

Performance, cardiovascular, and health behavior effects of an inhibitory strength training intervention

Siu-kuen Azor Hui · Rex A. Wright ·
Christopher C. Stewart · Angel Simmons ·
Bradley Eaton · R. Nicholas Nolte

Published online: 30 September 2009
© Springer Science+Business Media, LLC 2009

Abstract Female undergraduates were assigned to one of three groups, two involving regulatory training and one not. Training participants performed for 2 weeks tasks that required strong behavioral restraint (Strong Training) or weak behavioral restraint (Weak Training). Later, they took part in (1) a laboratory session in which they performed tasks with inhibitory components, and (2) a follow-up week in which they provided health behavior reports and used designated dental supplies. No Training participants took part only in the session and follow-up week. As expected, laboratory performance was improved for Strong- relative to No Training participants, with performance for Weak Training participants falling in between. Also as expected, Strong Training participants used more floss in the follow-up week than did the No Training participants, with floss for Weak Training participants falling between. Contrary to expectation, Strong Training participants used less toothpaste and reported having brushed less than the No Training participants. In addition, Strong Training participants evinced exaggerated—rather than diminished—cardiovascular responses during the laboratory tasks. The performance and floss use data support the suggestion that inhibitory system strength can be increased through use. The brushing and cardiovascular findings may be interpretable in inhibitory strength terms.

Keywords Regulatory control · Inhibition · Inhibitory training · Resource depletion · Ego depletion · Cardiovascular response

Introduction

Baumeister, Muraven and various colleagues have argued that self-regulation (behavioral restraint or inhibition) may involve a special performance system that functions like a muscle (Baumeister et al. 1998; Muraven and Baumeister 2000; Muraven et al. 1998, 2006). They have contended that, like a muscle, this system may draw on a performance (energy) resource that can be temporarily depleted through short-term use. Also like a muscle, the system may be strengthened through extended use.

Studies designed to evaluate the muscle analysis have produced abundant evidence favorable to the first suggestion above (Gailliot and Baumeister 2007; Schmeichel 2007; Schmeichel et al. 2003; Vohs and Schmeichel 2003; for a review, see Baumeister et al. 2006). More specifically, they have shown repeatedly that short term regulatory action tends to impair later regulatory task performance unless available performance incentives justify the extra effort required for success (Muraven and Slessareva 2003) or an energy source is introduced to restore regulatory capacity (Gailliot et al. 2007). By contrast, the studies have produced relatively limited evidence favoring the second suggestion.

To date, the strongest evidence for the second suggestion has come from an early experiment by Muraven et al. (1999) and a series of more recent experiments by Oaten and Cheng (2006a, b, 2007). Muraven et al. assigned participants to conditions in which they were directed to engage consistently in one of four regulatory behaviors (e.g., improving

S. A. Hui
University of Rochester Medical Center, Rochester, NY, USA

R. A. Wright (✉) · C. C. Stewart · A. Simmons ·
B. Eaton · R. N. Nolte
Department of Psychology, University of Alabama
at Birmingham, 415 Campbell Hall, 1300 University
Boulevard, Birmingham, AL, USA
e-mail: rwright@uab.edu

posture) over the course of 2 weeks. At the end of the training period, they assessed regulatory capacity operationalized as persistence in meeting a hand-grip challenge. Results indicated that regulatory treatment participants persisted longer following a depleting thought-suppression exercise than did no-treatment control participants.

Oaten and Cheng evaluated the influence of physical exercise, academic study, and financial monitoring programs on an array of outcomes involving inhibitory control, focusing on visual tracking following a thought-suppression exercise. Typical is their experiment involving financial monitoring (Oaten and Cheng 2007). Participants volunteered for a 4-week program that required them to meet with the experimenter, work out a management plan, and maintain spending records. Investigators measured tracking performance at the beginning of the study and in four subsequent laboratory sessions. They found a decline in errors across the training period within the experimental cohort, but not within a cohort of no-training controls.

Findings from the regulatory training studies above are encouraging with respect to the suggestion that inhibitory system strength can be improved through use. However, they call for replication, particularly in the context of protocols that involve markedly different training procedures and conceptually related, but operationally distinct, inhibitory strength outcomes. They also would benefit from extension via experiments that move beyond simple training/no-training comparisons.

One purpose of the present research was to address the preceding need for replication. A second purpose was to extend previous findings by including a condition that involved training comparable to, but less demanding than, that in the main training condition. Based on the Muraven and Baumeister reasoning, the central expectation was that resulting inhibitory strength in this “weak” training condition would be somewhere between that in the main (“strong”) training condition and that in the no training control condition.

Overview

Participants were assigned randomly to one of three groups, two involving training intervention and one not. Training participants performed for 2 weeks tasks that required strong behavioral restraint (Strong Training) or weak behavioral restraint (Weak Training). Later, they took part in a laboratory session and a follow-up “report” week. No Training participants took part only in the session and follow-up week. The full protocol time line for each group can be seen in Fig. 1.

The laboratory session began with a questionnaire period and a baseline rest period. Following the baseline, participants performed a moderately difficult mental

Time Line for Different Experimental Groups

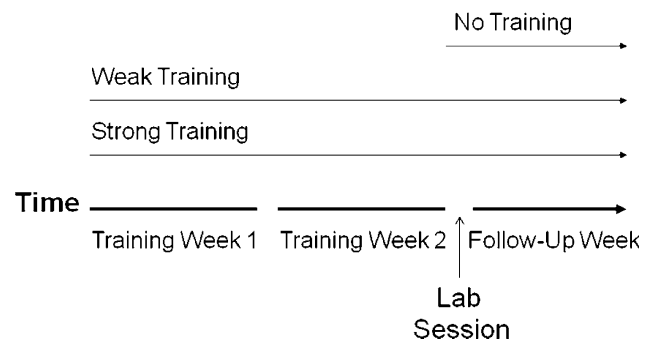


Fig. 1 Time line for the different experimental groups

concentration task (the d2, Brickenkamp 1981), rested, and then put their hand in near-freezing water with instructions to hold it there as long as possible, aiming to hold it for at least 60 s. For reasons described below, measures of systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial blood pressure (MAP), and heart rate (HR) were taken during the initial baseline and during the concentration and cold tolerance work periods. The time line for the laboratory session is in Fig. 2. During the follow-up week, participants completed daily inventories assessing health behaviors that involve self control (e.g., flossing) and used in their normal dental care supplies provided by the experimenter.

Central measures and predictions

Laboratory

Primary laboratory measures of inhibitory strength were concentration task scores and cold tolerance times. Concentration scores were linked to inhibitory strength because good performance on the task requires performers to resist tempting, but incorrect, response options. Tolerance times were linked on grounds that continued immersion requires resistance against a rising impulse to withdraw.

Additional laboratory measures were cardiovascular (CV) responses assessed during the concentration and cold

Time Line Within the Laboratory Session (All Conditions)

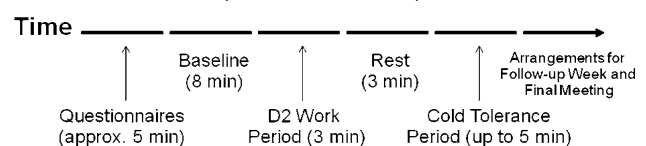


Fig. 2 Time line within the laboratory session

tolerance periods. We examined CV responses with two thoughts in mind. First, there is growing evidence that sympathetically mediated CV adjustment varies with effort (Brinkmann and Gendolla 2007, 2008; Gendolla and Krüsken 2002; Light and Obrist 1980; Obrist 1981; Richter et al. 2008; Smith et al. 1990). Second, there is reason to believe that people with higher ability sometimes expend different degrees of effort when confronted with a performance challenge than people with lower ability (Ford and Brehm 1987; Wright 1996; Wright and Kirby 2001).

Regarding the latter, so long as low- and high ability groups view success as possible and worthwhile, members of the low ability group should exert more effort to make up for their lack of performance capacity (Marcora et al. 2008). On the other hand, where a high ability group perceives success as possible and worthwhile, but a low ability group does not, members of the low ability group should exert less effort (Wright et al. 2007). The reason is because high ability group members should strive in proportion to challenge difficulty, whereas low ability group members should withhold effort to avoid expending energy resources futilely or inefficiently (Wright 2008).

An assumption was that participants with stronger inhibitory systems should be more likely to meet the concentration and cold tolerance challenges than participants with weaker inhibitory systems. This led us to expect better concentration and cold tolerance performances among Strong Training- than No Training participants, with performances for Weak Training participants falling in between. A further assumption was that all participants would view (1) success on the concentration task, and (2) *early* success on the cold tolerance task (i.e., tolerance for at least 60 s) as possible and worthwhile. This led us to expect *weaker* effort-related CV responses during the concentration period and early tolerance period among Strong Training (higher ability) participants than among No Training (lower ability) participants, with responses for Weak Training participants falling in between. Measures of SBP and HR tend to be more sensitive to sympathetic nervous system influence than DBP and MAP (Berntson et al. 1993). Consequently, we expected SBP and HR responses to be especially likely to reflect the expected (inverse) linear response pattern.

