some of these could control decisions regarding which of competing behavioral potentials would be realized at any moment (Pennartz *et al.*, 2009; Köver & Bao, 2010). Thus the size of representational area of a sound might give that stimulus critical access to motor planning and motor execution circuits.

On the other hand, when neural processes *reduce* a signal’s representation, it probably loses its prior advantages, at least in the absence of spontaneous recovery. Additionally, its associativity might likewise diminish. Learning for that stimulus could be retarded as it

becomes less able to command attention and maintain broad connections within various networks. This formulation provides a plausible mechanism in part for the retardation of subsequent acquisition by prior stimulus pre-exposure (i.e., “latent inhibition”: Lubow & Moore, 1959; of neural responses in A1: Westenberg & Weinberger, 1976; Condon & Weinberger, 1991). Also, reduced representational area might account for reduced levels of associativity consequent to conditioned inhibition as well as extinction (Rescorla, 2002), by diminishing the availability of specific “nodes”, i.e., specific cortical stimulus representations, to become incorporated into a network of associations with conditioned reinforcers (Wagner, 1981; Brandon *et al.*, 2003). Indeed, area of representation is related to strength of specific memory: the greater the area of signal-specific expansion in A1, the greater the resistance to extinction, i.e., the *stronger the memory* for that sound (Bieszczad & Weinberger, 2010c). Such a substrate might help explain both the intrusiveness and strength of traumatic memories, as in post-traumatic stress disorder (PTSD). If such memories incorporate more neurons, then not only will they be more difficult to forget, but the increased number of involved cells is more likely to activate underlying memory networks during spontaneous activity or a related experience that would be sub-threshold for the non-sufferers.

Overall, as the magnitude of signal-specific representational gain in the tonotopic map of A1 may promote the strength of specific excitatory auditory associations, so may the magnitude of areal loss promote the strength of specific inhibitory associations. In effect, specific gains and losses in representational area within primary sensory cortex could appropriately link specific important events to cognition and action.

Reduction of area below naïve level—Three subjects suffered reversal of signal-specific estimated areas gains to below the naïve level. This extra reduction of representational area suggests a mechanism for “below zero” extinction. Pavlov (1927) discovered that continued presentation of a signal without reinforcement after the signal no longer elicited any behavioral responses during extinction produced additional inhibitory learning, which could be revealed by retardation of subsequent new excitatory conditioning.

Perhaps behavioral sub-zero extinction develops *due to* continued loss of the area of signal representation below the naïve level, when extinction training continues after the absence of behavioral responses. Such a reduction of representational area might account, at least in part, for the impaired ability of the “over-extinguished” stimulus to enter into new excitatory associations. This hypothesis is based on the finding that the size of signal representational area seems to be related to behavioral significance by its value for reinforcement prediction. Thus, stimulus under-representation in cortical area may impair new excitatory associations because more training would be required to increase signal-specific area above the naïve level, to thereby encode reinforcement prediction. More generally, the initial sizes of representational areas of sensory stimuli might affect their associativity. Not only might excessively small areas contribute to retardation of learning, but also areas larger than naïve might promote new associations and perhaps help explain positive transfer effects.

Future Directions

For the learning task at hand, the findings support the “reinforcement prediction” hypothesis of expanded representation of a signal stimulus in the primary auditory cortex. It is noteworthy that consideration of such a cognitive function involves early sensory cortex, which had traditionally been assigned the function of analysis of stimulus parameters. A rapidly growing body of findings has required a broadened conceptualization of the functions of A1, and likely other primary sensory cortices, in both the fields of sensory physiology and learning/memory (reviewed in Scheich *et al.*, 2011; Weinberger, 2011).

Future investigations need to fully determine the long-term temporal dynamics of learning-related changes in sensory cortical representational area, as well as of other physiological indices of associative processes. Thus, it has long been known that neurophysiological correlates undergo various changes with stages of learning. After having developed during early stages of learning, they may weaken or even disappear after continued training, particularly when it continues for many days after achieving asymptotic performance, i.e., during over-training. This “waning” of associative plasticity has been known since the start of electrophysiological studies. For example, early research on the electroencephalogram (EEG) found that conditioned stimuli (CS) elicited widespread cortical activation as animals learned, but continued training produced a “shrinkage” of involved cortical fields (Morrell & Jasper, 1956). Direct study of changes in learning-induced cortical excitability, determined by electrical stimulation of the cortex, obtained the same pattern of results (Kogan, 1960). Such dynamics may reflect an initial necessary recruitment of more neurons to the signal stimulus followed by increased neural efficiency with continued training, transfer of behavioral control to other brain regions, re-coding of plasticity to an as yet undetected form at the initial site, or a combination of such processes. Recent studies in the motor (Molina-Luna *et al.*, 2008), somatosensory (Siucinska & Kossut, 1996) and auditory (Reed *et al.*, 2011) cortices have extended “waning” of neurophysiological correlates to representational plasticity. Given such findings, it is important for future studies to determine the complete temporal dynamics of map expansions and contractions, including the effects of extinction training and other inhibitory associations.

Related to this issue is the need for reductionistic studies of mechanisms underlying both acquisition-based expansion and extinction-based reduction of representational area. While evidence for involvement of the cholinergic system has been mentioned, other neuromodulators are likely to be involved; in particular, dopamine (Schicknick *et al.*, 2008) and nor-epinephrine (Manunta & Edeline, 2004) have been identified as modulators of the auditory cortical during learning, memory and various tasks.

