Failures to find suppression of episodic memories in the think/no-think paradigm

JOHN B. BULEVICH, HENRY L. ROEDIGER III, DAVID A. BALOTA, and ANDREW C. BUTLER
Washington University, St. Louis, Missouri

Anderson and Green (2001) had subjects learn paired associates and then selectively suppress responses to some of them. They reported a decrease in final cued recall for responses that subjects had been instructed not to think of and explained their data as resulting from cognitive suppression, a laboratory analog of repression. We report three experiments designed to replicate the suppression/repression results. After subjects learned a series of A–B word pairs (e.g., *ordeal–roach*), they were then asked to respond to some items and not to think of other items when shown their cues 1, 8, or 16 times. During a final recall test, they were cued with either a same (direct) probe (*ordeal–_____*) or an independent (indirect) probe (*insect–r_____*). None of our experiments showed reliable suppression effects with either the same or independent-probe tests. Suppression is apparently not a robust experimental phenomenon in the think/no-think paradigm.

Experimental psychologists have been interested in the suppression of memories almost since the birth of the field, spurred by Freud's theorizing (Freud, 1895/1957). The idea that unpleasant memories could be purged from consciousness has always been a tempting concept, both for its implications for clinical phenomena and as a target for research among experimentalists, starting with Rosenzweig and Mason (1934). The concept of repression has remained a popular psychological concept, despite very little experimental support for its existence (e.g., Holmes, 1990).

One possible problem with experimentation in repression is the lack of agreement on the definition of the concept (see Erdelyi, 1990, for a review). For example, it is unclear whether repression is the result of conscious or unconscious processes, as in Freud's original conceptualization. Following Anderson and Green (2001), in the present research we conceptualized repression as an active process—a strategy that people may use to forget—and we used the term *suppression* synonymously.

Laboratory analogs have been designed over the years to seek evidence that would support or discredit the theory of repression as either a conscious or an unconscious process (e.g., Glucksberg & King, 1967). Investigators attempted to develop experimental paradigms to directly address repression, which was mostly supported hereto-fore by anecdotal cases described by clinicians. The first attempt at investigating repression experimentally examined people's ability to recall puzzles that had been solved versus those that had not (Rosenzweig & Mason, 1934). The experimenters manipulated whether or not the subject finished the puzzle and found poorer recall for the unsolved puzzles than for the solved ones. Rosenzweig and Mason argued that not solving a puzzle would be a comparatively unpleasant memory and therefore would be less likely to be remembered, a result they did in fact obtain. The authors concluded that these findings supported Freud's concept of repression. Other researchers, however, reported the opposite pattern—that incomplete tasks are better recalled than are complete tasks. Zeigarnik (1927) is the first of these, and the *Zeigarnik effect* was named for her study. Patalano and Seifert (1994) have shown the same pattern more recently.

In the 1960s, the directed forgetting paradigm (Bjork, LaBerge, & Legrand, 1968; Weiner, 1968; see Basden & Basden, 1998, and MacLeod, 1998, for reviews) emerged as a method that could be used to study cognitive suppression. The general paradigm consisted of subjects learning a list of items with instructions telling them that they should actively forget some items and remember others or to forget one list of items and remember the next list. Subjects typically showed poorer retention for the forget items and less proactive interference for the items learned after the forget instruction, at least on recall tests (but only sometimes on recognition tests). Many researchers examined factors at work in this paradigm and the two primary ones implicated were rehearsal (to-be-forgotten items were rehearsed less often than to-be-remembered items) and segregation or grouping of the two types of items (forgotten and remembered items).

Few researchers considered directed forgetting as a possible analog of repression, but some did. For example,

Note—This article was accepted by the previous editorial team, when Colin M. MacLeod was Editor.

Weiner and Reed (1969) used a variant of this procedure to investigate repression by giving subjects a series of letter trigrams to learn with instructions either to remember and rehearse the trigrams, to remember but not rehearse them, or to forget them. They were given a secondary task of number shadowing during the retention interval after each trigram. Just before being tested on the material, subjects were instructed to recall as many items as possible no matter what instruction they had been given for the trial. Recall was poorest for items given a forget instruction, an effect that Weiner and Reed interpreted as analogous to repression. However, Roediger and Crowder (1972) proposed a simpler explanation. They showed that subjects who were instructed to forget engaged in the secondary task (counting backward, in their experiment) to a greater degree than those who were told to remember the items. The results are thus more easily explained by subjects expending more capacity during the secondary task on the forget trials than on the remember trials and making the interpolated task more difficult for that condition. The general conclusion from most observers of the directed forgetting literature is that the mechanisms at play are not akin to repression (see Johnson, 1994; MacLeod, 1998; MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003), although there are some exceptions (Geiselman, Bjork, & Fishman, 1983).

Although there has been little direct evidence for repression in this early research, Anderson and Green (2001) designed a paradigm that appears to capture repression in the laboratory. They developed the technique of instructing subjects to not think about recently learned events and then assessing their retention on a later test. In their experiments, subjects first learned a list of unrelated word pairs (e.g., *ordeal–roach*) to the criterion of being able to recall 50% of the items in the list. The first and second words of each pair were referred to as the *hint* and *response* words, respectively. Then, in the second phase of the experiment, subjects were instructed to recall aloud the response word (think or respond trials) or "not think" about the response word (suppress trials) when presented with the associated hint word. Subjects were not given a particular strategy to use when not thinking about the response words. The respond or suppress trials were given 0 (baseline), 1, 8, or 16 times.