Follow-up week

Follow-up report week measures of inhibitory strength included reports of health-related behavior and two behavioral measures of dental care: the amount of (1) dental floss, and (2) toothpaste, remaining at the end of the week. We reasoned that participants with stronger inhibitory systems should be more disciplined in their health habits than participants with weaker inhibitory systems.

Thus, we expected more favorable health behavior reports and less floss and paste among Strong- than No Training participants, with values for Weak Training participants falling between.

Method

Participants

Participants were 75 female undergraduates whom experimenters did not know personally. They signed up for participation on sheets that recruited women who (1) were right handed, (2) were free of circulatory problems, high blood pressure, diabetes, and epilepsy, and (3) had an active e-mail address and home access to a high speed internet connection. Recruitment sheets noted that participants would have the chance to earn four Psychology 101 research credits plus 10 USD, but did not provide study details. We recruited women because they were more available than men and we knew of no reason to believe they should respond differently to training than men should.

Not surprisingly, a number of intervention group participants (Weak Training $n = 7$; Strong Training $n = 13$) initially agreed to participate, but were unable or unwilling to complete their training. These were replaced to the degree that practical constraints allowed them to be. Because self-selection could be a concern, we determined the number who terminated for reasons within their control (e.g., missed work sessions) and the number who terminated for reasons outside of their control (e.g., software incompatibility). Results indicated that most terminations were due to factors outside the participants' control (13/20 = 65%). Three Weak Training participants terminated for controllable reasons and 4 Strong Training participants did so. The balance of controllable terminations suggests that the final training groups did not have pre-existing differences on relevant trait dimensions such as motivation or inhibitory ability.

The final sample consisted of 55 participants—24 in the no training condition, 13 in the weak training condition, and 18 in the strong training condition. Most were of European (56%) or African (38%) heritage. Age was not recorded. However, given the PY101 pool from which the participants were drawn, it is safe to assume age was typical for first and second year college students.

Cardiovascular measurement

CV measures were obtained with a Medwave Fusion monitor, which utilizes a wrist module with an embedded sensor. The wrist module was placed on the wrist of

participants' left arm where the radial artery passed over the flat portion of the radius bone. This allowed the sensor to measure the amplitude of the radial pulse and make SBP, DBP, and MAP estimates based on an analysis of pulse wave-form characteristics. HR was estimated based on a count of radial pulses. The Fusion can provide CV samples as frequently as every 15 s if left in its "continuous" sampling mode of operation. In this study, we sampled every 30 s in some periods and every 20 s in others.

Training tasks

Intervention participants performed two training tasks, (1) a computer task, and (2) an oral rinse task. For Strong Training participants, the computer task was a classic version of the Stroop color-word conflict task that lasted approximately 5 min.¹ On each of 170 trials, the program presented a color word (example: RED) in a conflicting color of print (example: the color yellow). Color words that could be presented were RED, GREEN, PINK, ORANGE, BLUE, and YELLOW. Print colors in which the words could be displayed were the same. Words were presented for .20 s and followed by a response period lasting 5 s. If participants failed to respond within the time allotted, the program moved to the next trial. Participants' goal on all trials was to identify the color of the print by clicking on the appropriate color word displayed at the bottom of the computer screen. To prevent participants from "cheating" by covering a portion of the color word displayed on each trial, the program presented new color words in random locations in the mid- to upper portion of the screen. The oral rinse task for Strong Training participants was to swish in their mouth for a full 30 s half an ounce of Listerine Antiseptic Mouthwash (original formula), a product that has a high alcohol content and produces a powerful burning sensation.

For Weak Training participants, the computer task was a no conflict version of the computer task described above. On each trial, the program presented a non-color word (example: HOUSE) in a particular color of print. Words that could be presented were TABLE, HOUSE, DOOR, CAR, CAT, and DOG. Print colors were the same as those in the conflict version. As was true for the conflict version, the goal was to identify the print color by clicking on the appropriate color word at the bottom of the screen. The program presented non-color words in random locations in the mid—to upper portion of the screen and had the same display and response times as the conflict version. The oral rinse task for Weak Training

participants was to swish for 30 s half an ounce of a diluted form of the Listerine product (two parts water, one part Listerine).

The strong training protocol described above was assumed to involve significant regulatory control in at least three respects. First, it required participants to maintain a performance regimen over the course of 2 weeks. Second, during its Stroop periods, it required participants to resist tempting, but incorrect, response options. Third, during its rinse periods, it required participants to resist a powerful urge to expel the Listerine. The weak training protocol required participants to maintain a performance regimen as well. However, its Stroop task required minimal resistance and its rinse task required resistance against a less powerful expulsion urge.

Prior to each computer session, participants downloaded their program from a server. They did so understanding that the server would record the time at which they performed and their performance score. Participants performed their oral rinse task understanding that the experimenter would see what remained at the end of the training period. This understanding did not ensure rinse compliance, but seems likely to have facilitated it. Instructions were to perform each training task twice a day for 2 weeks.

Laboratory tasks

Two tasks were administered, a version of the d2 mental concentration task (Brickenkamp 1981) and a cold tolerance task. The d2 involved arrays of "d" and "p" letters presented on four pages. Each "d" and "p" letter was paired with one, two, three, four, or no apostrophes. Instructions directed participants to scan the letter arrays and circle special "d" letters that they encountered, starting with the top row of letters and moving from left to right. The special d letters were ones linked to two, and only two, apostrophes. Instructions also indicated that, for research purposes, it was important for everyone to circle at the same pace. Therefore, the experimenter would set the pace by playing an audiotape that would call out the word "count" every 3 s. Each time participants heard "count", they were to circle a new "d". All participants were provided the same four pages of letters, each containing one 12 × 26 letter array; they were admonished to circle on all counts and only on counts. The d2 involves inhibition because it requires respondents to resist the impulse to circle tempting, but incorrect, letters (p letters and d letters with the wrong number of apostrophes). The cold tolerance task required the participants to immerse their hand in a circulating bath of 5°C water and hold it there as long as possible, aiming to hold it at least 60 s.

¹ The computer training programs were written by Aeron Gault, a computer systems analyst in the UAB Psychology department.

Questionnaires

The study included multiple questionnaires. Several were administered at the beginning of the laboratory session. One was completed at the end of each day during the follow-up report week and one was completed in a final interview session.

Laboratory session

We administered several questionnaires shortly after participants arrived (Fig. 2). One was a *Health Belief Inventory*. It was administered to gain a sense of participants' values with respect to physical exercise, consuming alcohol, flossing, and brushing. The questionnaire asked participants to indicate (1) how many hours a week they would exercise physically if they could exercise as much as they liked, (2) how many alcoholic drinks they thought they should have each week, (3) how many times they thought they should floss each week, and (4) how many times they thought they should brush each week. Response options were 0–2 (coded 1), 3–5 (coded 2), 6–8 (coded 3), 9–11 (coded 4), and 12 or more (coded 5).

A second questionnaire was an *Affect Checklist* administered to assess participants' feelings at the time they arrived. This asked participants to rate the extent to which they felt jittery, happy, nervous, fearful, angry, challenged, wide-awake, sad, threatened, weary, confused, and tired. Responses were made on 11-point scales with endpoints of 0 (not at all) and 10 (extremely).

A third questionnaire was a modified version of the *Fatigue Severity Scale* (Krupp et al. 1989) included to evaluate feelings of fatigue in the preceding 2 weeks. Fatigue in this period was of interest primarily because it might be expected to vary across conditions, being greatest for Strong Training participants. Participants were asked to consider the 2 weeks period and rate the extent to which (1) they had been easily fatigued, (2) fatigue had interfered with their functioning, (3) fatigue had caused problems for them, (4) fatigue had prevented sustained functioning, (5) fatigue had made it hard for them to carry out duties, (6) fatigue had been disabling, and (7) fatigue had interfered with their work, family and/or social life. They responded on 7-point scales that ranged from 1 (strongly disagree) to 7 (strongly agree). Fatigue Severity Scale scores were computed by averaging participants' responses, with higher values indicating more fatigue.²

² We also administered in the laboratory session a modified version of the *Insomnia Severity Index* (Morin 1993). We did so to evaluate sleep difficulties that participants might have had in the preceding 2 weeks. Analyses showed no differences among groups and no relations between index scores and measures of inhibitory strength. Therefore, this questionnaire will not be discussed further here.

