

Interactions Between Discrimination and Control of EEG Alpha

Jon A. Frederick^{1*}, Kelli N. Dunn¹, and Thomas F. Collura^{2,3}

¹Middle Tennessee State University, Murfreesboro, Tennessee, USA

²BrainMaster Technologies Inc., Bedford, Ohio, USA

³Brain Enrichment Center, Bedford, Ohio, USA

Abstract

The relationship between discrimination and control of physiological states is largely unexplored, although it is often suggested that this relationship is important for the mechanism of action of biofeedback. This pilot study examined 6 participants given seven sessions of alpha discrimination training combined with standard neurofeedback “control” training. Four subjects achieved five criterion (binomial $p < .05$) sessions in the discrimination task. The discrimination task performances correlated significantly with performance in the amplitude control task. Evidence that some subjects can use the intertrial interval (ITI) to predict the correct responses in the discrimination task led to an examination of how ITIs were distributed with respect to success (correct or incorrect) and type of trial (same or different from previous) in these and 40 additional subjects from archival data (Frederick, 2012). This analysis found that some information about the correct responses was conveyed by the ITI, but participants made relatively little use of this information. However, the criterion discrimination sessions showed dramatic changes in the distribution of ITIs in the present (but not the archival) study, suggesting that participants were controlling their electroencephalogram (EEG) during these sessions. These findings provide preliminary evidence of generalization of skills between these two tasks.

Keywords: discrimination learning; perceptual motor processes; consciousness states; electroencephalography; biofeedback; neurotherapy

Citation: Frederick, J. A., Dunn, K. N., & Collura, T. F. (2015). Interactions Between Discrimination and Control of EEG Alpha. *NeuroRegulation*, 2(3), 126–136. <http://dx.doi.org/10.15540/nr.2.3.126>

***Address correspondence to:** Dr. Jon A. Frederick, Middle Tennessee State University, P.O. Box 87, 1301 East Main Street, Murfreesboro, TN 37132-0001, USA. Email: Jon.Frederick@mtsu.edu

Edited by: John Davis, PhD, McMaster University, Ontario, Canada

Copyright: © 2015. Frederick et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC-BY).

Reviewed by: John Davis, PhD, McMaster University, Ontario, Canada
Larry Stevens, PhD, Northern Arizona University, Flagstaff, Arizona, USA

Introduction

One of the first demonstrations of operant conditioning of the electroencephalogram (EEG) was a study that showed how human subjects could be trained to discriminate high from low alpha (8–12 Hz) amplitude states (Kamiya, 1962, 1968, 2011). In an operant discrimination task, the experimenter uses a random schedule to decide whether the next trial will be high or low, and provides a prompt when the EEG spontaneously exceeds a threshold. Subjects respond *high* or *low* and are immediately informed whether their response is correct. Kamiya (1968) noticed how participants trained to discriminate alpha subsequently performed better than naive subjects at controlling their alpha amplitude in a standard neurofeedback task—where

subjects were rewarded when alpha amplitude exceeded a threshold. This observation was important, because it suggested a relationship between the different psychological constructs trained by the two tasks.

Frederick (2012, *in press*) argued that EEG operant discrimination generally trains observation or awareness of brainwave states, while standard neurofeedback (or “control training”) trains volition or control of these states. It is commonly argued that increasing awareness of subtle subjective correlates of physiological states is central to the mechanism of action of biofeedback (Brenner, 1974; Congedo & Joffe, 2007; Frederick, *in press*; Olson, 1987; Plotkin, 1981). However, despite a half a century of evolution in neurofeedback since Kamiya’s

discovery, there have been very few studies of EEG state discrimination (e.g., Cinciripini, 1984; Kotchoubey, Kubler, Strehl, Flor, & Birbaumer, 2002). Kamiya's initial result was only recently replicated (Frederick, 2012).

Thus, the relationship between the skills trained by EEG state discrimination training and control training remains poorly understood. Control training was observed to facilitate subsequent discrimination training for the sensorimotor rhythm (12–15 Hz; Cinciripini, 1984), and slow cortical potentials (Kotchoubey et al., 2002). However, the only known demonstration of discrimination training facilitating learning of a physiological control task was seen in a peripheral vasomotor response by Fudge and Adams (1985). Kamiya's (1968) report of facilitation of alpha control by prior discrimination training was anecdotal and did not include quantitative data.

The present report describes a pilot study to explore the relationship between alpha discrimination and control. Initially, the hypothesis was that dividing session time equally between state discrimination training and standard neurofeedback training would result in greater learning of both tasks than training of either task alone. This began with a preliminary study of the two-tasks combined condition. With limited resources to run subjects and sessions, it was reasoned that before running the individual tasks separately as controls, a high level of learning should first be seen in the combined condition. In fact, the level of learning was only moderate, prompting a redesigned study that is currently in progress. While facilitation of learning was not directly tested by the pilot study design, the study design did test a more general hypothesis, that there would be a correlation in performance of the two tasks. While such a correlation is not sufficient to demonstrate generalization between the skills involved in the two tasks, it is necessary. It was also predicted that subjects would show a learning curve for both tasks, with better performance on the final sessions than the early sessions.