After this procedure, Anderson and Green (2001) tested subjects' retention for the response words in two ways. In a same-probe test, they were prompted with the hint word that was previously associated with the response word (e.g., *ordeal–*?). In contrast, the independent-probe test provided subjects with a category name and the first letter of the response word (e.g., *insect–r___*). The order of administration for the same- and independent-probe tests was counterbalanced across subjects. Each test had a specific purpose: The same-probe test measured retention of the associative bond of the original pair, whereas the independent-probe test measured accessibility of the response word (*roach*, in our example). The authors argued that the independent-probe test is the more critical of the two because it provided a better measure of the ac-

cessibility of the response word. For example, when tested with the previously associated hint word (e.g., *ordeal–*?), subjects should draw on the information about the relation or link between the two words. However, testing with the independent probe (e.g., *insect–r___*) should reveal pure suppression or facilitation of the response word because the independent cue was never previously studied or associated with the response word during the experiment.

Anderson and Green (2001) found impaired recall for suppressed items in both the same- and independent-probe tests. For example, in Experiment 1, they found a decrease in same-probe test recall from 81% to 72% from 0 to 16 suppressions. Recall on the independent-probe test decreased from 88% to 81%, so that 16 suppression trials produced a 7% effect (about half a percent per suppression trial). Table 1 displays the results from Anderson and Green's first experiment. From these results, Anderson and Green argued that the think/no-think paradigm could function as a laboratory analog for the concept of repression. The independent-probe results indicated to Anderson and Green that the act of suppression had rendered the target item inaccessible.

Our aim in the present experiments was to replicate Anderson and Green's (2001) results and then to explore manifestations of the suppressed item's inaccessibility by various dependent measures. We had initially planned to use response latencies in recall, recognition memory tests, and implicit memory tests such as word stem or fragment completion to determine whether response words would show less accessibility by these measures. (Anderson and Green's, 2001, interpretation would predict that the response word's representation is simply less available by all measures of memory.) These plans went uncompleted, however, when we were unable to replicate the basic result with the same- and independent-probe cues.

At the suggestion of M. C. Anderson when we began this research (personal communication, October 2001), we changed one element of the procedure used by Anderson and Green (2001). Anderson suggested that we use a different way of precuing which items should be responded to and which should be suppressed; Levy and Anderson (2002) reported that the new technique, described below, produced more robust results and eliminated some problems inherent in the Anderson and Green precuing procedure. In Experiments 1 and 2, we used this Levy and Anderson precuing procedure. Experiment 3 was a direct

Table 1
Anderson and Green (2001, Experiment 1):
Mean Proportion of Words Recalled in the
Same-Probe and Independent-Probe Tests

	Number of Repetitions			
	0	1	8	16
Same Probe				
Respond	.84	.89	.98	.99
Suppress	.81	.81	.75	.72
Independent Probe				
Respond	.84	.89	.92	.91
Suppress	.88	.88	.83	.81

replication of Anderson and Green's (2001) procedure. In all three experiments, we failed to find reliable suppression in either the same- or independent-probe tests.

EXPERIMENT 1

Method

Subjects and Design. The subjects were 32 Washington University undergraduates who were tested individually and received credit toward a course requirement. (Anderson and Green, 2001, also used 32 subjects in all their experiments.) The experimental design was a 4 × 2 × 2 completely within-subjects design. The variables were number of presentations of the probe (0, 1, 8, or 16), type of trial (suppress or respond to the probe), and type of test (same probe or independent probe).

Materials. We used 50 unrelated A–B word pairs (e.g., *ordeal–roach*) taken from Anderson and Green (2001). Of those 50 pairs, 10 were used in practice trials and as fillers. The remaining 40 pairs were divided into eight (0, 1, 8, 16 respond and suppress conditions) five-pair groups and were rotated through conditions over subjects for counterbalancing. The stimuli were presented using E-Prime software (Schneider, Eschman, & Zuccolotto, 2002), and voice responses were recorded using Psychology Software Tools' serial response box and voice key.

Procedure. The experiment consisted of three phases: learning, respond/suppress, and testing, which are described in turn.

During the learning phase, subjects studied each word pair for 5,000 msec. A 400-msec interval separated the pairs. The pairs were presented three times in random order, with the restriction that the same pair could not be presented consecutively. The subjects were then given a paper-and-pencil cued recall test, with the left member of the pair serving as the hint or cue word. The experimenter scored the tests between presentations, and the subject had to obtain a score of 50% correct or higher. If the subject did not reach this criterion, he or she repeated this phase. All subjects completed the study phase in two study/test sessions.

In the respond/suppress phase, subjects were presented with a fixation point for 1,000 msec followed by the hint word for 4,000 msec. If the fixation point was presented in green, they were instructed to recall aloud the response word associated with the hint word as quickly as possible. If the subject failed to respond to the hint word within 4,000 msec, the response word was presented in blue-colored font for 2,000 msec. If the fixation point was presented in red, subjects were told to "not think" about the response word by any means possible while still focusing on the hint word.