Follow-up week and final interview

Each evening participants completed a *Health Behavior Inventory*. This began by asking participants to indicate on 11-point scales (1) how healthful their diet had been (0 = not at all, 10 = extremely), (2) the degree to which they followed their usual health regimen (e.g., took pills, applied medication—0 = not at all, 10 = very much), and (3) the degree to which they engaged in behavior that physicians would consider risky from a health standpoint (0 = not at all, 10 = very much). The Inventory also asked participants to indicate how much time in minutes they spent flossing, brushing, and exercising to improve their physical condition.

In the final interview, participants completed a 4-item *Post-Study Report*. The first item asked participants how closely they approached the goal of performing the task twice a day for 2 weeks. The second asked participants how closely they approached the goal of swishing for 30 s twice a day for 2 weeks. The third and fourth items asked participants how difficult their computer task was and how difficult it was to swish for 30 s, respectively. Responses were made on 11-point scales ranging from 0 to 10. Endpoints for the first two items were “did not comply” (0) and “complied completely” (10). Those for the second two were “not at all” (0) and “extremely” (10).

Procedure

A female experimenter contacted women who signed recruitment sheets and scheduled them for a first meeting. Just prior to the meetings, the experimenter consulted a randomized stack cards to determine condition assignment.

First meeting

In the weak- and strong training conditions, participants were met, escorted to an experimental chamber, and seated at a table on which was a computer and informed consent agreement. The experimenter began by providing a study overview, describing the study as concerned with the relation between people's health habits and their behavioral and physiological responses to different types of stress. She then asked participants to read and—if they agreed to participate—sign the consent form.

Following the consent procedure, the experimenter provided a set of written instructions that described the computer task and explained how to access and run it. For Strong Training participants, the written instructions began by describing the conflict version of the Stroop task. They stated that participants were to perform the task once in the morning (between 6 a.m. and 12 noon) and once in the evening (between 6 p.m. and 12 midnight), with each trial

period lasting approximately 5 min. They emphasized that it was critical for participants to perform twice a day and perform for full trial periods. Participants were told that if they were forced to interrupt a trial period, they should repeat it. Participants also were told that they should repeat periods in which they failed to attain a success rate of 80% or higher. Initial written instructions for Weak Training participants were identical to those for Strong Training participants except that they indicated that the task would be to identify the print color of words (not color words).

The experimenter reviewed the written computer instructions and then demonstrated how to download and run the assigned program. After demonstrating, the experimenter had participants access and run the program to confirm their understanding. While participants were practicing, the experimenter reminded them that they should strive to be successful at least 80% of the time and repeat trial series in which their performance fell short of this standard. She also noted that performance records would be kept on the server and that contingencies would be in place to encourage instruction compliance. Regarding the latter, two missed computer trial sessions would be allowed each week with no penalty. Three missed sessions in a week would result in loss of 5 USD and could result in termination.

Once participants understood, the experimenter provided written instructions describing the oral rinse task. For both Weak- and Strong Training participants, the instructions indicated that the task would be to swish for 30 s half an ounce of Listerine. The instructions told participants that they should do this once in the morning and once in the evening and that it was critical that they hold the solution in their mouth for a full 30 s. Participants were urged not to spit before 30 s unless they absolutely had to do so.

As she did with the written computer instructions, the experimenter reviewed the written oral rinse instructions. When participants understood, the experimenter gave them a bottle of Listerine, with Strong Training participants receiving the full strength rinse and Weak Training participants receiving the diluted rinse. The experimenter also scheduled a laboratory appointment and established a participation (i.e., inhibitory training) start date 14 days prior to it. Before dismissing participants, the experimenter instructed them to bring to their appointment their Listerine bottle. In exchange, they would receive a toothbrush, some dental floss, and some toothpaste to use during the week that would follow the laboratory session. They also would be given some sheets to complete during the follow-up recording week.

The first meeting procedure for No Training participants was similar to that for the Weak- and Strong Training participants, differing chiefly in three respects. First, it began with a study overview indicating that the study

would extend over 1 week (instead of three). Second, it required participants to consent only to activities associated with the laboratory session and the follow-up recording week. Third, it did not include computer and oral rinse instructions, but instead moved directly from the consent procedure to scheduling a laboratory appointment.

Laboratory session

Participants were met by a male or female experimenter who was blind to the condition to which they had been assigned. They were seated at a table on which was a computer, an intercom with a CALL button, and the questionnaires described earlier. On the floor to the right was a circulating tank (Neslab RTE 10.0) filled with water chilled to 5°C. A remote camera was mounted unobtrusively in a back corner. This allowed the experimenter to detect later the point at which participants began and ended their period of cold tolerance. The experimenter welcomed the participants, asked them to complete the questionnaires, and left the room so they could do so in private.

Participants pressed the intercom CALL button when they were finished. At this point, the experimenter returned to the room, attached the Fusion wrist module, and began an 8 min CV baseline period. During the baseline period, participants leafed through magazines while CV measures were taken at 30-s intervals, starting at 6 min and ending at 8. Magazines were selected for having non-arousing content and tended to be dated. Participants were directed to relax, being still but not uncomfortably stiff. Baseline for each measure was taken as the mean of the five readings for the measure.

When the baseline period was completed, the experimenter returned from the control room, removed the magazines, and placed in front of participants a d2 task packet. Participants read the instructions on the first page of the packet and pressed CALL when they were ready to begin. The call signal prompted the experimenter to play for 3 min an audiotape of a man's voice calling out the word "count" at 3-s intervals. The experimenter took CV samples at 30-s intervals while participants circled, starting at the 0.5 min mark and ending at the 2 min mark.

At the end of the d2 work period, the experimenter turned off the tape and told participants to relax for 3 min.³ After 3 min, the experimenter read the statement below.

It is now time for you to perform the tolerance task. A few seconds from now, I will say 'Please dip'. When I do, you should dip your right hand into the chilling

³ The experimenter took CV samples at 30 s intervals during this rest period, starting at 30 s and ending at 2 min. Values were comparable to those obtained at baseline and did not differ across conditions, $F_s < 1.0$, ns.

tank beside you, immersing your hand to the wrist. You will find that the water is cold, but not intolerable and not so cold that it will harm your hand in any way. I would like you to hold your hand in the water as long as you can, with the goal of holding it in the water for at least 60 seconds. To help you achieve this goal, I will tell you when 30 seconds have passed and when 60 seconds have passed. When you can tolerate the cold no more, you should withdraw your hand and use it to press the CALL button on the intercom. OK, now please dip.

The experimenter turned on a stop watch as soon as participants dipped their hand and stopped it as soon as they withdrew. He or she took CV samples at 20 s intervals so long as hands were immersed and announced “30 s” and “60 s” when appropriate. A few women held their hand in the water for 5 min and were asked to withdraw at this point. Although CV measures were taken for as long as 5 min, they were available for all participants only at the 20 s mark. Thus, we analyzed only the first set of CV values.

The experimenter returned to the chamber as soon as participants withdrew their hand and pressed CALL. He or she thanked the participants, gave them their health inventory sheets, and gave them a bag containing a new (Oral B) toothbrush, a small spool of (Johnson & Johnson) dental floss, and a small tube of (Crest) toothpaste. Before dismissing the participants, the experimenter scheduled the final meeting at which sheets and dental items would be collected and a debriefing would be conducted.

Recording week and final meeting

Participants completed a new health inventory sheet each evening and used the provided dental items in their routine dental care. At the final meeting, they returned the sheets and dental items and were debriefed. Those in the intervention conditions also responded to questions on the *Post-Study Report* and reported the number of computer sessions they completed. Following the debriefing, participants were given 10 USD and four research credits.

Floss use was assessed by measuring in centimeters the floss that remained on floss spools. Toothpaste use was assessed by weighing in grams paste in tubes (total weight minus tube weight). In both cases, higher values were taken to indicate less use. Our original intention was to assess compliance with the computer task instructions by maintaining session records on the server computer. However, unreliable data transmission prevented this. The computer task compliance question on the post-study questionnaire and participants’ reports of the number of sessions they completed were taken as alternative compliance measures.

Compliance with the oral rinse instructions was assessed via the oral rinse question on the post-study questionnaire and by weighing in milliliters the amount of rinse that remained in the (473 ml) Listerine bottles that participants turned in after their training.

Results

Training task difficulty and compliance

Responses to the questions asking intervention participants to rate the difficulty of their computer and oral rinse tasks (*Post Study Report*) are in Table 1. A one-way analysis of variance (ANOVA) on the oral rinse data indicated that ratings were higher for Strong Training participants, $F(1, 29) = 5.71, p = .02$. Computer difficulty ratings were weakly higher for those participants; however, the condition effect in that case was not significant, $F(1, 29) = .71, p = .41$.