The present study also sought to resolve technical issues in the design of the state discrimination protocol. A review of the data from an earlier discrimination study (Logsdon, Cox, West, & Frederick, 2013) found that one subject had spuriously achieved significant performance by repeating the previously correct response on trials with very short (4–8 s) intertrial intervals (ITIs). Orne and Wilson (1978) warned about how, with fixed thresholds, the length of the interval between prompts could convey some information about the

stimulus, allowing subjects to respond correctly without truly discriminating their physiological states. For instance, if a subject's EEG amplitude changes relatively slowly, then a brief time between trials could be used by subjects to correctly infer that a given trial would be the same (high or low) as the last one. In the Logsdon et al. (2013) data, this advantage was only seen for intervals of less than 8 s between trials (Jon Frederick, unpublished observations). Therefore, in the present study, the minimum ITI was set for 8.1 s, and the null hypothesis was that there would be no advantage to perseverative responding on short ITI trials. To allow further comparison, the same hypothesis was examined in archival data (Frederick, 2012) from 40 subjects.

Method

Participants

With approval of the Middle Tennessee State University Institutional Review Board, 6 participants (3 female; age 18–44) were recruited from students at Middle Tennessee State University and the surrounding community. Participants were compensated \$10 per session after their final session. Participants were required to have a peak alpha frequency (peak median alpha amplitude between 8–12 Hz; PAF) evident in a 60-s baseline recording.

Measurements and Apparatus

Skin at the recording sites was prepared to bring impedance below 10 k Ω , with no site greater than twice the others. Taking into account both the comfort of participants and modern amplifier input impedances (Feree, Luu, Russell, & Tucker, 2001), impedances up to 15 k Ω were sometimes accepted if repeated preparations did not bring them lower. Tin electrodes were attached to the parietal midline (Pz). Left and right ears were randomly assigned to reference and ground each session.

EEG was recorded with a BrainMaster Atlantis amplifier and BrainMaster 3.7i software using the default settings (Butterworth filter order 6; default passband 0.5–64 Hz; peak-to-peak amplitude scale; 60 Hz input notch filter; 256 samples per second). Rewards were controlled with a BrainMaster Event Wizard protocol where the alpha signal had a damping factor of 10, with sustain rewarded criterion and refractory period set to zero.

The alpha band was defined as each subject's peak alpha frequency (PAF) plus or minus 2 Hz (Klimesch, 1999). For example, if the PAF were 11

Hz, the alpha band was then defined as 9–13 Hz. Clear alpha peaks were seen in all subjects. If a consistent peak alpha frequency were seen across three or more sessions, that value was used in subsequent sessions even if the participant deviated from that value on a subsequent day.

For the alpha amplitude “control” task, the experimenter watched a 25-s event trend window of the filtered alpha amplitude and was allowed to change the reward threshold in real time. BrainMaster 3.0 displayed a running average of the percent time in reward for the most recent 60 s. The experimenter attempted to maintain the participant’s percent time in reward close to 30%. About every 20 seconds, the threshold was adjusted by 0.5 to 1.0 μV if the percent time in reward was less than 10% or greater than 40%. Effort was taken to avoid changing the threshold when the subject’s alpha amplitude was close to it.

Alpha amplitudes from the control task were visually examined as 1-min averages alongside averages in the delta (0–3 Hz) and hibeta (20–30 Hz) bands. One-minute segments with excessively high delta or hibeta amplitudes were assumed to result from muscle artifact and excluded from the analysis.

For the discrimination task, Fourier-transformed amplitudes for each 1 Hz band from 1–32 Hz were passed to a dynamic link library using shared memory, and then sampled in 10 times per second by custom real-time software (*Introspect*, written in C++), which recorded both EEG and task responses. The sum of amplitudes in lodelta (0.5–2 Hz) and hibeta (23–32 Hz) were each continuously monitored as artifact channels, and recording and task were automatically suspended (and an artifact warning tone played) whenever either value exceeded a threshold.

Frederick (2012) identified several signal parameters that resulted in significantly better discrimination of EEG alpha. Subjects discriminated absolute amplitudes better than relative amplitudes; 5 Hz bandwidths surrounding the peak alpha frequency better than 1 Hz bandwidths; 2- and 4-s stimulus durations (EEG smoothing averages) better than 1-s durations; and stimulus magnitudes far from the median (below the 10th and above the 90th percentile) better than more moderate stimulus magnitudes (near the 30th or 70th percentile). However, longer intertrial intervals are required when waiting for signals with extreme durations or magnitudes. This results in a trade-off between signal quality and the number of opportunities for

learning (trials per minute) that can be administered during a session. While the use of absolute amplitude and 5 Hz bandwidths were clearly indicated, the use of 2-s stimulus durations and moderate stimulus magnitudes (30th and 70th percentiles) were needed to achieve a desired 3–4 prompts per minute.

The dimensions of the two tasks were made as similar as possible. For instance, the target of 30% time in reward for the control task corresponded to the 30th and 70th percentile thresholds to trigger a prompt in the discrimination task. A 60-s baseline was used in the discrimination task because 60 s was the maximum sliding baseline for the percent time calculations provided by the BrainMaster event wizard.

Procedure

After obtaining informed consent, participants sat in a reclining chair with eyes closed in a dimly lit, sound-attenuated room. Participants were instructed about the nature of muscle artifact and strategies to relax and minimize it.

Each session consisted of 20 min of alpha control training followed by about 20 minutes of alpha discrimination training. The alpha control training started with 5 min during which high alpha amplitude was rewarded (the “enhance” condition), then 5 min during which low alpha amplitude was rewarded (the “inhibit” condition). These two conditions were then repeated. To control for spontaneous shifts in baseline alpha amplitude, the measure of performance in this task was defined as the percent amplitude difference between the high and low amplitude conditions.

Each run of 5 min in the control task was prefaced by saying, “Now you will be rewarded for increasing (or decreasing) alpha. Are you ready?” After at least one run of a condition, the threshold from the most recent run was re-used, and then adjusted as needed to bring the reward percentage near 30%.