As described in further detail later (see Experiment 3), this procedure was different from that in Anderson and Green (2001) in which they interposed another phase that required subjects to learn the stimuli or hint words for responses to be suppressed. This original Anderson–Green procedure therefore introduced a confound of additional study time to familiarize the subjects with the "to-be-suppressed" word pairs, because suppressed cue words were presented more often than respond cue words. As noted in the introduction, we made this change in the precuing procedure at the suggestion of M. C. Anderson (personal communication, October 2001) to eliminate the aforementioned confound, as discussed by Levy and Anderson (2002).

The specific instructions used in Experiment 1 were taken from materials provided by M. C. Anderson (personal communication, October 2001):

> When one of these critical HINT words appears on the screen [after the cue to suppress], we would like you to look at it in the same way you would look at any other Hint word. Keep attending to the Hint word—do not look away from it—for all 4 seconds that it will be on the computer screen. However, it is IMPERATIVE that you not say the associated Response aloud. We want you to do more than to NOT SAY the Response word—we want you to PREVENT the Response word (or even the idea

of it) <u>from coming to mind at all</u>, while looking at the Hint word. To be successful at this task, you should learn to not think about the Response word at all, not even for a second, and not even after the Hint word has gone off the screen, if that is possible. You can accomplish this in whatever way you see fit—in whatever way that works for you. The important thing is to learn to NOT THINK about the Response word associated with the Hint word on the screen. However, please be sure to LOOK at the Hint word for the entire time that it is on the screen.

If a subject responded to a suppress hint word, they received an error message. This occurred in less than 3% of all suppress trials. Because removing these trials did not change the outcome of any of the statistical analyses performed, they are included in our data.

Before the actual trials began, the subjects engaged in a set of 32 practice trials with the 10 filler word pairs to orient them to the task. These consisted of four pairs that were presented two times, four pairs that were presented three times, one pair that was presented four times, and one pair that was presented eight times. They were all respond trials except for the pair that was presented eight times, which subjects were told to suppress. After the practice, there was a short break (<3 min), and the subjects began the 377 actual trials. These consisted of the 250 critical trials, plus 127 trials that were all respond trials, consisting of the 10 filler items used during practice. For the 250 critical trials, the counterbalancing ensured that all stimuli were rotated through all conditions.

In the third and final phase, subjects were tested in two ways: a same-probe test and an independent-probe test. In the same-probe test, they were given the hint word (e.g., *ordeal*–?) for 4,000 msec and were asked to recall aloud the response word in order to activate the voice key. After their response triggered the voice key, subjects typed their response. In the independent-probe test, subjects were given a category cue and the first letter of the response word (e.g., *insect–r*____) for 4,000 msec. The same-probe test provided an intralist cue, whereas the independent-probe test provided an extralist cue. The order of the tests was counterbalanced across subjects. Only the 40 critical pairs were tested (i.e., the pairs that served as practice and filler items were not tested).

Results

The recall results are shown in Table 2. Consider first the same-probe data shown in the top half of the table. The level of recall for nonpracticed items in the same-probe condition was .75, which serves as our baseline. (This data point combines nontested items from both the respond and suppress trials). Recall for the respond items increased to .96 following repeated responding, demonstrating facilitation for the items that were repeatedly retrieved. However, the suppress items did not decrease significantly below baseline after either 8 or 16 repetitions, failing to replicate the inhibitory effect found by Anderson and Green (2001). An ANOVA revealed a main effect of type of response [$F(1,31) = 41.37$, $MS_e = 0.03$,

Table 2
Experiment 1: Mean Proportion of Words Recalled in the Same-Probe and Independent-Probe Tests

	Number of Repetitions			
	0	1	8	16
Same Probe				
Respond	.74	.93	.93	.96
Suppress	.77	.73	.81	.74
Independent Probe				
Respond	.83	.86	.92	.88
Suppress	.81	.80	.87	.84

$p < .001$], with respond items yielding higher levels of recall. There was also a main effect for number of presentations [$F(3,93) = 5.95$, $MS_e = 0.03$, $p < .01$], with more repetitions being associated with higher levels of recall. Finally, there was a significant interaction between type of response and the number of repetitions [$F(3,93) = 8.41$, $MS_e = 0.02$, $p < .001$]. However, this interaction appears to have been driven by the increased recall for respond items rather than any inhibitory effect for suppress items. A t test revealed a significant effect of responding between 0 and 16 repetitions [$t(31) = 6.66$, $p < .001$], whereas a t test comparing 0 and 16 suppressions failed to approach significance [$t(31) = 0.52$].

In the more critical independent-probe condition, subjects were cued with a category name and letter, with the results shown in the bottom part of Table 2. There was a tendency for a main effect of type of response [$F(1,31) = 3.10$, $MS_e = 0.03$, $p = .09$], with respond items showing numerically greater recall than suppress items. In addition, as the number of presentations increased, items that were repeated more frequently were better recalled [$F(3,93) = 3.90$, $MS_e = 0.01$, $p = .06$]. There were no significant interactions ($F < 1$). A t test comparing the baseline (0) condition (.82) with the mean of the 1, 8, and 16 respond conditions (.88) approached significance [$t(31) = 1.76$, $p = .09$]. There was no hint of inhibition; in fact, the mean of the three suppress conditions (.84) was slightly above that of the mean of the baseline conditions (.82).