Also in Table 1 are (1) participants’ responses to the questions asking the degree to which they complied with the computer and oral rinse instructions, (2) participants’

Table 1 Training task difficulty and compliance measures

Condition	Weak training	Strong training
Computer difficulty		
<i>M</i>	1.23	1.94
SD	1.79	2.65
Rinse difficulty		
<i>M</i>	1.69	4.28
SD	2.53	3.25
Computer compliance		
<i>M</i>	9.15	8.39
SD	0.80	1.34
Rinse compliance		
<i>M</i>	9.31	8.89
SD	1.03	1.49
Sessions completed		
<i>M</i>	26.62	23.83
SD	1.04	3.15
Remaining rinse		
<i>M</i>	66.92	114.72
SD	62.37	111.34
Cell <i>n</i>	13	18

Difficulty ratings were made on 11-point scales with endpoints of 0 (not at all) and 10 (extremely). Compliance ratings were made on 11-point scales with endpoints of 0 (did not comply) and 10 (complied completely). The maximum number of sessions participants could have completed was 28. Rinse was measured in milliliters and taken from a 473 ml bottle

reports of computer trial sessions completed, and (3) the amounts of rinse that remained in the Listerine bottles that were turned in. Values for all measures suggest that compliance was high. Preliminary analysis indicated that the measures were intercorrelated (p 's $\leq .10$). Consequently, we converted values to z-scores and combined (averaged) the scores to create a single compliance index. An ANOVA on the index indicated that compliance was better among Weak Training participants (weak training $M = .21$; strong training $M = -.15$), $F(1, 29) = 4.04$, $p = .05$.

The preceding compliance effect is inelegant insofar as it indicates a training condition confound. However, the effect is understandable, given the greater demands associated with the strong training protocol. Further, it does not limit our ability to interpret positive inhibitory strength results for two reasons. First, in theory, the effect should work against the emergence of group differences on the inhibitory strength measures. The reason is because reduced compliance among Strong Training participants should limit the effectiveness of their training. Thus, to the degree that group differences on strength measures emerge, the indication is that they have done so despite a counterforce.

Second, correlational analyses revealed no reliable relations between the compliance index and measures of inhibitory strength. The closest relation was a marginal ($p = .06$) positive correlation between compliance and reported flossing. Analysis of flossing ratings adjusted for compliance yielded the same condition effects as analysis of the flossing ratings not adjusted for compliance. The indication is that variations in compliance were dissociated with research outcomes of prime interest.

Laboratory session

Equipment failure resulted in the loss of laboratory data from one Strong Training participant. Thus, except where indicated, laboratory analyses were performed on data from 54 participants (No Training $n = 24$, Weak Training $n = 13$, Strong Training $n = 17$).

Questionnaires

One-way ANOVAs on responses to the initial questionnaires indicated no group differences in health beliefs (*Health Belief Inventory*), affect (*Affect Checklist*), or experienced fatigue in the preceding 2 weeks (*Fatigue Severity Scale*). As seen in Table 2, participants indicated that they (1) would like to exercise most days, (2) should have no more than two drinks a week, (3) should floss most days, and (4) should brush even more. They also indicated having experienced relatively little fatigue (no training $M = 2.93$, $SD = 1.34$; weak training $M = 2.84$, $SD = 1.36$; strong training $M = 2.51$, $SD = 1.00$).

Table 2 Health belief responses

Condition	No training	Weak training	Strong training
Exercise			
<i>M</i>	2.54	2.90	2.83
<i>SD</i>	1.22	1.18	1.38
Drinks			
<i>M</i>	1.17	1.08	1.28
<i>SD</i>	0.38	0.28	0.67
Floss			
<i>M</i>	3.23	3.85	3.39
<i>SD</i>	1.15	1.07	1.33
Brush			
<i>M</i>	4.67	5.00	4.94
<i>SE</i>	0.82	0.00	0.24
Cell <i>n</i>	24	13	18

Exercise hours (week) would exercise, *Drinks* drinks (week) should have, *Floss* times (week) should floss, *Brush* times (week) should brush, Response options: 1 = "0–2", 2 = "3–5", 3 = "6–8", 4 = "9–11", 5 = "12 or more"

Although *Fatigue Severity Scale* scores did not differ across groups, they did correlate with several of our inhibitory strength measures, suggesting that prior fatigue experience may have played a role in determining inhibitory strength outcomes. With the latter possibility in mind, we included *Fatigue Severity Scale* scores as covariates in analyses of inhibitory strength measures when the regression of inhibitory strength values onto these scores was reliable.

D2 performance

We intended originally to operationalize d2 performance in terms of the number of letters circled (attempts) and the number of circled letters that were correct (successes). However, virtually all (99%) attempts (circles) proved successful (correct). Therefore, we focused on successes, that is, d2s circled.

The success measure was examined with (1) contrasts that evaluated the linear and quadratic trends across the training conditions, and (2) focused comparisons of adjacent conditions. These procedures allowed evaluation of the hypothesis that inhibitory strength, and therefore performance, would be greater for Strong Training participants than for Control participants, with strength and performance for Weak Training participants falling in between.

Preliminary examination of the data revealed that one Strong Training participant had performance scores that were almost twice as high as the next highest scores and almost twice as high as the number of called counts. These

scores indicated a misunderstanding of instructions; therefore, we eliminated them from our main analyses. Preliminary examination also revealed that the regression of success values onto *Fatigue Severity Scale* scores was reliable ($p = .05$). Consequently, we performed appropriate contrasts and comparisons on values that were covariance-adjusted for experienced fatigue.

Analysis of the adjusted scores indicated a linear trend, $F(1, 49) = 4.83, p = .03$. As expected, values were greater for Strong Training participants than for No Training participants, with values for Weak Training participants falling in between (upper panel of Fig. 3). The quadratic trend did not approach significance, $F(1, 49) = 1.02, p = .32$, indicating that the linear trend accounted for all reliable variance. Focused comparisons of adjacent conditions indicated that values were greater for Strong Training participants than for Weak Training participants, $F(1, 49) = 3.10, p = .04$ (one-tailed), but equivalent for Weak- and No Training participants, $F(1, 49) = .02, p = ns$.⁴

Tolerance times

Cold tolerance times were lost for three No Training participants because their water temperature was inadvertently set above 5°C. Preliminary examination of the data indicated that the regression of tolerance values onto *Fatigue Severity Scale* scores did not approach significance. Therefore, times were not covariance-adjusted prior to analysis. Examination also revealed misleading outlier times in all conditions. In view of this, we analyzed the data by (1) determining the proportion of times in each experimental group that was above the median for the study as a whole (68 s), and (2) performing the contrasts described above on the *arc sine* transformed proportions (Langer and Abelson 1972; Winer 1971). Results were similar to those for the d2 scores (lower panel of Fig. 3). Once again, the linear trend proved reliable, $F(1, 48) = 5.72, p = .02$. By contrast, the quadratic trend did not approach significance, $F(1, 48) = .82, p = .37$. Focused comparisons indicated that values were higher for Strong- than Weak Training participants, $F(1, 48) = 3.11, p = .05$ (one-tailed), and equivalent for Weak- and No Training participants, $F(1, 48) = .15, ns$.

⁴ Results for attempts were redundant to those for successes. They indicated a linear trend ($p = .03$), with values being greater for Strong- than Weak Training participants, but equivalent for Weak- and No Training participants. Analysis of the success data without the covariate indicated a near-reliable linear trend, $F(1, 50) = 3.73, p = .059$, and no quadratic effect, $F < 1.0$.

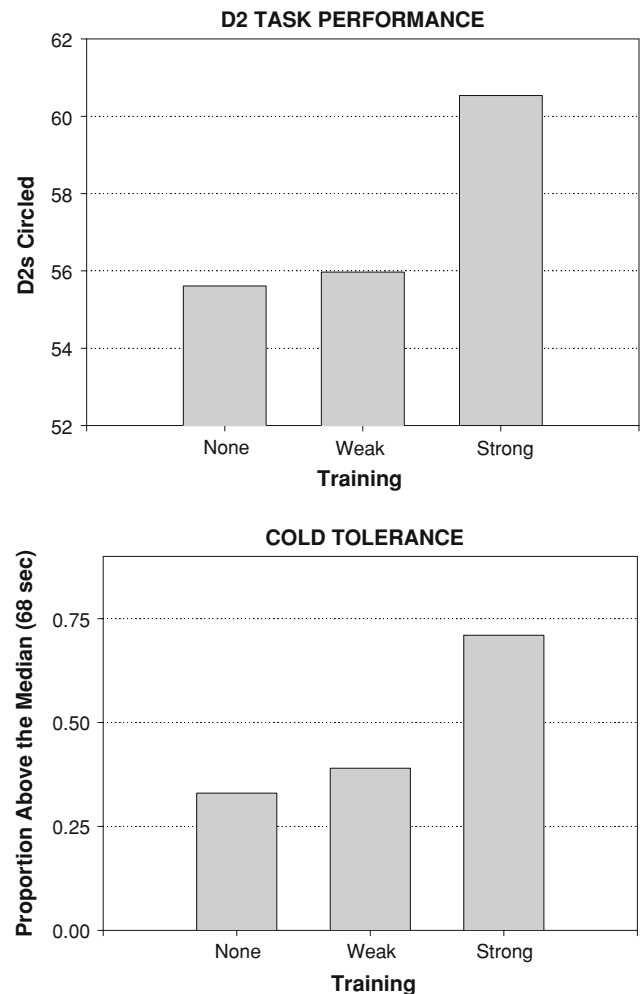


Fig. 3 Covariance-adjusted d2 (upper panel) scores and cold tolerance times (lower panel) for the different training groups. Standard deviations for the adjusted scores were 6.7 (no training), 6.7 (weak training), and 6.7 (strong training)

Cardiovascular data

Initial ANOVAs were performed on the baseline CV data to verify that the experimental groups did not differ in their resting levels of blood pressure and HR. These yielded no effects ($F_s \leq 1.96, p_s \geq .15$).