Before the discrimination task, participants were provided with a set of written instructions as described (Frederick, 2012). The instructions explained that EEG alpha usually means a relaxed but alert state with eyes closed; that high alpha might be increased by clearing, emptying, or quieting the mind, or disconnecting from mental contents. Low alpha was described as the opposite, the presence of imagination, attention to sensory details, thought, intention, or inner speech. It was emphasized that their own experiences before and

after each prompt were equally or more important, as these instructions were only rough guidelines. A strategy was suggested that if no discrimination prompt was received in a long time (e.g., 30 s), to try changing their mental state and see if that evoked a prompt.

Alpha state discrimination training consisted of a median of 40 (min 30, max 60) trials. A 60-s eyes-closed baseline EEG was recorded each session. During the task, each EEG epoch was ranked among a percentile distribution of alpha amplitudes of the most recent 60 s initially derived from the baseline recording. The baseline was updated with each response, or whenever the experimenter pressed the pause button. Triggering of new prompts was suspended for 8 s after each prompt or after resuming from a pause. A random number generator determined in advance whether each trial would be high or low. A prompt tone then sounded whenever the alpha band amplitude exceeded a critical threshold difference from the median of the baseline. Recording was then suspended until the subject responded.

The critical threshold for triggering a prompt was set at the 30th percentile for low alpha trials, and the 70th percentile for high alpha trials. Subjects responded “high” or “low” with a keypress response, and received immediate verbal feedback after each trial whether the response was correct. Software was programmed to exclude runs of six or more of

the same (high or low) trial type, although participants were not informed of this constraint.

Analysis

Performance in the EEG alpha amplitude control task was analyzed as the percent difference, or the amplitude difference between the high and the low conditions, divided by the overall average amplitude. The choice of a percent difference rather than a raw amplitude difference served to control for variance in factors such as skull thickness, which may have masked real differences in achievement in the task.

Performance in the EEG alpha discrimination task was analyzed in terms of probability under the binomial theorem, where a criterion performance was defined as a significant number correct with binomial $p < .05$. Four subjects achieved a total of five criterion sessions in the discrimination task.

Results

There was a strong association between performance in the discrimination task and percent difference between the high and low conditions of the amplitude control task. Among seven sessions, the five criterion performances on the discrimination task all occurred on sessions with the first, second, or third (median second, of 7) highest amplitude difference in the control task (Table 1).

Table 1

Percent Amplitude Difference Between High and Low Conditions in the Control Task, and Percent Correct in the Discrimination Task, Over Seven Sessions

Subject	Task	Session						
		1	2	3	4	5	6	7
mt003	control	MD	MD	-10	-16	22	7	11
	discrim	36	50	50	58	90**	48	60*
mt004	control	-4	-7	15	0	8	-2	1
	discrim	51	43	52	49	62.5*	52	57
mt005	control	7	-6	-14	20	15	6	13
	discrim	58	55	53	35	65*	53	40
mt007	control	-6	7	21	-2	8	24	18
	discrim	51	50	55	56	48	60	69**
average	control	-1	-2	3	0	13	9	11
	discrim	49	49	52	50	66	53	57

Note. *Denotes binomial $p < .05$; **Denotes binomial $p < .01$; MD, missing data.

The covariation in performance between the two tasks is clearly seen when the mean discrimination performance is plotted as a difference from 50 percent alongside the control task performance (Figure 1). Quantitatively, the mean within-subject correlation between the two tasks had Pearson's $r = .34$, which was not significantly different from zero, $t(3) = 1.49$, $p = .12$, $d = 0.744$. However, when these data were analyzed categorically (reducing the contribution of random error from below chance scores on the two tasks), the mean point-biserial correlation between criterion performances on the discrimination task and above average performance on the amplitude control task was $r = .57$, $t(3) = 3.96$, $p = .014$, $d = 1.98$.

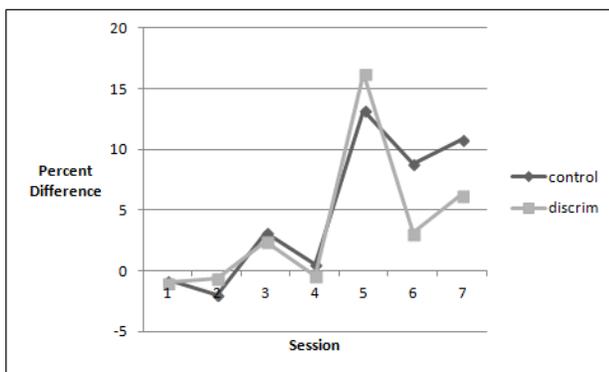


Figure 1. Covariation in performance of alpha control task (mean difference in percent amplitude between enhance and inhibit conditions) and alpha discrimination task (difference from 50% correct). Each line represents the mean of $n = 4$ subjects.

There was a strong learning curve effect, where performances in both tasks were nonrandom with respect to session number. The mean within subject correlation between session number and above average task performance for the control task was $r = .43$, $t(3) = 3.77$, $p = .016$, $d = 1.89$. The mean correlation between session number and criterion performance for the discrimination task was $r = .40$, $t(3) = 2.91$, $p = .031$, $d = 1.45$. Thus, it is possible that there was no causal relationship between the two performance variables, but an independent effect of learning caused both variables to covary over time.

However, in three cases performance declined in sessions after the first criterion performance (after session 5; see Table 1). In these cases, the mean point-biserial correlation between the two tasks across sessions 5–7 was $r = 0.67$, $t(2) = 4.00$, $p = .029$, $d = 2.31$.

As previously reported (Frederick, 2012), shorter times between sessions appeared to improve discrimination performance. The five criterion sessions had a mean 4.4 days since the previous session, compared to 8.1 for the remaining sessions, $t(31) = 1.806$, $p = .040$, $d = 0.877$. A similar effect was not seen for the control task.

