Joining the "How" and the "Why" to Combat Rumination:

A Novel Intervention Strategy

Eugenia I. Gorlin

B.A., Tufts University, 2008

M.A., University of Virginia, 2012

A Dissertation Presented to the Graduate Faculty of the University of Virginia

Department of Psychology

June 23rd, 2015

Committee Members

Bethany Teachman, Ph.D. (chair)

Gerald Clore, Ph.D.

Benjamin Converse, Ph.D.

Jason Freeman, Ph.D.

Table of Contents

Abstract

Rumination, or negative self-focused thought, is a well-established risk and maintaining factor for depression, anxiety, and a range of other highly prevalent mental health problems. Yet few interventions directly target rumination. To address this gap, the current study developed and tested a novel intervention strategy aimed at reducing rumination and associated negative outcomes. Drawing upon several theoretical frameworks within social and clinical psychology, this intervention paradigm trained participants to reflect on both the "why" and the "how" of a valued personal goal following an initial goal failure, by generating multiple alternative responses specifying both the value of (i.e., "why"), and the strategies by which (i.e., "how"), they might continue to pursue their goal.

Participants in this "proof-of-principle" intervention study were 298 undergraduate students (N=35 for the pilot stage, N=263 for the full intervention trial) with varying trait rumination levels. Participants were randomly assigned to a Why-only, How-only, or Combined (why+how) goal construal training condition or a "free-thinking" Control condition after receiving false negative feedback on an academic test battery. Participants then completed a second test battery, such that the effects of each training condition on subsequent cognitive performance, rumination, and emotional vulnerability could be evaluated.

It was hypothesized that participants in the Combined condition would show the least rumination and negative affect and the most improved cognitive performance, followed by those in the Why-only and How-only conditions, followed by the Control condition. We also tested whether trait rumination, anxiety and depressive symptoms, and working memory would moderate these effects. With respect to moderation effects, both reading comprehension performance and state rumination (as reported on a post-test questionnaire) improved most

significantly in the Combined condition, as expected, but only among highly ruminative or depressed individuals. However, with respect to training effects within the overall sample, our hypotheses were not supported: indeed, the Control condition unexpectedly performed as well or better than the other conditions on most outcomes, defying our assumption that this condition functioned as an inert control. Overall, results were somewhat mixed across outcome measures, and there was no evidence that post-training state rumination (controlling for state rumination at baseline) mediated the effects of training on cognitive performance, suggesting the mechanisms of this novel intervention strategy need to be further clarified.

These lingering questions notwithstanding, the current results help advance our theoretical understanding of the nuanced role of construal and goal-focused processes in rumination. By introducing a novel treatment mechanism that holds promise for reducing rumination and promoting more flexible, resilient goal pursuit, this project addresses a highly prevalent and pernicious mental health concern and paves the way for future, larger intervention trials.

Acknowledgements

With the usual caveat that it would be impossible to list the many friends, mentors, colleagues, and fellow travelers who have helped shape both the "why" and the "how" of my dissertation work, there are a few who bear special mentioning. First and foremost, I want to thank my advisor, Bethany Teachman. Bethany, it's no secret by now that you are the best advisor ever, but there is one thing I'm especially grateful for: Thank you for seeing the scientist hidden beneath my disorganized ramblings, and for helping me bring her out of hiding. Before graduate school, I viewed goals like "getting organized" and "being concise" as frustrating impositions that I was doubtful to ever master. Now, after five years of working with you, such goals no longer feel torturous or unattainable. Instead they seem to fit seamlessly with my own highest objective: to be a creative and thoughtful scientist excelling at the work I love. Clearly, something of you must have rubbed off on me. I'd heard the mantra before that "excellence is its own reward," but I didn't quite know what that meant or how to "live" it in practice until I got to know you. Thank you so much for showing me. Joining your lab was the best decision I've ever made, and I'm excited for many more years of collaboration and PACT family reunions ahead.

Speaking of the PACT family, I must thank my entire princess / arachnophile / superhero clan for all the valuable feedback, brainstorming, social support, and sheer hilarity you have provided over the years. Thank you to Meghan "Spider-Mom" Cody, my first and most formidable PACT lab role model, who has been a trailblazer for me in more ways than one; Shari Steinman, who bewitched me with her infectious enthusiasm about cognitive biases and spiders (and pretty much everything else in life) from the moment I first met her at ABCT back in 2009, and has never disappointed me since; and to all my other PACT lab "big sisters"—Jen Green, Ann Lambert, Meg Reuland—who both led by their impressive example and kindly helped show

me the way. Thank you to Jessica Beadel for being both a "big sister" and a "comrade-in-arms" to me during the sometimes-grueling internship application and dissertation-writing process, making it that much less grueling. Thank you to all my other fellow worshippers of tomfoolery, Jeff Glenn, Erin Maresh, Nauder Namaky, and Karl Fua; the laughs and flights of fancy we have shared have been among my most cherished moments in graduate school. And a special thank you to Alex Werntz, to whom I owe more different kinds of thanks than I can enumerate here; Alex, suffice it to say that you are an inspiration, and I am privileged to have you as a colleague, gossip-buddy, fellow cat-lover, confessor, and friend.