CV response during the work periods was construed as change from baseline (Jamieson 2004; Llabre et al. 1991). Change scores were computed by subtracting base values from the mean of values in the d2 task period and values from the first sampling period of the cold tolerance period. The change data were first examined with 3 (condition) \times 2 (work period) ANOVAs in which work period was a within subject factor. The regression of change onto *Fatigue Severity Scores* was non-reliable for all change measures; therefore, we did not include experienced fatigue as a covariate. Results indicated condition and period effects, but no interactions (means in Table 3). Because

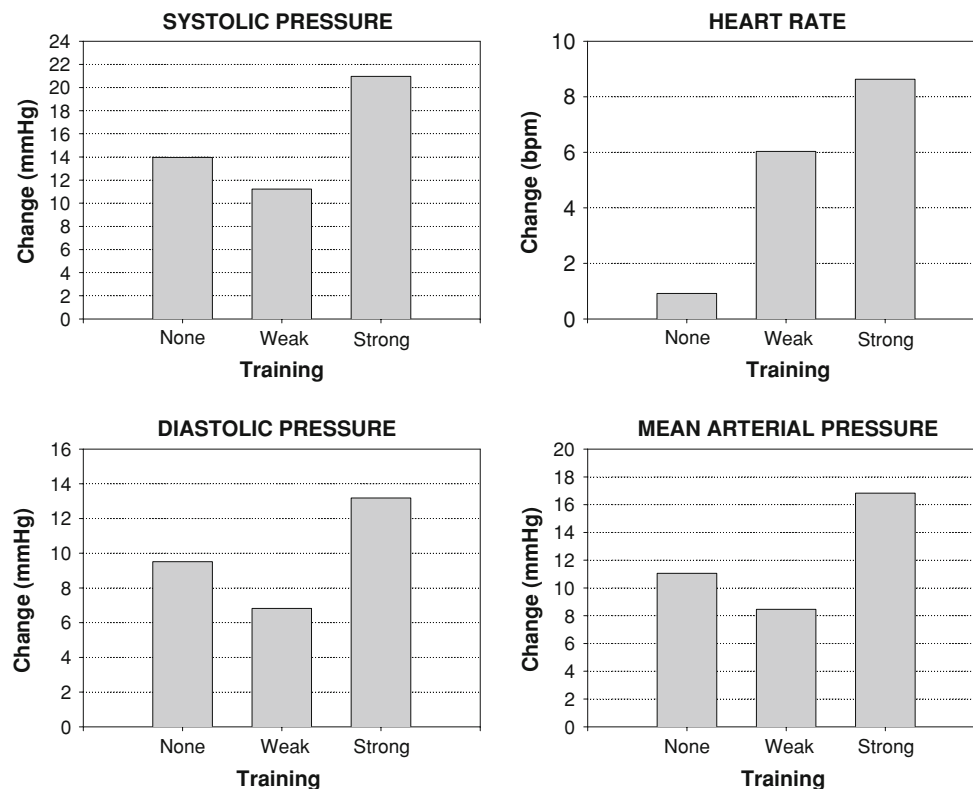


Fig. 4 Blood pressure and HR change scores for the different training groups collapsing across (d2 vs. cold tolerance) work period

Table 3 Blood pressure and heart rate change in the D2 and cold tolerance periods

Condition		No training	Weak training	Strong training
D2 period				
SBP	<i>M</i>	12.64 (10.33)	6.52 (7.05)	16.61 (10.20)
DBP	<i>M</i>	7.66 (6.57)	4.61 (8.12)	9.75 (8.59)
MAP	<i>M</i>	9.64 (7.64)	6.61 (8.33)	13.54 (9.70)
HR	<i>M</i>	0.15 (6.61)	2.81 (5.57)	5.93 (4.35)
Cold tolerance period				
SBP	<i>M</i>	15.28 (8.81)	15.93 (9.26)	25.31 (11.86)
DBP	<i>M</i>	11.34 (8.55)	9.04 (6.90)	16.61 (9.04)
MAP	<i>M</i>	12.45 (8.08)	10.31 (10.0)	19.32 (9.44)
HR	<i>M</i>	1.68 (9.68)	9.25 (9.79)	11.33 (11.00)

Standard deviations are in parentheses. *SBP* systolic blood pressure, *DBP* diastolic blood pressure, *MAP* mean arterial blood pressure, *M* mean, Cell ns for SBP were 20 (no training), 11 (weak training), and 15 (strong training). Those for DBP and MAP were 19 (no training), 11 (weak training), and 15 (strong training)

condition findings were of central interest and did not vary by period, we collapsed across period in evaluating training effects.⁵

⁵ To guard against measurement (e.g., movement) artifact, we omitted change values that deviated by more than two standard deviations from the mean of their group. The regression of change onto baseline values was non-reliable for all CV measures.

It will be recalled that we expected CV responses to be weaker for Strong Training participants than for No Training participants, with responses for Weak Training participants falling in between. For SBP and HR, we tested this using trend contrasts and focused comparisons. For DBP and MAP, we did so using simple ANOVAs and appropriate follow-up pair-wise comparisons. A priori tests were used in the cases of SBP and HR because those CV parameters are believed to be especially sensitive to effort and associated sympathetic responses.

Analysis indicated expected linear trends for both SBP, $F(1, 43) = 6.76, p = .01$, and HR, $F(1, 41) = 9.68, p = .003$. However, contrary to expectation, the trends reflected especially strong responses for Strong Training participants rather than especially weak responses for those participants (Fig. 4). The test of the quadratic trend did not approach reliability in the case of HR, $F(1, 41) = .31, ns$, indicating that the linear effect accounted for all variance. By contrast, the test did prove reliable for SBP, $F(1, 43) = 6.05, p = .02$. Focused comparisons on the HR data indicated greater change for Weak- than No Training participants, $F(1, 41) = 3.84, p = .03$ (one-tailed), and

Footnote 5 continued

Nonetheless, we also examined all change scores with analyses of covariance, including base as the covariate. Findings were virtually identical to those from the ANOVAs.

equivalent change for Weak- and Strong Training participants, $F(1, 41) = .92$, ns. Similar comparisons on the SBP data indicated greater change for Strong- than Weak Training participants, $F(1, 43) = 9.67$, $p = .03$ (one-tailed), and equivalent change for Weak- and No Training participants, $F(1, 43) = .88$, ns. Thus, whereas both CV measures indicated stronger responses in the strong training condition than in the no training condition, one indicated relatively pronounced responses in the weak training condition and the other indicated relatively diminished responses in that condition.

Analyses on the DBP and MAP data yielded results comparable to those for SBP. ANOVAs revealed reliable effects for condition [DBP: $F(2, 42) = 3.36$, $p = .04$; MAP: $F(2, 42) = 5.25$, $p = .009$]. Follow-up comparisons indicated greater change for Strong- than Weak Training participants [DBP: $F(2, 42) = 6.45$, $p = .02$; MAP: $F(2, 42) = 9.49$, $p = .004$] and equivalent change for Weak- and No Training participants [DBP: $F(2, 42) = 1.25$, ns; MAP: $F(2, 42) = 1.10$, ns].⁶

Recording week

One No Training participant terminated during the recording week. Further, some participants failed to complete all report measures. Thus, cell ns and degrees of freedom varied somewhat across analyses of the report week data.

Dental care

Dental care was assessed behaviorally and by way of self-report. Behavioral measures were the amounts of floss and toothpaste that remained at the end of the follow-up week. Preliminary examination of the behavioral data indicated that the regression of floss values onto *Fatigue Severity Scale* scores was reliable ($p = .02$). Therefore, these values were adjusted for experienced fatigue prior to analysis.