Information Conveyed by Intertrial Intervals

The distribution of intertrial intervals (ITIs) and same versus different trials (from previous trial types) was examined to determine whether ITIs were conveying information about the type of trial (Orne & Wilson, 1978). The 34 non-criterion sessions were first studied as a control condition. Not including first trials when the ITI was undefined, there were 1486 discrimination trials in these sessions. The mean ITI across all trials was 16.6 s (SD 10.8 s). There were 726 or 48.9% of trials that were the same versus 759 or 51.1% that were different from the previous (high or low) trial type. This difference was expected because runs of six or greater identical trial types were prevented by the Introspect software.

The pattern of same and different trials with respect to ITI was examined to see if there was an advantage to responding *same* as the previous correct trial type on short ITI trials, or to responding *different* on long ITI trials. To give all subjects equal weight (despite differing numbers of trials), the percentage of trials in each ITI was first computed within each subject and then the mean percentage was computed across subjects. This allowed for accurate degrees of freedom and valid statistical comparisons ($n = 6$ subjects, not $n = 1,486$ trials).

Consistent with the concern that information about the type of trial was conveyed by the ITI, the percent of same trials (% Same) was significantly higher in the ≤ 8.1 s category than the > 45.0 s category, one-tailed $t(5) = 4.28$, $p = .0039$, $d = 1.75$ (54.3 vs. 26.6% of trials, Table 2). The advantage (% Adv) of using this information was significantly different from zero for ITIs of 8.1 s, $t(5) = 2.09$, $p = .046$, $d = 0.85$; 30.1–45.0 s, $t(5) = 2.02$, $p = .049$, $d = 0.83$; and > 45.0 s, $t(5) = 3.65$, $p = .0074$, $d = 1.49$. One-tailed tests were used for these comparisons because first, there were theoretical reasons to predict an advantage to *same* responding on the shortest (8.1 s) ITI trials, and *different* on the longest (> 45.0 s) ITI trials (Cott, Pavloski, & Black, 1981; Orne & Wilson, 1978). Since it was unclear where the boundary between short and long was between these extremes, it was reasoned that an effective $p = .10$ alpha-level was justified if a one-tailed test was not, because type II error for these comparisons

amounts to type I error for the validity of the discrimination paradigm.

Including all three categories, then, the ITI may have conveyed useful information on 28.2% of trials. The total weighted advantage across all ITIs was 3.8%, meaning that a subject with perfect information and a perfect sense of timing could on have scored 53.8% using ITI information alone.

However, this information could only contribute to scores on the task if participants had a large response bias in the correct direction. The response bias (% Bias) is quantified as the difference between success on same trials (Same S) and different trials (Diff S). Overall, participants tended to show a bias toward *same* responding, where the average score on same trials was 57.1% and the average score on different trials was 41.8%. To use ITI information to

score effectively in the task, % Bias must approach 100%. For instance, a subject who responded 100% *same* on 8.1-s ITI trials where % Adv was 4.3% could expect an average score of 54.3% on those trials (100% correct on same trials, 0% on different trials). Subjects showed a significant response bias for ITIs of 8.1 s, one-tailed $t(5) = 2.16$, $p = .042$, $d = 0.88$; 8.2–9.0 s, $t(5) = 2.17$, $p = .041$, $d = 0.88$; 9.1–15.0 s, $t(5) = 2.69$, $p = .022$, $d = 1.10$; and 15.1–30.0 s, $t(5) = 2.15$, $p = .042$, $d = 0.88$. However, the percent points gained (% Gain) from use of the ITI is found by multiplying % Bias by % Adv times % Obs. As would be expected from noncriterion sessions where the overall mean score was 49.5%, % Gain was near zero for all ITIs. Table 2 could be summarized by saying that subjects had a large bias where it gained them no advantage, and no bias where the advantage was large.

Table 2

Information Conveyed by the Time Between Prompts and Discrimination Task Success for Noncriterion Sessions

ITI, s	% Obs	% Same	% Adv	Same S	Diff S	Avg S	% Bias	% Gain
8.1	17.8	54.3	4.3*	59.5	37.5	49.8	22.1*	0.2
8.2–9.0	7.5	51.1	1.1	57.2	30.7	44.2	26.6*	0.0
9.1–15.0	33.3	51.5	1.5	56.5	39.5	48.1	17.1*	0.1
15.1–30.0	31.0	47.7	-2.3	57.0	45.8	50.8	11.2*	-0.1
30.1–45.0	6.6	37.7	-12.3*	60.2	53.8	54.9	6.4	-0.1
> 45.0	3.8	26.6	-23.4**	46.1	50.0	50.9	-3.8	0.0

Note. 6 subjects; 1,486 trials. ITI, intertrial interval in seconds; % Obs, percent of trials in ITI category; % Same, percent of trials same as the correct response for the previous trial; % Adv, possible increase in score on these trials or advantage by using ITI—negative sign indicates advantage for responding *different*; Same S, percent success on same trials; Diff S, percent success on different trials; Avg S, average success; % Bias, difference between Same S and Diff S; % Gain, total percent points resulting from % Bias given %Adv and % Obs. *Denotes one-tailed t -test $p < .05$. **Denotes one-tailed t -test $p < .01$.