EXPERIMENT 2

In Experiment 1, we were unable to obtain the decrease in recall in either the same-probe or independent-probe test reported by Anderson and Green (2001). Our goal for Experiment 2 was again to obtain the suppression effect in both the same- and independent-probe tests by more closely following the method used by Anderson and Green. In Experiment 1, the learning phase was somewhat altered from their original methodology. We changed our procedure in Experiment 2 to replicate the original methodology. We also introduced two new questionnaires provided by M. C. Anderson (personal communication, September 2002) to try to identify and eliminate subjects who may not have been following instructions.

Method

Subjects and Design. The subjects were 40 Washington University undergraduates who were tested individually and received credit toward a course requirement. The design called for 32 subjects, but 8 were excluded on the basis of responses on the posttest questionnaire, which left 32 for analyses (with complete counterbalancing achieved).

Materials. The materials were the same as in Experiment 1 except that two questionnaires were included. The first questionnaire was given after a practice block in the respond/suppress phase to help identify any covert rehearsal of to-be-suppressed items and to allow the experimenter to give feedback to the subject to correct this problem. The entire questionnaire can be found in Appendix A. A sample item from the questionnaire is "When you saw the RED hint word, how often did you think of the word that went with it and simply didn't say it out loud?" The second questionnaire (in Appendix B), given after the sub-

ject had completed the entire experiment, was designed to determine whether the subject had adequately followed instructions during the experiment. A sample item was "I read the red HINT word, tried to not think of the associated response, but then <u>after the trial was over</u> I made sure that I still remembered the RESPONSE word." As noted, 8 subjects were eliminated because of positive responses to these types of items.

Procedure. The experiment consisted of three phases, as in Experiment 1, but the learning phase was changed to replicate exactly that of Anderson and Green (2001). In particular, subjects were presented the word pairs for 5,000 msec each, with a 400-msec interval between pairs. After one presentation of the pairs, they began the test–feedback part of the experiment. Subjects were presented with a fixation point for 200 msec. Then the hint word (e.g., *ordeal*) was presented for up to 5,200 msec, and the subject was asked to recall the associated response word (e.g., *roach*). Regardless of whether subjects answered correctly or remained silent, they were shown the response word in blue for 3,000 msec. This test–feedback procedure continued until the subjects were at least 50% accurate on the entire set of word pairs, which was Anderson and Green's criterion.

The respond/suppress phase was like that of Experiment 1 and used the precuing procedure of Levy and Anderson (2002). Briefly, subjects were presented with a fixation point for 1,000 msec followed by the hint word for 4,000 msec. If the fixation point was presented in green, they were instructed to recall the associated response word, whereas if the fixation was presented in red, they were told not to think about the response word. If a subject responded to a suppress cue, they received an error message. As in Experiment 1, this happened in less than 3% of suppress trials, and including or removing these trials had no effect on the results. The subjects were given a diagnostic questionnaire at this point and then were presented another practice block. After the practice, the subjects began the 377 critical trials. The test administration was the same as in Experiment 1.

Results

Table 3 displays recall results for both the same-probe and independent-probe tests. Considering the same-probe results first, recall in the baseline (0) repetition condition was .69 (averaged over the two counterbalanced groups of items). Recall of the respond items increased to .96 over 16 repetitions, demonstrating facilitation and replicating Anderson and Green (2001), as well as our own results from Experiment 1. Recall of the suppress items decreased only slightly, .04 below baseline in the condition with 16 suppression trials. An ANOVA revealed a main effect of type of response [$F(1,31) = 102.35$, $MS_e = 0.02$, $p < .001$], with respond items yielding higher levels of recall. There was also a main effect for number of presentations [$F(3,93) = 6.74$, $MS_e = 0.02$, $p < .001$], with, on average, more repetitions having higher levels of recall. Finally, there was a significant interaction between type

Table 3
Experiment 2: Mean Proportion of Words Recalled in the Same-Probe and Independent-Probe Tests

	Number of Repetitions			
	0	1	8	16
Same Probe				
Respond	.70	.89	.91	.96
Suppress	.68	.68	.66	.64
Independent Probe				
Respond	.78	.87	.84	.83
Suppress	.78	.76	.79	.76

of response and number of repetitions [$F(3,93) = 13.76$, $MS_e = 0.02$, $p < .001$]. A direct test of the 4% difference between the baseline (0) and 16 suppression conditions revealed no significant effect [$t(31) = 1.29$, $SEM = 0.04$].

In the critical independent-probe condition (see the bottom part of Table 3), an ANOVA revealed a main effect for type of response [$F(1,31) = 6.19$, $MS_e = 0.03$, $p < .05$], with respond items producing higher levels of recall than suppress items. In this experiment, there was no main effect for number of repetitions and no significant interactions, both yielding ($F < 1$). There was only a 2% decrease in recall (from .78 to .76) from the baseline (0) condition to the condition with 16 suppressions, which did not approach significance [$t(31) = 0.62$, $SEM = 0.04$]. As in Experiment 1, we found no statistically significant inhibition on either the same-probe or independent-probe tests even after 16 suppression trials.

EXPERIMENT 3

Experiments 1 and 2 deviated in slight ways from Anderson and Green's (2001) procedure, in particular the manner in which subjects were precued during the respond/suppress phase of the experiment. In Experiment 3, we followed Anderson and Green's procedure as closely as possible, using their "hint-training" method for cuing subjects for suppress trials.

Method

Subjects and Design. Altogether 39 Washington University undergraduates participated. Seven subjects were excluded on the basis of their responses on a posttest questionnaire; so 32 subjects' data (with complete counterbalancing) were included in the analyses. Anderson and Green (2001) did not report using such questionnaires, and this was intended to be a direct replication, but we decided to err on the side of caution. Subjects were tested individually and received credit toward an introductory psychology course.