Adjusted means for the remaining floss measure are in the upper panel of Fig. 5. Analysis indicated a linear trend, $F(1, 50) = 5.52$, $p = .02$, reflecting less floss (i.e., greater floss use) for Strong Training participants than for No Training participants, with floss for Weak Training participants falling in between. A test of the quadratic trend proved non-reliable, $F(1, 50) = .01$, ns, as did focused

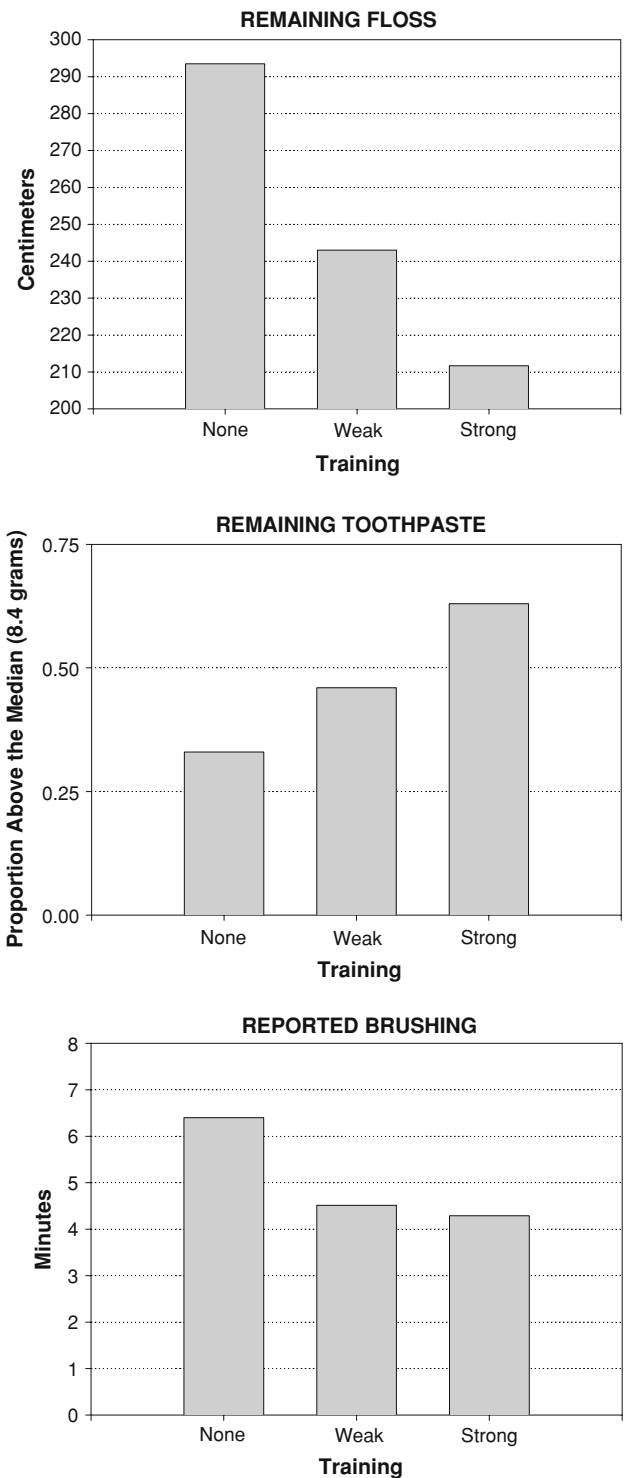


Fig. 5 Remaining dental floss (in centimeters), toothpaste weights (in grams) above the median, and average reported minutes brushing. Remaining floss and reported brushing scores are adjusted for experienced fatigue. Less floss and toothpaste suggest more flossing and brushing, respectively. Lower brushing reports indicate less brushing. Standard deviations for dental floss were 127.0 (no training), 126.9 (weak training), and 127.5 (strong training). Those for reported brushing were 2.8 (no training), 2.7 (weak training), and 2.8 (strong training)

⁶ Additional analyses were performed on the CV change scores excluding scores for the Strong Training participant who had extreme d2 task performance scores. Findings were the same as those obtained with scores for this person included. New Strong Training means for SBP were 17.87 (d2) and 26.10 (cold tolerance). New Strong Training means for DBP were 10.43 (d2) and 17.20 (cold tolerance). New Strong Training means for MAP were 14.64 (d2) and 19.93 (cold tolerance). New Strong Training means for HR were 6.18 (d2) and 12.80 (cold tolerance).

comparisons of the weak training condition with the strong- and no training conditions, $F_s \leq 1.32$, ns.⁷

Like the cold tolerance data, the remaining toothpaste data included misleading outliers, that is, values that diverged markedly from the central tendency of their group. Consequently, we analyzed them in the same way that we analyzed the tolerance data, specifically, by (1) determining the proportion of values in each group that was above the median for the study (8.4 g), and (2) performing contrasts on the *arc sine* transformed proportions. Proportions above the toothpaste median are in the middle panel of Fig. 5. It can be seen that they are ordered oppositely to the floss means. Analysis indicated a near-reliable linear trend, $F(1, 51) = 3.81$, $p = .056$, reflecting more remaining paste (i.e., less brushing) for Strong Training participants. Once again, neither the quadratic trend nor focused comparisons of the weak training condition with the strong- and no training conditions were reliable, $F_s < 1.0$.

Self-report measures of dental behavior were participants' ratings of how many minutes they flossed and brushed each day. To maximize report stability, reports were averaged across the 7 days that data were collected. Preliminary examination of the data indicated that the regression of reports onto *Fatigue Severity Scale* scores was reliable for both measures (p 's $< .02$); therefore, the reports were adjusted for experienced fatigue prior to analysis. Analysis of the floss reports revealed no effects, $F_s \leq 1.10$, ns. Report means were 2.55 (no training), 2.59 (weak training), and 1.85 (strong training). On the other hand, analysis of the brushing reports revealed a linear trend, $F(1, 50) = 5.95$, $p = .02$. Consistent with the behavioral brushing data, reports were lower for Strong Training participants than for No Training participants, with reports for Weak Training participants falling in between (lower panel of Fig. 5). A test of the quadratic trend did not approach reliability, $F(1, 50) = 1.05$, ns; focused comparisons indicated that Weak Training reports were lower than No Training reports, $F(1, 50) = 3.92$, $p = .03$ (one-tailed), but equivalent to Strong Training reports, $F(1, 50) = .05$, ns.⁸

Diet, health regimen, risky behavior and exercise

In addition to asking about dental care, the *Health Behavior Inventory* asked participants (1) to rate the healthfulness of their diet, their adherence to their usual

health regimen, and their engagement in risky health behaviors, and (2) to estimate how many minutes they exercised. Responses were averaged across days. Preliminary examination health behavior ratings indicated that values for the diet, health regimen, and risky behavior measures were compatibly arrayed across conditions, but only weakly intercorrelated (p 's $\geq .09$). Because of the latter, we examined the measures separately. Preliminary examination also indicated that the regression of risky behavior ratings onto *Fatigue Severity Scale* scores was reliable ($p = .03$). Consequently, we adjusted the risky behavior ratings prior to analysis.

Trend analyses indicated only weak linear trends for diet, $F(1, 50) = 2.43$, $p = .14$, and risky behavior, $F(1, 50) = 2.43$, $p = .13$. In both cases, the trends indicated somewhat healthier responses for Strong Training participants than for No Training participants, with responses for Weak Training participants falling in between. Means for diet were 5.68 (no training), 6.33 (weak training), and 6.35 (strong training); those for risky behavior were 1.67 (no training), .93 (weak training), and .83 (strong training).⁹

Initial examination of the exercise data indicated no relation between exercise estimates and *Fatigue Severity Scale* scores. Consequently, the estimates were not adjusted for fatigue. Analyses on the estimates revealed no effects, $F \leq 1.0$, ns.

Discussion

As expected, analysis of the d2 scores and cold tolerance times indicated improved performances in Strong Training participants relative to No Training participants. Performances for Weak Training participants fell between, differing from those of Strong Training participants, but not from those of No Training participants. Analyses also provided evidence of improved health behavior in Strong- relative to No Training participants. Specifically, they indicated that Strong Training participants had less floss remaining at the end of the follow-up week and tended very weakly to have more favorable diet and risky behavior reports than did the No Training participants. Remaining floss for Weak Training participants was intermediate between that for Strong- and No Training participants. Diet and risky behavior reports for Weak Training participants fell between as well, but were aligned closely with those of Strong Training participants. All of these findings comport with the suggestion that the inhibitory system can be strengthened through use (e.g., Muraven and Baumeister

⁷ Analysis of the remaining floss data without the covariate yielded a near-reliable linear trend, $F(1, 51) = 3.84$, $p = .056$, with no quadratic component, $F < 1.0$.