The overall percentages of same versus different trials were similar for the five criterion sessions (195 trials; 48.7% same, 51.3% different; Table 3). The mean ITI across all 195 trials was 19.1 s (SD 12.9 s). A much greater difference of % Same between short (8.1–9.0 s) and longer (> 9.0 s) ITIs was seen in the criterion sessions. The difference in % Same between 8.1 s and > 45 s was highly significant, $t(3) = 10.64$, $p < .001$, $d = 5.32$. There was a significant advantage to responding *same* on 8.1-s trials (33.6%), $t(3) = 5.80$, $p = .0051$, $d = 2.90$; and *different* for 15.1- to 30.0-s trials (11.9%), $t(3) = 5.80$, $p = .0051$, $d = 2.40$. The advantage approached significance for the 8.2- to 9.0-s trials (27.1%), $t(3) = 1.72$, $p = .092$, $d = 0.86$; and of 10.4% for the > 45-s trials (10.4%), $t(3) = 1.67$, $p = .097$, $d = 0.83$. These four categories included 59% of trials.

The distribution of % Same was also significantly different from the noncriterion sessions for 8.1 s (83.6 vs. 54.3%), two-tailed between groups $t(8) = 5.56$, $p = .00053$, $d = 3.59$; and 15.1–30.0 s (38.1 vs. 47.7%), $t(8) = 2.54$, $p = .034$, $d = 1.64$; and approached significance for 8.2–9.0 s (77.1 vs. 51.1%), $t(8) = 2.04$, $p = .075$, $d = 1.32$. The effect size was moderate to large ($d = 0.57$ to $d = 0.89$) for the others.

Given the distribution of % Same in Table 3, the total weighted advantage for responding based on ITIs alone was 13.6%. A participant with perfect information and perfect timing could, then, score 63.6% using ITI information alone.

Subjects in the criterion sessions showed a strong overall response bias toward perseverative or *same*

responding. Whereas the overall average score was 67.0% for the criterion sessions, the mean score was 82.7% on same trials (Same S) versus 52.8% on different trials (Diff S). The % Bias was significant for 9.1–15 s, one-tailed $t(3) = 2.36$, $p = .049$, $d = 1.18$; and approached significance for 8.1–9.0 s, $t(3) = 1.98$, $p = .070$, $d = 0.99$. Gain scores were computed for individual subjects. Three subjects had no Diff S trials for either the 8.1 or 8.2–9.0 categories. After combining these categories all

subjects had at least one trial, allowing for computation of % Bias and % Gain. When summed across all ITIs, the mean % Gain across 4 subjects was 2.0%, which was not significant, $t(3) = .56$, one-tailed $p = .31$, $d = 0.28$.

One subject, however, had a % Gain of 10.3 (explaining more than half of his overall score of 69%) of which 9.6 points were earned from positive advantage and positive bias from 8.1- to 15.0-s ITIs.

Table 3

Information Conveyed by the Time Between Prompts and Discrimination Task Success for Criterion Sessions

ITI, s	% Obs	% Same	% Adv	Same S	Diff S	Avg S	% Bias	% Gain
8.1	18.3	83.6	33.6**	79.9				
8.2–9.0	5.1	77.1	27.1	100.0				
8.1–9.0	23.4	80.2	30.2	83.7	37.5	74.5	46.2	3.3
9.1–15.0	31.7	43.1	-6.9	80.7	32.4	50.1	48.3*	-1.1
15.1–30.0	27.2	38.1	-11.9**	80.0	66.5	71.1	13.5	-0.4
30.1–45.0	9.3	53.0	3.0	79.2	71.7	72.6	7.5	0.0
> 45.0	8.3	39.6	-10.4	75.0	93.8	89.6	-18.8	0.2

Note. 4 subjects; 195 trials. ITI, % Obs, % Same, % Adv, Same S, Diff S, Avg S, % Bias, % Gain, see Table 2 caption. *Denotes one-tailed t -test $p < .05$. **Denotes one-tailed t -test $p < .01$.

It was of interest whether the distribution of ITIs had conveyed information about the stimuli in a previous study with larger sample size (Frederick, 2012). All noncriterion sessions were first examined from subjects who had achieved criterion in this archival data set. Data from one subject and 28 sessions were excluded because they had the minimum ITI set higher than 4.1 s, leaving 38 subjects, 152 sessions, and 14,279 trials (Table 4). The mean ITI for these sessions was 15.5 s (SD 16.1 s) and the average score was 50.1%.

These data had 6,969 or 48.8% same trials versus 7,310 or 51.2% different trials. These data also showed a pattern of greater % Same trials for short ITIs and lower percent same trials for longer ITIs. Although small, the difference in % Same between the shortest (4.1 s) and longest (> 45.0 s) ITIs bordered on significance with $n = 38$ (54.7 vs. 46.1%), $t(37) = 1.68$, $p = .051$, $d = 0.48$. There was

a significant advantage for responding based upon ITI for 4.1 s, $t(37) = 1.92$, $p = .031$, $d = 0.31$; 5.1–8.0 s, $t(37) = 1.93$, $p = .031$, $d = 0.31$; 8.1–15.0 s, $t(37) = 3.18$, $p = .0015$, $d = 0.52$; and 15.1–30.0 s, $t(37) = 2.46$, $p = .0092$, $d = 0.40$.

The total weighted advantage was 3.1%, meaning that information conveyed by ITIs could be used to score up to 53.1% correct.

The % Bias was significant for 4.1 s, $t(35) = 6.64$, $p < .001$, $d = 1.106$; 4.2–5.0 s, $t(34) = 3.40$, $p < .001$, $d = 0.57$; 5.1–8.0 s, $t(37) = 7.61$, $p < .001$, $d = 1.23$; 8.1–15.0 s, $t(37) = 5.87$, $p < .001$, $d = 0.95$; 15.1–30.0 s, $t(37) = 3.76$, $p < .001$, $d = 0.61$; and approached significance for > 45.0 s, $t(33) = 1.64$, $p = .055$, $d = 0.28$. However, owing to small % Adv requiring 100% bias to fully take advantage, this response bias resulted in near zero % Gain in score for these noncriterion sessions.