Conclusion

The current and related findings also indicate close relationships between the magnitude of representational area on the one hand and reinforcement prediction, associativity and memory strength on the other. The present clarification of *behavioral importance* as linked to reinforcement prediction is a beginning of explication, but further dissections of behavior and related neural plasticity are needed. Until a full “catalog” of the factors and forms of sensory cortical plasticity is achieved, it will be difficult, if not impossible to formulate a comprehensive model of the overall role of primary sensory cortex. Part of this process requires determination of the circumstances in which representational plasticity *does not form*, despite the acquisition and maintenance of discrimination and learning performance (e.g., Brown *et al.*, 2004; see also Recanzone *et al.*, 1993). There is accumulating evidence that *how* a task is solved can be critical for the development of associative plasticity in A1 (Berlau & Weinberger, 2008; Bieszczad & Weinberger, 2010a, 2010b). Thus, animals may solve some auditory problems without depending upon plasticity in the primary auditory cortex, yet may rely upon A1 plasticity at other times or circumstances. Therefore, it will be necessary to keep in mind all factors in order to adequately understand how sensory cortices sub-serve both their sensory-analytic and their cognitive functions.

Acknowledgments

This research was supported by research grants from the National Institutes of Health (NIDCD) DC-010013 to NMW and DC-009163 to KMB. We would like to thank Jacquie Weinberger, Nataliya M. Gross and Gabriel K. Hui for technical assistance.

References

- Bakin JS, Kwon MC, Masino SA, Weinberger NM. Suprathreshold auditory cortex activation visualized by intrinsic signal optical imaging. *Cereb Cortex*. 1996; 6(2):120–130. [PubMed: 8670644]
- Bakin JS, Weinberger NM. Classical conditioning induces CS-specific receptive field plasticity in the auditory cortex of the guinea pig. *Brain Res*. 1990; 536(1–2):271–286. [PubMed: 2085753]
- Bakin JS, Weinberger NM. Induction of a physiological memory in the cerebral cortex by stimulation of the nucleus basalis. *Proc Natl Acad Sci USA*. 1996; 93(20):11219–11224. [PubMed: 8855336]
- Berlau KM, Weinberger NM. Learning strategy determines auditory cortical plasticity. *Neurobiol Learn Mem*. 2008; 89(2):153–166. [PubMed: 17707663]
- Bieszczad KM, Weinberger NM. Learning strategy trumps motivational level in determining learning-induced auditory cortical plasticity. *Neurobiol Learn Mem*. 2010a; 93(2):229–239. [PubMed: 19853056]
- Bieszczad KM, Weinberger NM. Remodeling the cortex in memory: increased use of a learning strategy increases the representational area of relevant acoustic cues. *Neurobiol Learn Mem*. 2010b; 94(2):127–144. [PubMed: 20434577]
- Bieszczad KM, Weinberger NM. Representational gain in cortical area underlies increase of memory strength. *Proc Natl Acad Sci USA*. 2010c; 107(8):3793–3798. [PubMed: 20133679]
- Blake DT, Strata F, Churchland AK, Merzenich MM. Neural correlates of instrumental learning in primary auditory cortex. *Proc Natl Acad Sci USA*. 2002; 99(15):10114–10119. [PubMed: 12119383]
- Brandon SE, Vogel EH, Wagner AR. Stimulus representation in SOP: I. Theoretical rationalization and some implications. *Behav Processes*. 2003; 62(1–3):5–25. [PubMed: 12729966]
- Brosch M, Selezneva E, Scheich H. Representation of reward feedback in primate auditory cortex. *Front Syst Neurosci*. 2011; 5:5. [PubMed: 21369350]
- Brown M, Irvine DR, Park VN. Perceptual learning on an auditory frequency discrimination task by cats: association with changes in primary auditory cortex. *Cereb Cortex*. 2004; 14(9):952–965. [PubMed: 15115736]
- Carpenter-Hyland E, Plummer T, Vazdarjanova A, Blake D. Arc expression and neuroplasticity in primary auditory cortex during initial learning are inversely related to neural activity. *Proc Natl Acad Sci USA*. 2010; 107(33):14828–14832. [PubMed: 20675582]
- Condon CD, Weinberger NM. Habituation produces frequency-specific plasticity of receptive fields in the auditory cortex. *Behav Neurosci*. 1991; 105(3):416–430. [PubMed: 1863363]
- Diamond DM, Weinberger NM. Classical conditioning rapidly induces specific changes in frequency receptive fields of single neurons in secondary and ventral ectosylvian auditory cortical fields. *Brain Res*. 1986; 372(2):357–360. [PubMed: 3708366]
- Dimyan MA, Weinberger NM. Basal forebrain stimulation induces discriminative receptive field plasticity in the auditory cortex. *Behav Neurosci*. 1999; 113(4):691–702. [PubMed: 10495078]
- Disterhoft JF, Stuart DK. Trial sequence of changed unit activity in auditory system of alert rat during conditioned response acquisition and extinction. *J Neurophysiol*. 1976; 39 (2):266–281. [PubMed: 1255223]
- Downman CB, Woolsey CN, Lende RA. Auditory areas I, II, and Ep: cochlear representation, afferent paths and interconnections. *Bull Johns Hopkins Hosp*. 1960; 106:127–142. [PubMed: 13817915]
- Edeline JM, Pham P, Weinberger NM. Rapid development of learning-induced receptive field plasticity in the auditory cortex. *Behav Neurosci*. 1993; 107(4):539–557. [PubMed: 8397859]
- Fritz J, Shamma S, Elhilali M, Klein D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci*. 2003; 6(11):1216–1223. [PubMed: 14583754]
- Fritz JB, David SV, Radtke-Schuller S, Yin P, Shamma SA. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nat Neurosci*. 2010; 13(8):1011–1019. [PubMed: 20622871]
- Fritz JB, Elhilali M, Shamma SA. Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *J Neurosci*. 2005; 25(33):7623–7635. [PubMed: 16107649]