Thank you to my committee members, Jerry Clore, Ben Converse, and Jason Freeman, for their helpful insights and generous feedback on this project, not to mention their moral support. Thank you also to the army of research assistants without whom this project would not exist: Sean Norton, Sarai Arbus, Sarah Carroll, Lelia Schutzenhofer, Gina Vuu, Virginia Clemo, Eugene Kim, Allie Soroka, Emily Meissel, Tara Saunders, Nha-Han Pham, and Austin St. John.

Thank you to my long-time friends and kindred spirits who, despite being scattered across the country, have unfailingly been there to advise, challenge, feed, clothe, shelter, and entertain me as needed. Greg and Karen Salmieri, Debi and Onkar Ghate, this goes especially for you.

Last but not least, I want to thank my parents, Boris and Sophia Gorlin, who had the love and courage and perseverance to uproot their lives 22 years ago so that I could start mine anew. Mama, papa, thank you for giving me this chance at the American dream, while imbuing me with those book-loving, fiddle-playing "Evreyskaya dyevochka" qualities that allowed me to make that dream distinctively my own. Thank you for always being sincere with me, Russian-style, but also for being proud of me, American-style. You have truly given me the "best of both worlds," and I am better for it. This dissertation is dedicated, with love and gratitude, to you.

Joining the "How" and the "Why" to Combat Rumination:

A Novel Intervention Strategy

Depressive rumination, most commonly defined as the tendency to get "stuck" in negative, self-focused thought (Nolen-Hoeksema, 1987), is a long-established risk and maintaining factor for depression (Nolen-Hoeksema, 1991)—a disorder that affects approximately 121 million people worldwide and is estimated to be the most economically burdensome disease in high- and middle-income countries (World Health Organization, 2001, 2008). Moreover, rumination has recently been identified as a transdiagnostic predictor of numerous other highly prevalent forms of psychopathology—including social and generalized anxiety, obsessive-compulsive disorder, borderline personality disorder, alcohol and substance abuse, and eating disorders (Aldao & Nolen-Hoeksema, 2010; Nolen-Hoeksema & Watkins, 2011). Experimental inductions of rumination have further established its causal role in prolonging and amplifying negative mood, impairing task concentration and problem-solving, and disrupting goal-directed behavior (Nolen-Hoeksema, 1987; Lyubomirsky, Kasri, & Zehm, 2003), and rumination has been found to mediate the predictive relationship between stressful life events and later development of psychopathology (Michl, McLaughlin, Shepherd, & Nolen-Hoeksema, 2013). Thus, by treating rumination, there is potential to reduce the cost and suffering associated with a wide range of mental illnesses.

Yet, rumination is not directly targeted by most existing evidence-based interventions, and rumination levels frequently remain elevated following treatment for depression (Roberts, Gilboa, & Gotlib, 1998; Riso et al., 2003), with elevated rumination, in turn, predicting higher rates of relapse (Hood, 2007). The few interventions specifically targeting rumination, such as mindfulness-based cognitive therapy (Teasdale et al., 2002) and rumination-focused cognitive-

behavioral therapy for depression (Watkins et al., 2007), have shown promise in shifting rumination and associated outcomes in the context of depression. However, these interventions remain largely untested in other clinical samples, and, even more crucially, they appear to be least effective in circumstances where they may be most needed, such as when depressive symptoms are preceded by stressful life events (Ma & Teasdale, 2004). Thus, there is a pressing need for new, transdiagnostic interventions that target the common underlying mechanisms of rumination, while allowing for more inclusive implementation across individuals and circumstances. To address this need, the current research developed and tested a novel goal construal training paradigm designed to shift rumination and associated negative outcomes following a goal failure experience. This project also aimed to clarify and extend prior research on the cognitive mechanisms involved in rumination. Toward these ends, a preliminary "proof-of-principle" intervention study was conducted to examine the effects of a brief version of how+why goal-focused training on post-failure cognitive performance, rumination, and emotional vulnerability in a sample with varying trait rumination levels.