Table 4*Information Conveyed by the Time Between Prompts and Discrimination Task Success for Noncriterion Sessions*

ITI, s	% Obs	% Same	% Adv	Same S	Diff S	Avg S	% Bias	% Gain
4.1	10.3	54.7	4.7*	63.5	33.8	51.4	28.6**	0.1
4.2–5.0	6.0	53.3	3.3	62.0	40.5	51.1	24.9**	0.0
5.1–8.0	23.1	52.8	2.8*	64.4	38.6	52.7	20.8**	0.3
8.1–15.0	27.9	46.9	-3.1**	58.7	41.4	49.2	17.3**	-0.2
15.1–30.0	21.3	47.2	-2.8**	54.6	43.3	48.4	11.3**	-0.1
30.1–45.0	6.3	48.0	-2.0	52.2	48.7	49.0	4.9	-0.1
> 45.0	5.0	46.1	-3.9	45.7	55.9	50.8	-9.7	0.1

Note. Archival data, $n = 38$ subjects; 14,279 trials (Frederick, 2012). ITI, % Obs, % Same, % Adv, Same S, Diff S, Avg S, % Bias, % Gain, see Table 2. *Denotes one-tailed t -test $p < .05$. **Denotes one-tailed t -test $p < .01$. % Bias and % Gain were computed only from subjects who had at least one same and one different trial in a given ITI category, unlike %Obs, %Adv, Same S, or Diff S. Thus the % Bias and % Gain reported in Tables 4 and 5 are not exactly equal to those computed from other columns in the table.

The distribution of ITIs versus same/different trials and performance was also examined in the criterion sessions from the same archival subject group (Frederick, 2012; Table 5). Two sessions and 23 sets (parts of sessions) were excluded because the minimum ITI was set higher than 4.1 s, leaving 76 sets and 16 sessions from 36 subjects and 4,532 trials. The mean ITI was 14.7 s (SD 14.8 s), and the mean score was 69.2%. The difference in % Same between the shortest (4.1 s) and longest (> 45.0 s) ITIs approached significance (54.2 vs. 46.2%), $t(28) = 1.34$, $p = .094$, $d = 0.25$.

Unlike the four criterion subjects from the present study, in the archival subjects there were no significant differences in % Same between the criterion and noncriterion data. Only the 15.1–30.0 s ITIs showed a significant difference from 50% (45.5%), $t(36) = 1.77$, $p = .042$, $d = 0.30$.

The total weighted advantage was 3.3%. A subject with perfect information and a perfect sense of timing could thus score 53.3% using ITI information alone.

Participants showed a strong bias toward *same* responding, scoring 74.3% overall on same trials and 61.6% on different trials. % Bias was significant for 4.1 s, one-tailed $t(31) = 3.03$, $p = .0025$, $d = 0.53$; 4.2–5.0 s, $t(25) = 3.24$, $p = .0017$, $d = 0.63$; and 5.1–8.0 s, $t(33) = 3.60$, $p = .00052$, $d = 0.62$; and

approached significance for 30.1–45.0 s, $t(23) = 1.49$, $p = .075$, $d = 0.30$.

The % Gain was significantly different from zero for 4.1 s, $t(31) = 2.44$, $p = .010$, $d = 0.43$; and 4.2–5.0 s, $t(25) = 2.01$, $p = .027$, $d = 0.39$. The overall mean % Gain was 0.9%, which was significant, $t(35) = 1.69$, $p = .0499$, $d = 0.28$. Among 36 gain scores, there were an equal number (18) positive and negative. If the distribution of negative gain scores is assumed to represent random variation, the absolute value of scores beyond the 90th percentile (-2.23%) could be used as a one-tailed $p < .05$ test for suspicious positive gain scores. Six subjects had gain scores in this category: 3.0, 3.6, 4.6, 4.8, 11.3, and 11.9% (by contrast, the six most negative gain scores were -1.4, -1.7, -2.1, -2.2, -2.2, and -2.7%). Among the six subjects with suspicious positive gain scores, subtracting the positive gain would have resulted in performances below a binomial $p = .05$ for two subjects and between $p = .05$ and $p = .01$ for two subjects.

The 11.9% score belonged to a subject who scored 94.1% overall (explaining at most 27% of his score; $p < .001$ after removing this effect). However, the 11.3% gain score could explain as much as 48% of that subject's score of 73.5%. Nearly all of the 11.3% was earned from positive advantage and positive bias on short (4.1–8.0 s) ITI trials. Subtracting this gain resulted in binomial $p = .09$.

Table 5
Information Conveyed by the Time Between Prompts and Discrimination Task Success for Criterion Sessions

ITI, s	% Obs	% Same	% Adv	Same S	Diff S	Avg S	% Bias	% Gain
4.1	14.9	54.2	4.2	78.2	56.8	71.6	23.3**	0.5*
4.2–5.0	6.1	56.8	6.8	80.0	54.9	71.0	26.0**	0.3*
5.1–8.0	21.0	52.2	2.2	79.7	59.6	71.2	19.3**	0.1
8.1–15.0	26.0	52.3	2.3	69.7	63.4	67.4	7.3	0.1
15.1–30.0	22.2	45.5	-4.5*	69.5	65.4	66.8	3.6	0.0
30.1–45.0	6.2	50.3	0.3	77.9	63.2	71.0	11.2	-0.1
> 45.0	3.6	46.2	-3.8	72.5	64.0	69.9	3.1	0.1

Note. Archival data, $n = 36$ subjects; 76 sets, 16 sessions; 4,532 trials. ITI, % Obs, % Same, % Adv, Same S, Diff S, Avg S, % Bias, % Gain, see Table 2. *Denotes one-tailed t -test $p < .05$. **Denotes one-tailed t -test $p < .01$.