- Fritz JB, Elhilali M, Shamma SA. Adaptive changes in cortical receptive fields induced by attention to complex sounds. *J Neurophysiol.* 2007; 98(4):2337–2346. [PubMed: 17699691]
- Galambos R, Sheatz G, Vernier VG. Electrophysiological correlates of a conditioned response in cats. *Science.* 1956; 123(3192):376–377. [PubMed: 13298694]
- Gao E, Suga N. Experience-dependent corticofugal adjustment of midbrain frequency map in bat auditory system. *Proc Natl Acad Sci USA.* 1998; 95(21):12663–12670. [PubMed: 9770543]
- Gao E, Suga N. Experience-dependent plasticity in the auditory cortex and the inferior colliculus of bats: role of the corticofugal system. *Proc Natl Acad Sci USA.* 2000; 97 (14):8081–8086. [PubMed: 10884432]
- Ghose GM, Bearl DW. Attention directed by expectations enhances receptive fields in cortical area MT. *Vision Res.* 2010; 50(4):441–451. [PubMed: 19819253]
- Gonzalez-Lima F, Scheich H. Neural substrates for tone-conditioned bradycardia demonstrated with 2-deoxyglucose. II. Auditory cortex plasticity. *Behav Brain Res.* 1986; 20 (3):281–293. [PubMed: 3741589]
- Gottfried JA, O'Doherty J, Dolan RJ. Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science.* 2003; 301(5636):1104–1107. [PubMed: 12934011]
- Hackett TA, Barkat TR, O'Brien BM, Hensch TK, Polley DB. Linking topography to tonotopy in the mouse auditory thalamocortical circuit. *J Neurosci.* 2011; 31(8):2983–2995. [PubMed: 21414920]
- Harrison RV, Kakigi A, Hirakawa H, Harel N, Mount RJ. Tonotopic mapping in auditory cortex of the chinchilla. *Hear Res.* 1996; 100(1–2):157–163. [PubMed: 8922990]
- Heil P. Auditory cortical onset responses revisited: I. First-spike timing. *J Neurophysiol.* 1997a; 77(5): 2616–2641. [PubMed: 9163380]
- Heil P. Auditory cortical onset responses revisited: II. Response strength. *J Neurophysiol.* 1997b; 77(5):2642–2660. [PubMed: 9163381]
- Heil P, Irvine DR. The posterior field P of cat auditory cortex: coding of envelope transients. *Cereb Cortex.* 1998; 8(2):125–141. [PubMed: 9542892]
- Hoshino O. Enhanced sound perception by widespread-onset neuronal responses in auditory cortex. *Neural Comput.* 2007; 19(12):3310–3334. [PubMed: 17970655]
- Hui GK, Wong KL, Chavez CM, Leon MI, Robin KM, Weinberger NM. Conditioned tone control of brain reward behavior produces highly specific representational gain in the primary auditory cortex. *Neurobiol Learn Mem.* 2009; 92(1):27–34. [PubMed: 19249380]
- Jaramillo S, Zador AM. The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nat Neurosci.* 2011; 14(2):246–251. [PubMed: 21170056]
- Ji W, Gao E, Suga N. Effects of acetylcholine and atropine on plasticity of central auditory neurons caused by conditioning in bats. *J Neurophysiol.* 2001; 86(1):211–225. [PubMed: 11431503]
- Kilgard MP, Merzenich MM. Cortical map reorganization enabled by nucleus basalis activity. *Science.* 1998; 279(5357):1714–1718. [PubMed: 9497289]
- Kisley MA, Gerstein GL. Daily variation and appetitive conditioning-induced plasticity of auditory cortex receptive fields. *Eur J Neurosci.* 2001; 13(10):1993–2003. [PubMed: 11403693]
- Knight PL. Representation of the cochlea within the anterior auditory field (AAF) of the cat. *Brain Res.* 1977; 130(3):447–463. [PubMed: 890444]
- Kogan, AB. Manifestations of processes of higher nervous activity in the electrical potentials of the cortex during free behavior of animals *Moscow Colloquium on Electroencephalography of Higher Nervous Activity (Electroencephalography and Clinical Neurophysiology series, Suppl 13).* In: Jasper, HH.; Smirnov, GD., editors. EEG Journal, Montreal. 1960. p. 51–65.
- Konorski, J. Conditioned Reflexes and Neuron Organization. Cambridge University Press; New York: 1948.
- Köver H, Bao S. Cortical plasticity as a mechanism for storing Bayesian priors in sensory perception. *PLoS One.* 2010; 5(5):e10497. [PubMed: 20463967]
- Kringelbach ML. The human orbitofrontal cortex: linking reward to hedonic experience. *Nat Rev Neurosci.* 2005; 6(9):691–702. [PubMed: 16136173]