Materials. The materials were the same as in Experiment 2.

Procedure. The training phase was identical to that of Experiment 2 and Anderson and Green (2001). After the training phase, subjects received the extra hint-training phase used by Anderson and Green. During this phase, subjects learned to recognize which 15 hint words would require suppression during the respond/suppress phase. The full list of all 15 suppress hint words was displayed in a column in the center of the screen for 1 min. After the minute had elapsed, subjects were given a recognition test to determine whether they could identify all of the critical hint words that they had just studied. They received a sheet of paper with 29 words from the experiment, of which 15 were the critical hint words. The other 14 items served as distractors and consisted of both hint and response words taken from the filler pairs. The subjects were told to circle the critical hint words that they had just studied. They were required to achieve perfect performance on this test before moving to the next phase. If subjects failed to achieve perfect performance, they repeated this phase up to two additional times. If they had still failed to reach perfect performance on the recognition test, they would have been excluded from the experiment, but none did.

The respond/suppress phase was the same as in Experiment 2 except that there were no explicit cues (the color of the fixation point in Experiments 1 and 2) to indicate in what type of trial the subject was about to engage. The test administration was the same as in Experiment 2.

Table 4
Experiment 3: Mean Proportion of Words Recalled in the Same-Probe and Independent-Probe Tests

	Number of Repetitions			
	0	1	8	16
Same Probe				
Respond	.70	.86	.98	.99
Suppress	.70	.70	.71	.69
Independent Probe				
Respond	.79	.85	.89	.86
Suppress	.80	.77	.77	.79

Results

Again, subjects were tested in two ways to assess their recall. These results are presented in Table 4; we describe the same-probe results first. Recall for the respond items increased from .70 in the baseline condition to .99 after 16 repetitions. Again, there was no suppression for items following the no-think instructions. An ANOVA revealed a main effect of type of response [$F(1,31) = 61.09$, $MS_e = 0.03$, $p < .001$], with respond items yielding higher levels of recall. There was also a main effect for number of presentations [$F(3,93) = 12.84$, $MS_e = 0.02$, $p < .001$], with, on average, more repetitions having higher rates of recall. Finally, there was a significant interaction between type of response and number of repetitions [$F(3,93) = 13.76$, $MS_e = 0.02$, $p < .001$]. When directly comparing the 1% difference between 0 and 16 repetitions in the suppress condition, however, a t test failed to reach significance [$t(31) = 0.11$, $SEM = 0.06$].

In the more important independent-probe data shown at the bottom of Table 4, an ANOVA revealed a main effect for type of response [$F(1,31) = 6.26$, $MS_e = 0.05$, $p < .05$], with respond items producing higher levels of recall than suppress items. In this experiment, there was no main effect for number of repetitions and no significant interaction. As expected, the 1% difference between 0 and 16 repetitions in the suppression condition did not approach significance [$t(31) < 0.10$]. So, as in our previous experiments, we found no support for suppression in the crucial independent probe recall results. In sum, we again failed to obtain evidence of inhibition in either the same-probe or independent-probe tests, despite this time using the exact Anderson and Green (2001) procedure.

DISCUSSION

In three experiments, we used the conditions, the materials, and, for the most part, the procedures described by Anderson and Green (2001), and we were unable to replicate the result of suppression in cued recall following no-think instructions. Experiment 3 constituted a direct replication. We did occasionally find small numerical differences between baseline and 16 suppressions in the same-probe condition, but none approached significance. More importantly, we found no evidence for suppression in recall in any experiment with the crucial independent-probe test.

To examine the effect across all experiments, we combined the suppression data from all three experiments. There were minor differences in the methodologies of our experiments, but the procedural differences did not much affect our results. For the same-probe conditions, the mean recalls for the 0, 1, 8, and 16 suppression conditions were .71, .70, .73, and .69. Even the 2% decline after 16 suppressions did not differ from that of the baseline condition [$t(95) = 0.79$, $SEM = 0.03$]. The comparable data for the more important independent probe were .80, .78, .81, and .80. Obviously, the proportions of cued recall in the baseline and 16 suppression conditions were identical. Because there was no evidence of suppression in the independent-probe test, there is no evidence that the availability or accessibility of items was decreased following 16 no-think trials.

In addition to combining the data from our experiments, we also examined the data from the first test that the subjects took. Recall that, in all three experiments, the subjects took both the same- and independent-probe tests and the order was counterbalanced. The data presented above, as in Anderson and Green's (2001) experiments, combined data from the first and second tests. This leaves open the possibility that the act of taking the first test may alter performance on the second test. Therefore, we examined the data from the first test of each experiment separately, and we also analyzed the data from the first test combined across the three experiments. There were no significant effects of suppression in any of these additional analyses for either the same- or independent-probe tests. We also collected latency data in our experiments, which are not reported. These data also did not provide evidence of inhibition following no-think trials.