Discussion

This study showed a strong association between performance in an EEG alpha control task and an EEG alpha discrimination task over the course of seven sessions. The best performances on the discrimination task tended to occur on the same days when participants achieved the greatest differences between the enhance and inhibit conditions in the control task.

However, performance also showed a significant correlation with session number for both tasks, suggesting a learning curve effect. This learning curve effect might have been a confound for interpreting some generalization of skills between the two tasks. However, results for several subjects after the peak of the learning curve served as a control, actually showing a higher correlation when there was no learning effect.

The use of categorical data for computing correlations (criterion performance on the discrimination task and above average performance on the amplitude control task) rather than raw interval data could be argued to be a “cherry-picking” of analytical methods, since the correlation between raw performance scores was not significant. The rationale was that discrimination scores below 50% and amplitude differences below 0% are functionally equal, so nominalizing these data reduces error variance. However, all statistics from a pilot study with $n = 4$ are to be interpreted with caution and only as suggestions for further research.

The overall performance on the discrimination task in this subject group seemed to be less successful than that observed in Frederick (2012). Whereas 56% of

subjects achieved $p < .01$ performance by the seventh session in that study, only 2 out of 6 subjects achieved $p < .01$ in this study.

There were several differences between the two studies that may have contributed to reduced performance in the discrimination task. Most notably, only half rather than the entire session time was spent practicing the discrimination task in this study. Further, Frederick (2012) varied the prompt threshold within the 0–30 and 70–100 percentile ranges (mean difference from 50th percentile, 34.3, $SD 8.9$), whereas they were set at constant 30th and 70th percentiles in the present study (mean difference 28.7, $SD 9.5$).

However, several other factors were expected to improve performance in this study. The present study used absolute (not relative) amplitude for all trials rather than about half of trials, and 5-Hz alpha bandwidths for all trials rather than about one third of trials. This study also consistently used the same parameters for absolute amplitude, bandwidth, stimulus duration, and location, whereas varying these parameters in Frederick (2012) might have confused participants.

Another factor that may have influenced performance was that participants received verbal rather than automated feedback on the discrimination task in this study. Modeling the importance of the client-therapist relationship, it was reasoned that having the experimenter say whether a response was correct would improve motivation by conveying that a person was in the room who cared. However, it was noted that this verbal feedback often took 200 or more milliseconds to initiate after the response; and created opportunities for variance in how this information

was perceived. Sherlin et al. (2011) noted the importance of minimizing the delay of reinforcement after the behavior, citing studies showing how learning can be adversely affected by latencies exceeding 250–350 ms (Felsing, Gladstone, Yamaguchi, & Clark, 1947; Grice, 1948). The motivational advantage of personal feedback may, then, be outweighed by the need to present feedback quickly and consistently.

The observation that ITIs conveyed significant information about the types of trials in both the present and the archival study—the tendency for % Same to be greater for shorter ITIs and lower for the longer ITIs—represents a flaw in the design of both studies. Frederick (2012) argued that the use of the sliding baseline avoided this complication. The relatively small (3.1 to 3.3%) advantage of using this information in that study suggests that the sliding baseline reduced, but did not eliminate this problem. The advantages seen in the present data set (3.8 and 13.6%) show that increasing the minimum ITI to 8.1 s did not reduce the problem. Subjects in both studies tended to have the greatest response biases for *same* responding for the shortest ITIs and the greatest response biases for *different* responding for the longest ITIs. However, these biases were rarely large enough to result in gains in score that were both significant and substantial relative to the total score.

Future research should take steps to reduce the possible advantage of responding based on the ITI. For instance, our discrimination task software has now been revised to automatically increase the minimum ITI whenever the difference in number of trials between same and different trial types is greater than one for 3.1–5.0, 5.1–10.0, and 10.1–15.0 s intervals. One benefit to this analysis has been the elimination of the 8.1 minimum ITI, allowing for more trials per minute or a higher prompt threshold.

An unexpected discovery in this study was the significant difference in the distribution of same versus different trials between the criterion and noncriterion sessions (Tables 2 and 3). Shorter ITIs had a much higher % Same and longer ITIs had much lower % Same. This difference was not seen in the archival study, which had not included control task training (Tables 4 and 5). One possible interpretation is that participants were successfully stabilizing and controlling their EEG—generalizing their control task skill to the discrimination session. Participants in the criterion sessions tended to persevere, not only in reporting the previously

correct state (for all ITIs), but also in maintaining it (for up to 9.0 s).

The overall bias toward *same* responding might also reflect an honest strategy for guessing one's internal state in the absence of a clear perception. Thus, if one's subjective perceptions are no different on one trial compared to the previous one, it is reasonable to assume that one's alpha amplitude hasn't changed either.

This study has provided preliminary evidence of generalization between the skills involved in EEG state discrimination and standard neurofeedback control tasks, both in the correlation of task performances and in participants' tendency to stabilize their alpha amplitude between trials in the criterion discrimination sessions (greater % Same on short ITI trials).

Future studies in this laboratory will explore more directly whether there is a causal interaction between the skills involved in these two tasks. Currently, we are assessing whether the skills generalize or transfer, by measuring performance on one task after seven training sessions on the other task.