- Lauwereyns J, Takikawa Y, Kawagoe R, Kobayashi S, Koizumi M, Coe B, Sakagami M, Hikosaka O. Feature-based anticipation of cues that predict reward in monkey caudate nucleus. *Neuron*. 2002; 33(3):463–473. [PubMed: 11832232]
- Linden JF, Liu RC, Sahani M, Schreiner CE, Merzenich MM. Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. *J Neurophysiol*. 2003; 90(4):2660–2675. [PubMed: 12815016]
- Lubow RE, Moore AU. Latent inhibition: the effect of nonreinforced pre-exposure to the conditional stimulus. *J Comp Physiol Psychol*. 1959; 52(4):415–419. [PubMed: 14418647]
- Manunta Y, Edeline JM. Noradrenergic induction of selective plasticity in the frequency tuning of auditory cortex neurons. *J Neurophysiol*. 2004; 92(3):1445–1463. [PubMed: 15084638]
- Menning H, Roberts LE, Pantev C. Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *Neuroreport*. 2000; 11(4):817–822. [PubMed: 10757526]
- Merzenich MM, Brugge JF. Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Res*. 1973; 50(2):275–296. [PubMed: 4196192]
- Merzenich MM, Knight PL, Roth GL. Cochleotopic organization of primary auditory cortex in the cat. *Brain Res*. 1973; 63:343–346. [PubMed: 4764302]
- Merzenich MM, Knight PL, Roth GL. Representation of the cochlea within the primary auditory cortex in the cat. *J Neurophysiol*. 1975; 38(2):231–249. [PubMed: 1092814]
- Merzenich MM, Nelson RJ, Kaas JH, Stryker MP, Jenkins WM, Zook JM, Cynader MS, Schoppmann A. Variability in hand surface representations in areas 3b and 1 in adult owl and squirrel monkeys. *J Comp Neurol*. 1987; 258(2):281–296. [PubMed: 3584541]
- Metherate R, Weinberger NM. Cholinergic modulation of responses to single tones produces tone-specific receptive field alterations in cat auditory cortex. *Synapse*. 1990; 6(2):133–145. [PubMed: 2237776]
- Milad MR, Quirk GJ. Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature*. 2002; 420(6911):70–74. [PubMed: 12422216]
- Molchan SE, Sunderland T, McIntosh AR, Herscovitch P, Schreurs BG. A functional anatomical study of associative learning in humans. *Proc Natl Acad Sci USA*. 1994; 91(17):8122–8126. [PubMed: 8058767]
- Molina-Luna K, Hertler B, Buitrago MM, Luft AR. Motor learning transiently changes cortical somatotopy. *Neuroimage*. 2008; 40(4):1748–1754. [PubMed: 18329289]
- Morrell F. Electrophysiological contributions to the neural basis of learning. *Physiol Rev*. 1961; 41:443–494. [PubMed: 13772805]
- Morrell F, Jasper HH. Electrographic studies of the formation of temporary connections in the brain. *Electroenceph Clin Neurophysiol*. 1956; 8(2):201–215. [PubMed: 13317812]
- Morris JS, Friston KJ, Dolan RJ. Experience-dependent modulation of tonotopic neural responses in human auditory cortex. *Proc Biol Sci*. 1998; 265(1397):649–657. [PubMed: 9608726]
- Ohl FW, Scheich H. Differential frequency conditioning enhances spectral contrast sensitivity of units in auditory cortex (field AI) of the alert Mongolian gerbil. *Eur J Neurosci*. 1996; 8(5):1001–1017. [PubMed: 8743748]
- Ohl FW, Scheich H. Learning-induced dynamic receptive field changes in primary auditory cortex of the unanaesthetized Mongolian gerbil. *J Comp Physiol A*. 1997; 181(6):685–696. [PubMed: 9449827]
- Pascual-Leone A, Hamilton R. The metamodal organization of the brain. *Prog Brain Res*. 2001; 134:427–445. [PubMed: 11702559]
- Pavlov, IP. *Conditioned Reflexes*. Dover Publications; New York: 1927.
- Pedreira ME, Pérez-Cuesta LM, Maldonado H. Mismatch between what is expected and what actually occurs triggers memory reconsolidation or extinction. *Learn Mem*. 2004; 11(5):579–585. [PubMed: 15466312]
- Pennartz CM, Berke JD, Graybiel AM, Ito R, Lansink CS, van der Meer M, Redish AD, Smith KS, Voorn P. Corticostriatal interactions during learning, memory processing, and decision making. *J Neurosci*. 2009; 29(41):12831–12838. [PubMed: 19828796]