⁸ Analysis of the brushing report measure without the covariate yielded only a linear trend, $F(1, 51) = 6.81$, $p = .01$ (quadratic trend: $F < 1.0$).

⁹ Analysis of the risky behavior measure without the covariate yielded a marginal linear trend, $F(1, 51) = 3.07$, $p = .086$, with no quadratic component, $F < 1.0$.

2000). Of special significance are findings in the Weak Training conditions, which both document the suggestion that reduced regulatory experience should yield reduced inhibitory strength and confirm that regulatory training per se is not sufficient to produce substantial strength improvement.

Unexpected findings

Although the preceding findings fit, or tended to fit, with expectations, other findings did not. Most significantly, the remaining toothpaste and reported brushing data indicated less dental care among Strong Training participants than among No Training participants, with care for Weak Training participants falling in between. One possible explanation follows from the supposition that participants may have had available only a certain amount of time each day for dental care. If this was so, then extra time spent flossing would have necessarily cut into time normally allotted for brushing.

In relation to this interpretation, it may be useful to consider participants' flossing and brushing responses to the *Health Belief Inventory*. The responses suggest that participants valued flossing, but flossed less than they brushed. If Strong Training participants flossed more reliably than usual during the recording week, some dental care periods would have been converted from full time brushing to part time brushing, which would have reduced total brushing time.

A second explanation for the toothpaste and reported brushing effects would follow from the idea that flossing may have reduced the perceived *need* to brush. If heavier flossing Strong Training participants felt a reduced need, they may have brushed less often and for shorter periods.

Also unexpected in this study were participants' CV responses during the d2 and cold tolerance periods. We hypothesized originally that effort-related responses would be reduced for Strong Training participants relative to No Training participants, with responses for Weak Training participants falling in between. In fact, blood pressure and HR responses were stronger for Strong Training participants than for No Training participants. Responses for Weak Training participants varied somewhat, approximating those of Strong Training participants in the case of HR and approximating those of No Training participants in the cases of SBP, DBP, and MAP.

A reasonable interpretation of the unexpected CV responses would draw on the ability/effort logic that drove the original CV predictions (Introduction). In theory, effort should be lower for high- than for low ability people so long as both groups view success as possible and worthwhile. However, where a high ability group perceives success as possible and worthwhile, but a low ability group

does not, members of the high ability group should exert *more* effort because they should engage, whereas members of the low ability group should not. This suggests that Strong Training participants may have displayed more pronounced responses because they (1) had greater inhibitory system strength, and consequently, (2) attempted consistently to meet the set challenges (circling correctly every 3 s and tolerating for at least 60 s), whereas the Weak- and No Training participants did not.

Goal disengagement on the part of Weak- and No Training participants is easy to imagine in the case of the d2 task. Strong Training participants might have accepted the challenge and aimed to circle correctly on every count. Weak- and No Training participants might have rejected the challenge and been willing to circle less frequently and accurately. Notably, the protocol did not provide an explicit incentive for meeting the challenge, which means there was no overt cost associated with failure. Also notably, the protocol implicitly allowed participants to accrue benefit by performing at a lower level. Specifically, it allowed participants to gain approval and avoid disapproval by at least doing something. This could have prevented Weak- and No Training participants from withdrawing effort entirely.

Goal disengagement on the part of Weak- and No Training participants is more difficult to envision in the case of the cold tolerance task, but still plausible. Values in the lower panel of Fig. 3 show that most Weak- and No Training participants tolerated for less than 68 s, whereas most Strong Training participants tolerated for more than 68 s. Thus, most Weak- and No Training participants may have rejected the 60 s tolerance challenge and been preparing to withdraw at or near the 20 s CV sampling point. By contrast, Strong Training participants may have accepted the challenge and been braced for further tolerance when CV responses were assessed.

Difficulty, compliance, and laboratory questionnaires

Analysis of the difficulty and compliance data provided convincing evidence that intervention participants followed their training instructions and some evidence that regulatory demand was greater in the strong training condition. As expected, oral rinse difficulty ratings were higher for Strong Training participants. Computer difficulty ratings were not higher for these participants. However, compliance index scores—which incorporated trial reports and computer compliance ratings—were lower for these participants. This is consistent with the idea that the full experience was more taxing when the training was strong.

Examination of questionnaire responses provided at the beginning of the laboratory session showed no group differences in health beliefs, affect, or fatigue experienced in

the preceding 2 weeks. The health belief data argue against the possibility that effects observed in the follow-up week reflected pre-existing or induced group differences in health behavior values. The other data suggest that the laboratory effects were unlikely to be the result of group differences in affect and that training in the intervention conditions was not so taxing that it tired participants to a noticeable degree.

Although *Fatigue Severity Scale* scores did not differ among conditions, they did correlate with several inhibitory strength measures, including d2 scores, remaining floss, floss reports, brushing reports, and reports of risky health behavior. As noted previously, the correlations present the possibility that fatigue in the 2 weeks prior to the laboratory session was involved in the determination of inhibitory strength outcomes. An intuitive expectation might be that fatigue reports would be negatively associated with inhibitory strength measures. This would follow from the assumptions that (1) 2-week reports implied relatively chronic states with respect to fatigue, and (2) more fatigued participants would tend to do less well on tasks requiring inhibitory strength than less fatigued participants. However, in fact, most *Fatigue Severity Scale* regressions indicated *positive* associations between fatigue scores and inhibitory strength outcomes. The sole exception was the regression involving risky health behavior, which indicated that risky behavior rose as fatigue experience fell. Additional analysis indicated that fatigue scores were only marginally related to a laboratory state tiredness index comprised of the tired and weary items on the Affect Checklist ($r = .24, p = .08$) and that tiredness index scores were even more loosely related to laboratory task performance (number correct: $r = .15, p = .29$; tolerance time: $r = .22, p = .13$).

Mechanisms that could have produced this full set of outcomes are by no means certain. However, it is reasonable to suppose that a three-step process might have been involved. First, there might have been within-condition variation in the degree to which participants spontaneously engaged in regulatory behavior during the critical 2 weeks period. Second, this variation might have translated into variation in felt fatigue and later inhibitory strength, with participants who engaged more experiencing greater fatigue and subsequent strength increments. Third, the improved strength increments could have largely neutralized fatigue by the end of the 2-week period and influenced subsequent inhibitory strength outcomes.

Potentially inconsistent with the process described above is the null group effect that was obtained in the ANOVA performed on the *Fatigue Severity Scale* scores. That is, it could be argued that if spontaneous regulatory efforts generated subjective fatigue effects, then the inhibitory training programs should have generated

subjective effects as well. However, this would overlook the fact that regulatory behaviors are likely to vary in the degree to which they evoke fatigue feelings. Some behaviors (e.g., disciplined jogging) are likely to be strongly evocative, whereas others (e.g., swishing Listerine) may be weakly evocative, at best. Thus, it is possible that within-group fatigue variations in this study were reflective of corresponding variations in highly evocative regulatory behavior.

Regardless of the mechanism or mechanisms responsible for the observed relations between *Fatigue Severity Scale* scores and measures of inhibitory strength, it is important to understand that the experimental effects for the strength measures were not strongly dependent on those relations. Linear effects that proved reliable with the fatigue covariate included either were reliable or very closely approached reliability (p 's < .06) when the covariate was removed (see relevant footnotes).

Alternative interpretations and limitations

Our main findings fit in multiple respects with the idea that regulatory training improves inhibition strength and, thus, inhibitory task performance. However, alternative interpretations can and should be considered. One alternative follows from Festinger's theory of cognitive dissonance (Brehm 1956; Festinger 1957). It would assume that intervention participants experienced dissonance to the degree that their training was demanding (Aronson and Mills 1959) and reduced the dissonance by altering their beliefs about how important it was to do what was asked of them. If Strong Training participants had especially high importance appraisals, they might have been especially likely to accept the d2 and cold tolerance challenges and exert the effort required to meet them. They also might have been especially diligent in flossing and maintaining good health habits during the follow-up week.

The dissonance alternative comports with the fact that intervention participants were provided detailed information prior to the point at which they agreed to participate. It also agrees with the fact that participants had strong participation choice. On the other hand, it hinges on the questionable assumption that participants reduced dissonance by enhancing their appraisals of success importance. More straightforward modes of reduction would have involved (1) elevating the value of the money and credits being earned, (2) enhancing the value of the data being generated, and (3) trivializing the unpleasantness of the training activities (Simon et al. 1995). These modes have been documented repeatedly, whereas the importance mode is speculative. Consequently, the dissonance view should be considered guardedly at present.