Asking "how" versus "why": Construal and rumination

Traditional accounts of rumination largely attribute its impairing, recurrent nature to an overly abstract processing mode, characterized by an excessive focus on the "causes, meanings, and consequences" of one's negative mood states and failed personal goals (Nolen-Hoeksema, 1987; Watkins, 2008). By contrast, shifting to a more concrete, less "why"-focused and more "how"-focused processing mode is typically posited to have beneficial effects (Watkins, 2011). Recent evidence from across disciplines, however, suggests a more nuanced role for both "why"-focused and "how"-focused processing modes in rumination.

With respect to the benefits of relatively concrete, "how-focused" processing, preliminary

findings from a "concreteness training" intervention (Watkins et al., 2012) indicate that thinking more concretely about one's negative emotional states—i.e., focusing on the concrete, moment-to-moment details (such as the sensations in one's body, the specific steps one could take to move forward from the situation, etc.)—helps reduce rumination and associated negative outcomes (such as despondent mood, overgeneral biographical memory, negative global self-evaluations, and impaired problem-solving) in dysphoric individuals (Rimes & Watkins, 2005; Watkins & Teasdale, 2004). Thus, it may be that processing negative emotions or personal failures at a more concrete level helps reduce the tendency to make global, overgeneralized conclusions about the causes and consequences of any given goal failure, while also helping the individual to identify concrete action-steps that might restore goal progress (Watkins, 2008, 2011).

On the other hand, some researchers have obtained seemingly opposite findings regarding the effects of abstract versus concrete thought inductions on rumination and emotional reactivity. For instance, among undergraduate students trained to recall a negative interpersonal event from a distanced, third-person perspective, those induced into a relatively abstract, "why"-focused construal mode (wherein they focused on the reasons underlying their feelings) experienced significantly less anger and negative affect than those induced into a relatively concrete construal mode (wherein they focused on the specific emotions and sensations they had felt; Kross, Ayduk, & Mischel, 2005). Moreover, the degree to which they construed their emotional experience more abstractly (versus more concretely) was shown to mediate the effectiveness of a "distancing" strategy (versus an "immersed" strategy) for down-regulating negative affect (Kross et al., 2005). Similarly, high socially anxious participants induced into a relatively abstract, "analytic" mode of self-focus subsequently reported fewer negative, self-evaluative beliefs than

those induced into a relatively concrete, "experiential" self-focus mode (Wong & Moulds, 2012). In light of this unexpected finding, the researchers speculated that relatively abstract, "analytic" self-focus may have permitted participants to cognitively reappraise the meaning and potential consequences of their performance on the social stressor task, whereas the more concrete, "experiential" condition may have only heightened awareness of their anxiety symptoms.

Taken together, the mixed findings across studies suggest that both "why"-focused (or abstract) and "how"-focused (or concrete) processing may be either harmful or beneficial, depending on the context. However, given the variable content targets to which "abstract" and "concrete" processing have previously been applied in the rumination literature—from emotional states, to negative past events, to a current social stressor, etc.—it is somewhat difficult to draw generalized conclusions about the role of these processes in rumination.

Selecting the content of training: Goal theories and rumination

Given the heterogeneity in the targeted contents of prior manipulations of mental construal mode, we took a theoretically driven approach to selecting a content focus for the present intervention. In particular, we drew from a broad theoretical framework within the goal literature (see "control theory," Carver & Scheier, 1982, and "current concerns theory," Klinger, 1975) that conceptualizes rumination as the involuntary processing of unresolved personal goals, which tend to persist and remain highly accessible until a) the goal is abandoned, or b) goal progress is restored.

Consistent with this account, past findings have shown that depressed and rumination-prone individuals (versus healthy control participants) have difficulty strategically disengaging from failed or unattainable goals (e.g., Kuhl & Helle, 1986). Moreover, they are more likely to engage in unintentional processing of prior goal failures and associated negative emotional states,

which in turn can impair their performance on current goal-pursuits (e.g., Koole et al., 1999;

Koole & Fockenberg, 2011). Thus, an intervention strategy for rumination will likely be

effective to the extent that it promotes disengagement from a failed goal, restores goal progress,

and/or facilitates engagement in a new goal. In some cases, this can be accomplished by mere

distraction; that is, by shifting one's focus to a new goal-pursuit that is wholly unrelated to the

goal failure. This is not always a helpful strategy in practice, however, given that goal-relevant

problem-solving is often required to restore progress (e.g., when a student does poorly on an

exam and then needs to review her errors in order to do better on the next one).