It is argued that awareness or explicit processing is important to early stages of learning (Fitts & Posner, 1967; Gentile, 2000), but can actually decrease performance on highly practiced tasks (Beilock & Carr, 2001). It stands to reason, then, that adding discrimination training to the early stages of standard neurofeedback might increase this explicit processing and enhance learning of the standard neurofeedback task. Thus, an additional subject group in our study will receive both tasks in the same session—similar to this pilot study—to assess whether combining the two types of training results in better learning than either task alone.

References

- Beilock S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General*, *130*(4), 701–725. <http://dx.doi.org/10.1037/0096-3445.130.4.701>
- Brener, J. (1974). A general model of voluntary control applied to the phenomena of learned cardiovascular change. In P. Obrist, A. H. Black, J. Brener, & L. V. DiCara (Eds.), *Cardiovascular psychophysiology*. Chicago: Aldine.
- Cinciripini, P. M. (1984). Discrimination of sensorimotor EEG (12–15 Hz) activity: A comparison of response, production, and no-feedback training conditions. *Psychophysiology*, *21*(1), 54–62. <http://dx.doi.org/10.1111/j.1469-8986.1984.tb02317.x>
- Congedo, M., & Joffe, D. (2007). Multichannel tomographic neurofeedback: Wave of the future? In J. R. Evans (Ed.),

- Handbook of neurofeedback*. New York: The Hayworth Medical Press.
- Cott, A., Pavloski, R. P., & Black, A. H. (1981). Operant conditioning and discrimination of alpha: Some methodological limitations inherent in response-discrimination experiments. *Journal of Experimental Psychology: General*, *110*(3), 398–414. <http://dx.doi.org/10.1037/0096-3445.110.3.398>
- Felsing, J. M., Gladstone, A. I., Yamaguchi, H. G., & Clark, L. (1947). Reaction latency (StR) as a function of the number of reinforcements (N). *Journal of Experimental Psychology*, *37*(3), 214–228. <http://dx.doi.org/10.1037/h0055587>
- Feree, T. C., Luu, P., Russell, G. S., & Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Clinical Neurophysiology*, *112*(3), 536–544. [http://dx.doi.org/10.1016/S1388-2457\(00\)00533-2](http://dx.doi.org/10.1016/S1388-2457(00)00533-2)
- Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Oxford, England: Brooks and Cole.
- Frederick, J. A. (2012). Psychophysics of EEG alpha state discrimination. *Consciousness and Cognition*, *21*(3), 1345–1354. <http://dx.doi.org/10.1016/j.concog.2012.06.009>
- Frederick, J. A. (in press). EEG state discrimination and the phenomenal correlates of brainwave states. In T. F. Collura & J. A. Frederick (Eds.), *Clinician's Companion to QEEG and Neurofeedback* (annotated and with an introduction by J. Kiffer). New York: Taylor & Francis.
- Fudge, R., & Adams, H. D. (1985). The effects of discrimination training on voluntary control of cephalic vasomotor activity. *Psychophysiology*, *22*(3), 300–306.
- Gentile, A. M. (2000). Skill acquisition: Action, movement, and neuromotor processes. In J. Carr & R. Shepherd (Eds.), *Movement science: Foundations for physical therapy in rehabilitation* (2nd ed., pp. 111–187). Gaithersburg, MD: Aspen.
- Grice, G. R. (1948). The relation of secondary reinforcement to delayed reward in visual discrimination learning. *Journal of Experimental Psychology*, *38*(1), 1–16. <http://dx.doi.org/10.1037/h0061016>
- Kamiya, J. (1962). *Conditioned discrimination of the EEG alpha rhythm in humans*. Proceedings of the Western Psychological Association, San Francisco, California.
- Kamiya, J. (1968). Conscious control of brain waves. *Psychology Today*, *1*, 57–60.
- Kamiya, J. (2011). The first communications about operant conditioning of the EEG. *Journal of Neurotherapy*, *15*(1), 65–73. <http://dx.doi.org/10.1080/10874208.2011.545764>
- Klimesch, W. (1999). Full-length review: EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, *29*(2–3), 169–195. [http://dx.doi.org/10.1016/S0165-0173\(98\)00056-3](http://dx.doi.org/10.1016/S0165-0173(98)00056-3)
- Kotchoubey, B., Kübler, A., Strehl, U., Flor, H., & Birbaumer, N. (2002). Can humans perceive their brain states? *Consciousness and Cognition*, *11*(1), 98–113. <http://dx.doi.org/10.1006/ccog.2001.0535>
- Logsdon, H., Cox, C., West, K., & Frederick, J. A. (2013). First evidence of 14–18 Hz EEG state discrimination at Fz and anterior cingulate in human subjects. Presented at ISNR, Dallas, TX.
- Olson, P. (1987). Definitions of biofeedback. In M. S. Schwartz (Ed.), *Biofeedback: A practitioner's guide* (pp. 33–38). New York: The Guilford Press.
- Orne, M. T., & Wilson, S. K. (1978). On the nature of alpha feedback training. In G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 2). New York: Plenum Press.
- Plotkin, W. B. (1981). A rapprochement of the operant-conditioning and awareness views of biofeedback training: The role of discrimination in voluntary control. *Journal of Experimental Psychology: General*, *110*(3), 415–428. <http://dx.doi.org/10.1037/0096-3445.110.3.415>
- Sherlin, L. H., Arns, M., Lubar, J., Heinrich, H., Kerson, C., Strehl, U., & Sterman, M. B. (2011). Neurofeedback and basic learning theory: Implications for research and practice. *Journal of Neurotherapy*, *15*(4), 292–304. <http://dx.doi.org/10.1080/10874208.2011.623089>

Received: July 30, 2015

Accepted: September 20, 2015

Published: October 8, 2015