Other alternative possibilities are that the main findings reflect (1) the influence of participant self-selection, and (2) enlightenment on the part of Strong Training participants. Regarding the former, one could speculate that some intervention participants were disingenuous when reporting the reason for their termination, suggesting it was due to a factor outside their control when it was not. If this was so and participants with low motivation and/or poor inhibitory control were likely to drop out to the degree their training was difficult, then the intervention conditions could have been weighted with participants who possessed high motivation or inhibitory control, with the strength of these factors being greatest in the strong training condition.

Although we cannot dismiss the self-selection possibility, it seems unlikely. A careful review of reasons for termination suggests that the uncontrollable reasons were credible. Several uncontrollable terminations were due to experimenter or server error. Most others were due to computer difficulties that could not be resolved. One participant reported a death in her family and another reported that she had to leave town to attend a university event.

Regarding the enlightenment possibility, it could be that Strong Training participants gained special insight into what they could endure and, as a result, were willing to withstand more in the laboratory session and follow-up week than they would have in the absence of the insight. This is plausible and worthy of further investigation. However, it assumes that knowledge of endurance capacity inspires endurance, which has not been demonstrated. It also may be true only under certain conditions, which may or may not have been present in this study. Until this possibility is better fleshed out conceptually and empirically, it too should be considered cautiously.

An obvious limitation of this study is its failure to retain all participants who agreed initially to participate. Other limitations are its relatively small cell ns and utilization of a highly select participant population. Smaller ns would be expected to produce less stable findings than larger ns. They also would be expected to increase the difficulty of detecting experimental effects. Thus, for example, we might have detected more differences between the no training and weak training conditions had our ns been larger. The use of the select participant population necessarily restricts our ability to generalize findings to larger groups, including older people, people in poor health, and people outside the university community. Hopefully these limitations can be addressed effectively in future training investigations.

Acknowledgments This research was supported by a grant from the University of Alabama at Birmingham (UAB) Gregory Fleming James Cystic Fibrosis Research Center. Siu-ken Azor Hui is now at the University of Kansas Medical Center, Kansas City, Missouri.

References

- Aronson, E., & Mills, J. (1959). The effect of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology*, *59*, 177–181.
- Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego-depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*, 1252–1265.
- Baumeister, R. F., Gailliot, M., DeWall, C. N., & Oaten, M. (2006). Self-regulation and personality: How interventions increase regulatory success, and how depletion moderates the effects of traits on behavior. *Journal of Personality*, *74*, 1773–1801.
- Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1993). Cardiac psychophysiology and autonomic space in humans: Empirical perspectives and conceptual implications. *Psychological Bulletin*, *114*, 296–322.
- Brehm, J. W. (1956). Post-decision changes in the desirability of alternatives. *Journal of Abnormal and Social Psychology*, *52*, 384–389.
- Brickenkamp, R. (1981). *Test d2* (7th ed.). Göttingen, Germany: Hogrefe.
- Brinkmann, K., & Gendolla, G. H. (2007). Dysphoria and mobilization of mental effort: Effects on cardiovascular reactivity. *Motivation and Emotion*, *31*, 71–82.
- Brinkmann, K., & Gendolla, G. H. (2008). Does depression interfere with effort mobilization? Effects of dysphoria and task difficulty on cardiovascular response. *Journal of Personality and Social Psychology*, *94*, 146–157.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
- Ford, C. E., & Brehm, J. W. (1987). Effort expenditure following failure. In C. R. Snyder & C. E. Ford (Eds.), *Coping with negative life events: Clinical and social psychological perspectives* (pp. 51–79). New York: Plenum.
- Gailliot, M. T., & Baumeister, R. F. (2007). Self-regulation and sexual restraint: Dispositionally and temporarily poor self-regulatory abilities contribute to failures at restraining sexual behavior. *Personality and Social Psychology Bulletin*, *33*, 173–186.
- Gailliot, M. T., Baumeister, R. F., Schmeichel, B. J., DeWall, C. N., Maner, J. K., Plant, E. A., et al. (2007). Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. *Journal of Personality and Social Psychology*, *92*, 325–336.
- Gendolla, G. H. E., & Krüsken, J. (2002). The joint effect of informational mood impact and performance-contingent incentive on effort-related cardiovascular response. *Journal of Personality and Social Psychology*, *83*, 271–285.
- Jamieson, J. (2004). Analysis of covariance (ANCOVA) with difference scores. *International Journal of Psychophysiology*, *52*, 277–283.
- Krupp, L. B., LaRocca, N. G., Muir-Nash, J., & Steinberg, A. D. (1989). The fatigue severity scale: Application to patients with multiple sclerosis and systemic lupus erythematosus. *Archives of Neurology*, *46*, 1121–1123.
- Langer, E. J., & Abelson, R. P. (1972). The semantics of asking a favor: How to succeed without really dying. *Journal of Personality and Social Psychology*, *24*, 26–32.
- Light, K. C., & Obrist, P. A. (1980). Cardiovascular response to stress: Effects of the opportunity to avoid shock, shock experience, and performance feedback. *Psychophysiology*, *17*, 243–252.
- Llabre, M. M., Spitzer, S. B., Saab, P. G., Ironson, G. H., & Schneiderman, N. (1991). The reliability and specificity of delta versus residualized change as measures of cardiovascular

- reactivity to behavioral challenges. *Psychophysiology*, 28, 701–711.
- Marcora, S. M., Bosio, A., & de Morree, H. M. (2008). Locomotor muscle fatigue increases cardiorespiratory responses and reduces performance during intense cycling exercise independent of metabolic stress. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, 294, 874–883.
- Morin, C. M. (1993). *Insomnia: Psychological assessment and management*. New York, London: Gilford Press.
- Muraven, M. B., & Baumeister, R. F. (2000). Self-regulation and the depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, 126, 247–259.
- Muraven, M. B., Baumeister, R. F., & Tice, D. M. (1999). Longitudinal improvement of self-regulation through practice: Building self-control through repeated exercise. *Journal of Social Psychology*, 139, 446–457.
- Muraven, M. B., Shmueli, D., & Burkley, E. (2006). Conserving self-control strength. *Journal of Personality and Social Psychology*, 91, 524–537.
- Muraven, M. B., & Slessareva, E. (2003). Mechanisms of self-control failure: Motivation and limited resources. *Personality and Social Psychology Bulletin*, 29, 894–906.
- Muraven, M. B., Tice, D. M., & Baumeister, R. F. (1998). Self-control as a limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology*, 74, 774–789.
- Oaten, M., & Cheng, K. (2006a). Improved self-control: The benefits of a regular program of academic study. *Basic and Applied Social Psychology*, 28, 1–16.
- Oaten, M., & Cheng, K. (2006b). Longitudinal gains in self-regulation from regular physical exercise. *British Journal of Health Psychology*, 11, 717–733.
- Oaten, M., & Cheng, K. (2007). Improvements in self-control from financial monitoring. *Journal of Economic Psychology*, 28, 487–501.
- Obrist, P. A. (1981). *Cardiovascular psychophysiology: A perspective*. New York: Plenum Press.
- Richter, M., Friedrich, A., & Gendolla, G. H. E. (2008). Task difficulty influences on cardiac reactivity. *Psychophysiology*, 45, 869–875.
- Schmeichel, B. J. (2007). Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *Journal of Experimental Psychology: General*, 136, 241–255.
- Schmeichel, B. J., Vohs, K. D., & Baumeister, R. F. (2003). Intellectual performance and ego depletion: Role of the self in logical reasoning and other information processing. *Journal of Personality and Social Psychology*, 85, 33–46.
- Simon, L., Greenberg, J., & Brehm, J. W. (1995). Trivialization: The forgotten mode of dissonance reduction. *Journal of Personality and Social Psychology*, 68, 247–260.
- Smith, T. W., Baldwin, M., & Christenson, A. J. (1990). Interpersonal influence as active coping: Effects of task difficulty on cardiovascular reactivity. *Psychophysiology*, 27, 429–437.
- Vohs, K. D., & Schmeichel, B. J. (2003). Self-regulation and the extended now: Controlling the self alters the subjective experience of time. *Journal of Personality and Social Psychology*, 85, 217–230.
- Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.
- Wright, R. A. (1996). Brehm's theory of motivation as a model of effort and cardiovascular response. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 424–453). New York: Guilford.
- Wright, R. A. (2008). Refining the prediction of effort: Brehm's distinction between potential motivation and motivation intensity. *Social and Personality Psychology Compass: Motivation and Emotion*, 2, 682–701.
- Wright, R. A., Junious, T. R., Neal, C., Avello, A., Graham, C., Herrmann, L., et al. (2007). Mental fatigue influence on effort-related cardiovascular response: Difficulty effects and extension across cognitive performance domains. *Motivation and Emotion*, 31, 219–231.
- Wright, R. A., & Kirby, L. D. (2001). Effort determination of cardiovascular response: An integrative analysis with applications in social psychology. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 33, pp. 255–307). San Diego, CA: Academic Press.