- Phillips DP, Hall SE, Boehnke SE. Central auditory onset responses, and temporal asymmetries in auditory perception. *Hear Res.* 2002; 167(1–2):192–205. [PubMed: 12117542]
- Polley DB, Read HL, Storace DA, Merzenich MM. Multiparametric auditory receptive field organization across five cortical fields in the albino rat. *J Neurophysiol.* 2007; 97 (5):3621–3638. [PubMed: 17376842]
- Polley DB, Steinberg EE, Merzenich MM. Perceptual learning directs auditory cortical map reorganization through top-down influences. *J Neurosci.* 2006; 26(18):4970–4982. [PubMed: 16672673]
- Rasmusson D, Smith S, Semba K. Inactivation of prefrontal cortex abolishes cortical acetylcholine release evoked by sensory or sensory pathway stimulation in the rat. *Neuroscience.* 2007; 149(1): 232–241. [PubMed: 17850979]
- Rauschecker JP. Auditory cortical plasticity: a comparison with other sensory systems. *Trends Neurosci.* 1999; 22(2):74–80. [PubMed: 10092047]
- Reale RA, Imig TJ. Tonotopic organization in auditory cortex of the cat. *J Comp Neurol.* 1980; 192(2): 265–294. [PubMed: 7400399]
- Recanzone GH, Schreiner CE, Merzenich MM. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J Neurosci.* 1993; 13(1): 87–103. [PubMed: 8423485]
- Reed A, Riley J, Carraway R, Carrasco A, Perez C, Jakkamsetti V, Kilgard MP. Cortical map plasticity improves learning but is not necessary for improved performance. *Neuron.* 2011; 70(1): 121–131. [PubMed: 21482361]
- Rescorla RA. Savings tests: separating differences in rate of learning from differences in initial levels. *J Exp Psychol Anim Behav Process.* 2002; 28(4):369–377. [PubMed: 12395494]
- Rolls ET. The orbitofrontal cortex and reward. *Cereb Cortex.* 2000; 10(3):284–294. [PubMed: 10731223]
- Rose JE, Woolsey CN. The relationships of thalamic connections, cellular structure and evokable electrical activity in the auditory region of the cat. *J Comp Neurol.* 1949; 91(3):441–466. [PubMed: 15409490]
- Rutkowski RG, Miasnikov AA, Weinberger NM. Characterisation of multiple physiological fields within the anatomical core of rat auditory cortex. *Hear Res.* 2003; 181(1–2):116–130. [PubMed: 12855370]
- Rutkowski RG, Weinberger NM. Encoding of learned importance of sound by magnitude of representational area in primary auditory cortex. *Proc Natl Acad Sci USA.* 2005; 102(38):12664–13669.
- Sakurai Y. Involvement of auditory cortical and hippocampal neurons in auditory working memory and reference memory in the rat. *J Neurosci.* 1994; 14(5 Pt 1):2606–2623. [PubMed: 8182430]
- Sally SL, Kelly JB. Organization of auditory cortex in the albino rat: sound frequency. *J Neurophysiol.* 1988; 59(5):1627–1638. [PubMed: 3385476]
- Scheich H, Brechmann A, Brosch M, Budinger E, Ohl FW, Selezneva E, Stark H, Tischmeyer W, Wetzel W. Behavioral semantics of learning and crossmodal processing in auditory cortex: the semantic processor concept. *Hear Res.* 2011; 271(1–2):3–15. [PubMed: 20971178]
- Schicknick H, Schott BH, Budinger E, Smalla KH, Riedel A, Seidenbecher CI, Scheich H, Gundelfinger ED, Tischmeyer W. Dopaminergic modulation of auditory cortex-dependent memory consolidation through mTOR. *Cereb Cortex.* 2008; 18(11):2646–2658. [PubMed: 18321872]
- Schreiner CE. Spatial distribution of responses to simple and complex sounds in the primary auditory cortex. *Audiol Neurotol.* 1998; 3(2–3):104–122. [PubMed: 9575380]
- Schultz W. Reward signaling by dopamine neurons. *Neuroscientist.* 2001; 7(4):293–302. [PubMed: 11488395]
- Schultz W, Tremblay L, Hollerman JR. Reward processing in primate orbitofrontal cortex and basal ganglia. *Cereb Cortex.* 2000; 10(3):272–284. [PubMed: 10731222]
- Siucinska E, Kossut M. Short-lasting classical conditioning induces reversible changes of representational maps of vibrissae in mouse SI cortex — a 2DG study. *Cereb Cortex.* 1996; 6(3): 506–513. [PubMed: 8670676]