One might argue that there were individual differences between the samples of subjects at the University of Oregon and Washington University and that this fact could possibly explain the differences in results. Anderson and Green (2001) noted that the suppression task reflects an executive process and so it should be facilitated in subjects higher in executive control. However, we think it is unlikely that our sample contained a disproportionate number of low executive control subjects relative to the University of Oregon sample. We did not include any measure of executive control to screen our subjects, nor did Anderson and Green. However, there is no a priori reason to believe this would be true. If Verbal + Quantitative SAT scores are used as a proxy for intelligence, data from these universities for entering classes in 2003 showed that Washington University undergraduates scored 1390 (Washington University, 2003) and University of Oregon students scored 1113 (University of Oregon, 2003), a difference of nearly 280 points favoring our subjects.

Anderson and Kuhl (2004) reported evidence that another possible individual difference variable plays a role in the think/no-think paradigm. Specifically, they compared individuals who reported that they experienced a relatively high number of traumatic events in the past with individuals who reported experiencing relatively fewer traumatic events in the think/no-think paradigm. Their results indicated that only those individuals who reported a relatively high level of past traumatic events showed suppression in the think/no-think paradigm. Interestingly, and consistent with the present results, there was relatively little evidence of suppression for the low-trauma subjects. According to Anderson and Kuhl, students who had suffered trauma in the past are better at suppressing in the think/no-think paradigm, due to practice suppressing traumatic events outside the laboratory. Obviously, this hypothesis is an intriguing target for future research, but the result also indicates that many (and perhaps most) subjects will not show the effect. As applied to our present data, one would need to argue that our Washington University students suffered fewer traumatic events than did typical University of Oregon students (the original subjects in Anderson and Green's, 2001, research), which seems unlikely.

Since we initiated this project, there have been other relevant results that have appeared in the literature. Hertel and Gerstle (2003) also reported an experiment using the think/no-think methodology comparing dysphoric and nondysphoric subjects. They did obtain differences between the two groups of subjects for the suppress trials (with dysphoric subjects recalling more targets than nondysphoric subjects). However, they did not obtain a significant decrease below baseline for the suppression trials for either group; nondysphoric subjects did show a marginally significant trend for positive valence cues to produce below-baseline suppression (but negative valence cues produced a difference in the opposite direction, albeit not significantly so). Also, Hertel and Gerstle used only the same-probe test (not the critical independent-probe test), and so any suppression could have been due to the weakening of associative bonds rather than a loss of availability of the target word (e.g., *roach*). However, to reiterate, there was no significant suppression even in the same-probe test.

Anderson et al. (2004) conducted a neuroimaging study in which they used a variation of the Anderson and Green (2001) procedure suitable for imaging. Considering only the behavioral results, the authors did obtain a small but significant suppression effect—roughly a 7% difference between the baseline and 16 suppression items in the independent-probe test. We used this new procedure in a purely behavioral experiment, not reported here, but again failed to find any suppression in either a same-probe or independent-probe test. However, because the Anderson et al. (2004) procedure did not require subjects to provide responses during either the respond or the suppress trials (to prevent subjects from speaking while in the magnet in their experiment), we worried that our subjects might simply have not been following directions. Even in our think conditions, later cued recall did not behave in the typical fashion, so the failure to find inhibition in the no-think condition may not be diagnostic.

While working on this project, we have become aware of other groups of researchers who have failed to replicate the original Anderson and Green (2001) results, although most have given up and not attempted to publish their results. Still, to reiterate, we are not claiming in this article

that the results cannot be replicated; obviously, Anderson and Green obtained them in several experiments, as did Anderson et al. (2004). (Of course, Hertel & Gerstle, 2003, did not, but they used more strongly related pairs and attributed their failure to this change.) Rather, our claim is that our results show, at the least, that suppression results are not particularly robust and are not easily obtained even by researchers following the original procedures as closely as possible. We had originally hoped to extend the results to other materials, to other measures besides cued recall, and so on; however, of course, we abandoned these plans when we could not replicate the basic results.

Even in the original Anderson and Green (2001) research, the suppression effects were small and somewhat erratic. With use of 0, 1, 8, and 16 suppression attempts, recall in either the same- or independent-probe conditions was never below baseline in the 1 suppression attempt conditions (perhaps not surprisingly), and performance following 8 no-think trials was only reliably below baseline in some experiments. Only the 16 suppression condition reliably showed inhibition below baseline in the Anderson and Green study in both the same- and independent-probe conditions, and then the effect was about 8%, or a half percent loss in cued recall per suppression trial attempt.

Hertel and Calcaterra (2005) reported results that may help to resolve the mystery of when and why inhibition is sometimes obtained in the think/no-think paradigm. In their experiment, one group of subjects was tested under think/no-think conditions as in our experiments and those of Anderson and Green (2001), a condition they called *unaided suppression*, for reasons that will become clear. In another condition, subjects were given an aid to help in suppressing associates by being provided, before the suppression trials, another word to think of in response to the cue word. So, for example, if subjects had learned the pair *porcelain–doll* in the original acquisition phase, in the aided suppression condition they were taught the pair *porcelain–goblet* before the suppression phase so that when they saw the cue *porcelain* with instructions to suppress, they could substitute practice on *goblet* as an aid to suppressing *doll*. The aided suppression condition therefore amounts to an A–B, A–D interference paradigm (Barnes & Underwood, 1959). During the final test, subjects were given the cue words (e.g., *porcelain*) and instructed to recall the word associated with it in the original list (*doll*) but were also told that if two words came to mind, they should respond with both words without worrying about which was correct. (Therefore, the instructions constituted what Barnes and Underwood called *modified–modified free recall*.)

In the unaided suppression condition that was comparable to the standard think/no-think condition of Anderson and Green (2001), Hertel and Calcaterra (2005) obtained no suppression effects on the cued recall test even in the direct-probe condition, thereby replicating the present pattern. (They did not include the critical independent-probe condition in their study.) However, in the aided suppression (or A–B, A–D interference) condition, they did find inhibition on the same-probe test. The conclusion

that might be drawn is that suppression instructions to not think of an unwanted response may succeed if subjects are given the strategy (or themselves hit upon the strategy) of always thinking of some other item when they are trying to suppress an unwanted response.

Hertel and Calcaterra (2005) gave questionnaires after their experiment to query subjects about their strategies. They showed that students in the unaided condition who hit upon the strategy of thinking about something irrelevant while trying to suppress a response did indeed show suppression in cued recall. On the other hand, subjects who did not hit upon such a strategy actually showed facilitation in the suppression condition. This situation led to the null result overall in the suppression condition, due to averaging of the suppression in subjects using an interference/substitution of responses strategy and the facilitation in the data from people who did not use this substitution strategy. The upshot of the Hertel and Calcaterra results is that suppression may only occur when subjects suppress responses by using the strategy of substituting a new response for the unwanted one, thereby creating retroactive interference.

Of course, even assuming that Hertel and Calcaterra's (2005) conclusion is correct, the mystery of why we (and also Hertel and Calcaterra in the unaided direct-probe condition) failed to replicate the results of Anderson and Green (2001) is not solved. Anderson and Green did not instruct subjects to use a substitution strategy to induce retroactive interference, and neither did we. If the substitution strategy is the crucial factor, it is unclear why many of Anderson and Green's subjects hit upon the strategy spontaneously but ours did not, given that the same instructions were used in both sets of experiments. Still, the Hertel and Calcaterra work does point to a promising lead in explaining why suppression occurs when it does occur—namely, due to retroactive interference via a substitution of one response for another. Of course, it would be a welcome development to determine whether their conclusion will also hold when an independent-probe (indirect-probe) test is used.

In some ways, it is not terribly surprising that we failed to replicate Anderson and Green's (2001) basic phenomenon. The effect was small in the original study, and some inhibitory phenomena are notoriously difficult to pin down. It is also worth noting that some manipulations that bear a surface similarity to those of Anderson and Green yield very different results. Consider work done by Wegner and colleagues (Wegner, Schneider, Carter, & White, 1987; see Wegner, 1989, 1994, for reviews). They asked subjects not to think about a particular item (e.g., white bears) and then measured the tendency to think about the forbidden object (by pressing a key whenever the thought came to mind). Over many experiments, Wegner and his colleagues have reported a rebound phenomenon wherein subjects were more likely to think about white bears when trying to suppress them than under other control conditions. Although these paradigms certainly are different, both involve the attempt to suppress thoughts and yet lead to opposite outcomes.

As one reviewer of the manuscript pointed out to us, if one is interested in generalizing results to phenomena outside the laboratory, the Wegner (1994) thought suppression paradigm seems more apt than the Anderson and Green (2001) paradigm. Specifically, when attempting to stop thinking about some unpleasant experience (a rejected paper, some failure in sports, an argument) or some other, more dramatic event (a death of a loved one, a sexual trauma, an accident or injury), the thought one is trying to suppress is a single thought in central, focal attention. Wegner's paradigm has subjects focus on one thought and try to suppress it; under this condition, he finds the result that many of us know through experience—the mind becomes locked in a positive feedback loop such that attempts to suppress the thought have the ironic effect of making the thought even more accessible to consciousness (Wegner, 1994).

Attempts to find experimental evidence for Freud's (1895/1957) concept of repression (in the sense of making memories less accessible to consciousness) have suffered false starts over the past 70 years, ever since Rosenzweig and Mason's (1934) first attempt. We conclude that the journey is not over yet, unless one considers suppression via retroactive interference as in Hertel and Calcaterra's (2005) research to have cut through this Gordian knot. If this concept of repression is held to be correct, then training people in A–B, A–D interference techniques (or counterconditioning techniques) is teaching them to repress thoughts—to substitute a pleasant (wanted) thought for an unpleasant one, for example. However, this definition of repression is not likely to be accepted by most scholars. Rather, researchers will continue to seek evidence for a decrease in availability (or at least accessibility) of memories through thought suppression as provided in Anderson and Green's (2001) original interpretation of their findings.

REFERENCES

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, **410**, 366-369.

Anderson, M. C., & Kuhl, B. A. (2004, November). *Inhibitory control and the suppression of unpleasant events*. Paper presented at the Annual Meeting of the Psychonomic Society, Minneapolis, MN.

Anderson, M. C., Ochsner, K. N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S. W., et al. (2004). Neural systems underlying the suppression of unwanted memories. *Science*, **303**, 232-235.

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology*, **58**, 97-105.

Basden, B. H., & Basden, D. R. (1998). Directed forgetting: A contrast of methods and interpretations. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 139-172). Mahwah, NJ: Erlbaum.

Bjork, R. A., LaBerge, D., & Legrand, R. (1968). The modification of short-term memory through instructions to forget. *Psychonomic Science*, **10**, 55-56.

Erdelyi, M. H. (1990). Repression, reconstruction, and defense: History and integration of the psychoanalytic and experimental frameworks. In J. L. Singer (Ed.), *Repression and dissociation: Implications for personality theory, psychopathology, and health* (pp. 1-31). Chicago: University of Chicago Press.