- Skinner, BF. *The Behavior of Organisms*. Prentice Hall; Englewood Cliffs, NJ: 1938.
- Song EY, Boatman JA, Jung MW, Kim JJ. Auditory cortex is important in the extinction of two different tone-based conditioned fear memories in rats. *Front Behav Neurosci*. 2010; 4:24. [PubMed: 20631823]
- Stiebler I, Neulist R, Fichtel I, Ehret G. The auditory cortex of the house mouse: left right differences, tonotopic organization and quantitative analysis of frequency representation. *J Comp Physiol A*. 1997; 181(6):559–571. [PubMed: 9449817]
- Takahashi H, Funamizu A, Mitsumori Y, Kose H, Kanzaki R. Progressive plasticity of auditory cortex during appetitive operant conditioning. *BioSystems*. 2010; 101(1):37–41. [PubMed: 20398726]
- Teich AH, McCabe PM, Gentile CC, Schneiderman LS, Winters RW, Liskowsky DR, Schneiderman N. Auditory cortex lesions prevent the extinction of Pavlovian differential heart rate conditioning to tonal stimuli in rabbits. *Brain Res*. 1989; 480(1–2):210–218. [PubMed: 2713651]
- Thomas H, Tillein J, Heil P, Scheich H. Functional organization of auditory cortex in the mongolian gerbil (*Meriones unguiculatus*). I. Electrophysiological mapping of frequency representation and distinction of fields. *Eur J Neurosci*. 1993; 5(7):882–897. [PubMed: 8281300]
- Thompson BM, Baratta MV, Biedenkapp JC, Rudy JW, Watkins LR, Maier SF. Activation of the infralimbic cortex in a fear context enhances extinction learning. *Learn Mem*. 2010; 17(11):591–599. [PubMed: 21041382]
- Tsytsarev V, Yamazaki T, Ribot J, Tanaka S. Sound frequency representation in cat auditory cortex. *Neuroimage*. 2004; 23(4):1246–1255. [PubMed: 15589090]
- Villa AE, Tetko IV, Hyland B, Najem A. Spatiotemporal activity patterns of rat cortical neurons predict responses in a conditioned task. *Proc Natl Acad Sci USA*. 1999; 96 (3):1106–1111. [PubMed: 9927701]
- Wagner, A. SOP: a model of automatic memory processing in animal behavior. In: Spear, NE.; Miller, RR., editors. *Information Processing in Animals: Memory Mechanisms*. Erlbaum; Hillsdale, N.J: 1981. p. 5-47.
- Wang X, Lu T, Snider RK, Liang L. Sustained firing in auditory cortex evoked by preferred stimuli. *Nature*. 2005; 435(7040):341–346. [PubMed: 15902257]
- Watanabe M. Prefrontal unit activity during associative learning in the monkey. *Exp Brain Res*. 1990; 80(2):296–309. [PubMed: 2358045]
- Weinberger NM. Auditory associative memory and representational plasticity in the primary auditory cortex. *Hear Res*. 2007; 229(1–2):54–68. [PubMed: 17344002]
- Weinberger, NM. Reconceptualizing the primary auditory cortex: learning, memory and specific plasticity. In: Winer, JA.; Schreiner, CE., editors. *The Auditory Cortex*. Vol. 22. Springer; New York: 2011. p. 465-491.
- Weinberger NM, Diamond DM. Physiological plasticity in auditory cortex: rapid induction by learning. *Prog Neurobiol*. 1987; 29(1):1–55. [PubMed: 3295997]
- Weinberger NM, Miasnikov AA, Chen JC. The level of cholinergic nucleus basalis activation controls the specificity of auditory associative memory. *Neurobiol Learn Mem*. 2006; 86(3):270–285. [PubMed: 16750916]
- Westenberg IS, Weinberger NM. Evoked potential decrements in auditory cortex. II. Critical test for habituation. *Electroencephalogr Clin Neurophysiol*. 1976; 40(4):356–369. [PubMed: 56262]
- Wetzel W, Wagner T, Ohl FW, Scheich H. Categorical discrimination of direction in frequency-modulated tones by Mongolian gerbils. *Behav Brain Res*. 1998; 91(1–2):29–39. [PubMed: 9578437]
- Williams SB. Resistance to extinction as a function of the number of reinforcements. *J Exp Psychol*. 1938; 23(5):506–522.
- Yang WW, Zhou XM, Zhang JP, Sun XD. Modulation of frequency receptive field plasticity in rat auditory cortical neurons by electrical stimulation of medial prefrontal cortex. *Acta Physiol Sinica*. 2007; 59(6):784–790. [PubMed: 18157472]
- Zellner MR, Ranaldi R. How conditioned stimuli acquire the ability to activate VTA dopamine cells: a proposed neurobiological component of reward-related learning. *Neurosci Biobehav Rev*. 2010; 34(5):769–780. [PubMed: 19914285]

- Zhou X, de Villers-Sidani E, Panizzutti R, Merzenich MM. Successive-signal biasing for a learned sound sequence. *Proc Natl Acad Sci USA*. 2010; 107(33):14839–14844. [PubMed: 20679210]
- Zhou X, Merzenich MM. Intensive training in adults refines A1 representations degraded in an early postnatal critical period. *Proc Natl Acad Sci USA*. 2007; 104(40):15935–15940. [PubMed: 17895375]

Fig. 1.

Experimental protocols and timeline. There were two behavioral training phases, “Acquisition” and “Extinction”, followed by electrophysiological recording mapping from A1. *Acquisition*. Rats were trained to bar-press (B) to a 5.0 kHz signal tone (T) for a water reward (R). At least one rewarded bar-press during the tone allowed an additional rewarded bar-press during a “Free” period up to 7 s immediately after the tone ended. The first bar-press after the reward period generated a time-out error-period signaled by a flashing light (E). All other erroneous responses during the silent inter-trial intervals generated an error time-out period 50% of the time. *TOTE Test*. Rats were tested without Free periods to determine the degree to which they exhibit such a pattern of response in which bar-presses are made beginning at tone onset and continue until the delivery of the first error. This pattern (as shown) reveals use of a learning-strategy called TOTE (“tone-onset-to-error”). The use of TOTE is quantified by a TOTE learning index (TLI) which is used to predict the amount of frequency-specific area gain in A1 (see Fig. 2). *Extinction*. Typical extinction trials present only a single signal-frequency without reinforcement. Here, we expanded extinction training to include trials that presented non-signal frequencies called “generalization-test” tones. This could reveal the frequency-specificity of extinction by comparison of the decrement in bar-pressing to the signal-tone vs. non-signal probe tone frequencies. *Spontaneous recovery*. The amount of recovery of bar-press responses to tones was tested 2 days after extinction to assess the strength of extinction. This session was identical to that of extinction training. *Electrophysiology*. Frequency areas of representation were determined by microelectrode mapping procedures once, at the end of all training. *Top right*, bar indicates the scale for a 5 s time period in the illustration. *Right column references equation indices provided in Methods that used data from the indicated part of the experiment.*

Fig. 2.