Freud, S. (1957). *The standard edition of the complete psychological works of Sigmund Freud*. London: Hogarth. (Original work published 1895)

Geiselman, R. E., Bjork, R. A., & Fishman, D. L. (1983). Disrupted retrieval in directed forgetting: A link with posthypnotic amnesia. *Journal of Experimental Psychology: General*, **112**, 58-72.

Glucksberg, S., & King, L. J. (1967). Motivated forgetting mediated by implicit verbal chaining: A laboratory analog of repression. *Science*, **158**, 517-519.

Hertel, P. T., & Calcaterra, G. (2005). Intentional forgetting benefits from thought substitution. *Psychonomic Bulletin & Review*, **12**, 484-489.

Hertel, P. T., & Gerstle, M. (2003). Depressive deficits in forgetting. *Psychological Science*, **14**, 573-578.

Holmes, D. S. (1990). The evidence for repression: An examination of sixty years of research. In J. L. Singer (Ed.), *Repression and dissociation: Implications for personality theory, pathology, and health* (pp. 85-102). Chicago: University of Chicago Press.

Johnson, H. M. (1994). Processes of successful intentional forgetting. *Psychological Bulletin*, **116**, 274-292.

Levy, B. J., & Anderson, M. C. (2002, November). *Executive control and the mechanisms of encoding*. Paper presented at the Annual Meeting of the Psychonomic Society, Kansas City, MO.

MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1-58). Mahwah, NJ: Erlbaum.

MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 43, pp. 163-214). San Diego: Elsevier, Academic Press.

Patalano, A. L., & Seifert, C. M. (1994). Memory for impasses during problem solving. *Memory & Cognition*, **22**, 234-242.

Roediger, H. L., III, & Crowder, R. G. (1972). Instructed forgetting: Rehearsal control or retrieval inhibition (repression)? *Cognitive Psychology*, **3**, 244-254.

Rosenzweig, S., & Mason, G. (1934). An experimental study of memory in relation to the theory of repression. *British Journal of Psychology*, **24**, 247-265.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburgh: Psychology Software Tools.

University of Oregon (2003). *UO Homepage News Archive*. Retrieved October 21, 2003 from duckhenge.uoregon.edu/hparchive/display .php?q=enroll.html.

Washington University (2003). *Academic life—Parents handbook*. Retrieved October 21, 2003 from parents.wustl.edu/handbook/stud .html.

Wegner, D. M. (1989). *White bears and other unwanted thoughts: Suppression, obsession, and the psychology of mental control*. New York: Penguin.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, **101**, 34-52.

Wegner, D. M., Schneider, D. J., Carter, S. R., III, & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality & Social Psychology*, **53**, 5-13.

Weiner, B. (1968). Motivated forgetting and the study of repression. *Journal of Personality*, **36**, 213-234.

Weiner, B., & Reed, H. (1969). Effects of the instructional sets to remember and to forget on short-term retention: Studies of rehearsal control and retrieval inhibition (repression). *Journal of Experimental Psychology*, **79**, 226-232.

Zeigarnik, B. V. (1927). Das Behalten erledigter und unerledigter Handlungen. *Psychologische Forschung*, **9**, 1-85.

APPENDIX A
Preprocedure Questionnaire

<u>**GREEN HINT WORDS**</u>

1. For the green hint words, how often did you try to come up with the associated response as FAST as possible?

Never		Half of the time		Always
0	1	2	3	4

<u>**RED HINT WORDS**</u>

1. Typically, for how many seconds did you look at a RED hint word when it was presented on the screen?

0	1	2	3	4

2. When you looked at the RED hint word, how often did you read and understand it?

Never		Half of the time		Always
0	1	2	3	4

3. When you saw the RED hint word, how often did you think of the word that went with it and simply didn't say it out loud?

Never		Half of the time		Always
0	1	2	3	4

4. When you saw the RED hint word, how often were you able to avoid thinking about the word that went with it?

Never		Half of the time		Always
0	1	2	3	4

5. When the RED hint word went off the screen, how often, did you then think about the word that went with it?

Never		Half of the time		Always
0	1	2	3	4

6. When you saw the RED word, did you ever think about the associated word "just for a second" to see if you still knew it?

Never		Half of the time		Always
0	1	2	3	4

APPENDIX B
Postprocedure Questionnaire

Sometimes people suspect that their memory will be tested on response words for red HINT words later on, even though they have been asked to not think about the RESPONSE words. Below you find examples of strategies that people have used in the past to make sure that they still knew the RESPONSE for the red HINT words. Please read all the statements first and then indicate on the scale below how often—if ever—you used any of these strategies (Note: it is extremely important for the experiment that you be as honest and accurate as possible with your answers—please, just describe your experience during that phase).

	Never	Rarely	Sometimes	Frequently	Very Frequently
I read the red HINT word, made sure I still knew the associated word first and <u>then</u> tried to <u>not think</u> of this associated word.	0	1	2	3	4
I read the red HINT word, tried to not think of the associated response, but then <u>after the trial was over</u> I made sure that I still remembered the RESPONSE word.	0	1	2	3	4
I read the red HINT word and kept myself from <u>saying</u> the RESPONSE word aloud, but kept repeating the response word to myself to improve my memory for it.	0	1	2	3	4