Relationship between the acquisition of a tone-reward association and area of representation of the signal frequency in A1. A significant linear relationship between behavior and brain has been previously established for the tone-reward instrumental task used in this study (equation indicated in figure; [EQ. 1]). The predictive behavior is the degree to which animals solve the task by use of a “tone-onset-to-error” (TOTE) strategy, *viz.*, bar-pressing from onset of a 10 s tone until receiving an error signal (cued by a flashing light during a time-out error-period) for responding in silence after tone offset. The quantification of the degree of use of the TOTE strategy by the TOTE Learning Index (TLI; see Methods, “*TOTE Learning Index*”) (x-axis), predicts the percent of A1 area representation of the 5.0 kHz signal frequency (within a ± 0.25 octave band) (y-axis). The more use of the TOTE learning strategy (i.e., and the higher the TLI), the larger the area in A1. *Reproduced from Fig. 7, Bieszczad & Weinberger (2010b).*

Fig. 3.

Performance and cortical gains in acquisition. (A) Animals learned to bar-press to tones for rewards to a stable level of performance within 17 days of training. Dashed line indicates transition in difficulty from a protocol with short inter-trial-intervals (first four days = 4–12 s, random schedule), to long inter-trial interval durations (5–25 s, random schedule) (Methods, “Acquisition”). (B) Cortical map data from animals trained identically in a prior study (Bieszczad and Weinberger 2010b) were used to estimate the gains in representational area in the current study. The learning curves for the current group and the previous group were the same (*inset*). Bieszczad and Weinberger (2010b) found a performance correlate (TLI, x -axis) of signal-specific area gain in A1 (y -axis, amount of A1 area for the 5.0 kHz signal frequency $\pm$ 0.25 octave) (open circles, $n = 7$; as in Fig. 2). Here, the established linear relationship was used to estimate the magnitude of signal-specific increase in

representational area with the acquisition of the tone–reward association (closed diamonds, $n = 11$; jittered on the x -axis dimension to avoid overlap). Table 2 provides all TLI and estimated area values. The size of A1 area representation for the signal frequency band in naïve animals is shown by the horizontal solid line (mean $\pm$ standard error) ($7.7\% \pm 1.6$). The current group mean area is indicated by the dashed line ($12.1\% \pm 3.2$; see also Fig. 6).

Fig. 4.

Extinction behavior. Extinction training resulted in a decrease in the number of responses to all tones. A significant group $\times$ day interaction ($F_{4,98} = 5.97, p < 0.001$) revealed the decrement in behavioral response was dependent on tone frequency. The greatest decrease in responses was for the signal frequency (solid line, closed circles) and less to frequencies that were more spectrally distant from the signal frequency (Neighbor tones: 2.4 and 10.6 kHz, dashed line, open circles; Far tones: 1.1 and 22.4 kHz, dotted line, open circles).

Fig. 5.

Magnitudes of signal-specific response during extinction and spontaneous recovery. The signal (5.0 kHz) and four novel non-signal tone frequencies were presented during extinction and spontaneous recovery sessions. Signal-specific responses in extinction (“SSRE”) values reveal the amount of decrement in *signal-specific* behavioral responses in extinction (see Methods, “*Extinction*”). Three examples are illustrated. (a) On the 1st and 3rd extinction days, peak responses were at the flanking high frequency while they were at the signal frequency on day 2 and during the test for spontaneous recovery. (b) Peak responses were at the signal frequency throughout extinction training and spontaneous recovery. (c) Peak responses were at the flanking higher frequency on Day 1 of extinction. Although the peak was at the signal frequency on Day 2, the frequency response profile was still broad, as it was to a lesser extent during spontaneous recovery. The range of SSRE values in these

examples indicates a high degree of inter-individual variability of extinction learning. Nonetheless, the general pattern was a decrease in SSRE (i.e., decreased responses to the signal-frequency) across extinction training and some degree of spontaneous recovery centered at the signal frequency thereafter. Abbreviations: Ext D1, Ext D2, Ext D3 = Days 1, 2 and 3 of extinction training, respectively; Spon. Rec = spontaneous recovery test session; SSRE = Signal-specific responses in extinction (see text; [EQ. 4]). SSRE values for all animals during all extinction and spontaneous sessions are provided in Table 1.

Fig. 6.

Representation of frequency in A1 following all training. (A) The amount of characteristic frequency (CF) representational area for the signal frequency after extinction was determined relative to the total size of A1 (y-axis, % of total). CF distributions in half-octave bands (centered at the 5.0 kHz signal frequency) revealed a significant difference in the interpolated A1 representation for the signal frequency (5.0 kHz $\pm$ 0.25 octaves) from 7.7% ($\pm$ 1.6 s.e.m.) in naïves to 12.5% ($\pm$ 3.4 s.e.m.) in trained subjects after acquisition (filled circle; $t_{18} = 3.58$, $*p < 0.001$). However, extinction reduced the signal area from 12.1% $\pm$ 3.2 to 7.5% ($\pm$ 3.7 s.e.m.) (filled symbols). In contrast, the areas of non-signal frequencies were not significantly reduced (all $F_{1,20}$ ratios, $p > 0.05$). (B) A1 maps of trained subjects that varied in the amount of cortical representation of the signal frequency. Examples show that subjects could have signal areas that were not significantly different in size from naïve (e.g., 7.7%), or with significantly more area (e.g., 10.6%) or less signal area (e.g., 2.2%) than naïve. Thus, the effects of extinction on the representation of frequency in A1 could differ between subjects.

Fig. 7.

Behavioral extinction is significantly related to the magnitude of area reduction in A1. There was a significant relationship between the estimated area of change (mainly loss) of signal-specific representational area in A1 and the strength of behavioral extinction: the greater the reduction in signal area, the stronger the extinction. Best fit regression was curvilinear: $y = -1.37x^2 + 35.17x - 121.94$ ($r = 0.82$, $p < 0.01$; $n = 11$). The x -axis shows the magnitude of reversal in the signal-specific area estimated to have been gained in A1 for the signal-frequency (in a half-octave band centered at the signal-frequency). Larger positive values along the x -axis indicate greater reductions in signal area with extinction. Larger negative values indicate lesser reductions. The magnitudes of individual extinction-induced area loss were calculated as the difference between the estimated gain in area due to acquisition (Fig. 3B) and the signal-specific area determined electrophysiologically after extinction (Fig. 6). The y -axis shows the strength of extinction for the signal tone (5.0 kHz) as the percentage of the total decrement of behavioral responses during the three days of extinction that recovered two days later during the spontaneous recovery test. The significant relationship between signal area and memory strength is curvilinear likely because the amount of possible loss in A1 area reaches a floor because the amount of signal-specific representational area cannot decrease to a negative value. Annotated diamonds (a, b and c) reference individual animals shown in Fig. 5A, 5B and 5C in extinction and spontaneous recovery, respectively.

Table 1

The strength of extinction as indexed by the amount of signal-specific spontaneous recovery of the extinguished response

Subject	(A) Extinction Day 1 SSRE	(B) Extinction Day 3 SSRE	(C) Spontaneous Recovery SSRE	(D) Final Strength of Extinction*
Amb 10	0.9	4.6	8.8	-113 *
Amb 11	14.1	3.4	2.0	-13
Amb 13	1.4	-0.6	0.3	+42
Amb 14	1.9	0.7	-0.5	-99
Amb 15	15.1	4.7	35.5	+297
Amb 16	26.8	6.9	10.6	+19
Amb 17	2.7	-0.4	3.3	+119
Amb 18	35.6	15.3	14.8	-2
Amb 19	12.7	1.2	-0.6	-16
Amb 20	6.7	3.3	1.3	-62
Amb 21	17.1	2.8	0.4	-17

Values in columns A–C are “SSRE” indices [see EQ. 3 and EQ. 4]. Actual Extinction (Column D) is the percentage of the extinguished response (i.e., the change in SSRE values in extinction) that returns in spontaneous recovery (i.e., the change in SSRE values in spontaneous recovery) [see EQ. 5; expressed by column heading: $D = (C - B / B - A) \times 100\%$].

* “Final Strength of Extinction” is expressed as the percentage of the behavioral extinguished response that spontaneously recovers after a 2-day delay. For example, 0% indicates no behavioral recovery and 100% indicates complete response recovery. A positive value greater than 100% indicates that the bar-pressing response magnitude during the spontaneous recovery test exceeded the original amount of bar-pressing during the first extinction session (Day 1). Negative values indicate continued extinction rather than response recovery. Thus, the more negative the value, the stronger the extinction.

Table 2Representational plasticity of signal-specific A1 area (5.0 kHz $\pm$ 0.25 octaves)

Subject	(A) TLI	(B) Estimated amount of A1 signal-area* i.e., <i>before</i> Extinction	(C) Determined amount of A1 signal-area i.e., <i>after</i> Extinction	(D) Estimated loss in signal-area i.e., column B – column C
Amb 10	0.9	19.6	5.8	13.9
Amb 11	0.3	12.1	8.0	4.1
Amb 13	0.2	10.8	10.3	0.5
Amb 14	0.5	14.6	6.1	8.5
Amb 15	0.4	13.3	15.5	–2.2**
Amb 16	0.0	8.3	2.2	6.1
Amb 17	0.2	10.8	8.2	2.6
Amb 18	0.2	10.8	4.5	6.3
Amb 19	0.0	8.3	3.6	4.7
Amb 20	0.4	13.3	7.9	5.5
Amb 21	0.2	10.8	10.6	0.2

* Estimated from the brain–behavior correlation shown in Fig. 2, (EQ. 1): Signal-Specific A1 Area = $0.1262 \times \text{TLI} + 0.0828$.

** Negative value identifies the only case in which signal-area was not lost after extinction, but rather gained 2.2% of A1 area.