An alternative strategy, then, may be to reframe the nature of the goal being pursued, so

that one's focus shifts from the prior goal failure to those aspects of the goal that are currently

actionable. Indeed, promising findings from one prior study with an unselected sample (Ciarocco,

Vohs, & Baumeister, 2010) indicated that reframing one's current goal in terms of improvement

on a new task (versus in terms of succeeding on the specific task one has just failed) helps reduce

negative rumination and improve task performance following an initial failure. Thus, it may be

that goal-focused thought can facilitate rather than hinder ongoing goal-pursuit following a

failure, depending on how the goal is framed. To examine this possibility, the current study

developed and tested three variants of a goal-focused intervention strategy, whose common aim

was to reduce rumination and restore adaptive goal-pursuit after a failure experience.

Cognitive benefits of abstract and concrete goal construal

Having identified goals as a relevant and potentially fruitful content domain to target in the

context of a recent failure experience, we are in a better position to consider how different

mental construal modes might facilitate a more or less constructive framing of one's goal(s). On

this question, the Construal Level Theory (CLT; Trope & Liberman, 2003) literature in social

psychology offers valuable evidence regarding the contexts in which thinking either about the "how" or the "why" of one's goals may be beneficial. At the broadest level, the CLT framework posits that most objects of human knowledge can be represented at relatively more abstract (i.e., general, superordinate) or relatively more concrete (i.e., narrow, specific, subordinate) construal levels. Thus, for instance, an apple can be construed as "a source of nutrition" (relatively abstract) or "a type of fruit" (less abstract) or "this ripe and juicy snack" (least abstract / most concrete). Similarly, a behavior like "studying" can be construed as an act of "pursuing academic success" (relatively abstract) or "preparing for a test" (less abstract) or "sitting in the library" (least abstract / most concrete). In this conceptual application of mental construal theory to goal-directed behavior, one behavioral construal (A) is more "abstract" than another (B) if A identifies a wider, more superordinate goal of which B is a specific sub-goal or means (e.g., A = "preparing for a test," whereas B = "sitting in the library"). By contrast, A is more "concrete" than B if A identifies a narrower action or instance of goal-pursuit subsumed within the wider goal domain B (e.g., A = "preparing for a test," whereas B = "pursuing academic success"). In other words, A is more abstract than B if it answers the question of "why" B, whereas it is more concrete than B if it answers the question "how" B.

In studies that have experimentally induced a relatively more abstract versus concrete goal construal mode—typically by instructing participants to focus on "why" (abstract) versus "how" (concrete) a behavior is performed—each mode has been shown to benefit certain aspects of cognitive task performance. For instance, inducing an abstract (versus concrete) goal construal mode appears to override the effects of ego depletion (Schmeichel & Vohs, 2009) and improve self-control, presumably by making the value of one's wider goals (e.g., "live a fit and healthy life") more accessible and motivationally salient. In the same vein, inducing an abstract construal

mode helps shield individuals' long-term goal pursuits from short-term temptations (e.g., Fujita et al., 2006). On the other hand, concrete, "how"-focused goal construal (e.g., planning out the steps by which one might reach a particular goal) has been shown to reduce procrastination (McCrea, Liberman, Trope, & Sherman, 2008) and improve reactive inhibitory control (Schmeichel, Vohs, & Duke, 2010). Moreover, with respect to performance on cognitive tasks, abstract and concrete construal modes have been shown to differentially facilitate performance on creative and analytical reasoning tasks, respectively (Foerster, Friedman, & Liberman, 2004).

In light of these mixed findings within both the rumination and broader goal construal literatures, it is likely that both "why"-focused and "how"-focused construal modes may be helpful or harmful in the context of post-failure rumination, depending on current task demands and on the value and actionability of the goal being construed. Furthermore, given evidence for the potential benefits of each construal mode, there may be circumstances when training *both* modes will yield greater benefits than either mode in isolation. This proposal aligns well with the growing body of evidence that cognitive flexibility—i.e., the ability to shift flexibly between different attentional or mental sets in response to changing situational demands—is a core human competency, given its ties to numerous aspects of cognitive, psychological, and even physical health (e.g., Baas, Dreu, Nijstad, 2008; Brockett, LaMarca, & Gould, 2015; Farrant, Fletcher, & Maybery, 2014). Indeed, of direct relevance to the current research, cognitive flexibility is known to be impaired in depressed and especially highly ruminative samples (e.g., Airaksinen, Larsson, Lundberg, & Forsell, 2004; Davis & Nolen-Hoeksema, 2000).

Thus we theorized that the ability to alternate between "why-focused" and "how-focused" goal construal processes after a failure experience, rather than getting "stuck" in one particular mode, might be doubly beneficial. This idea draws further support from Action Identification

Theory (AIT; Wegner & Vallacher, 1987), which suggests that construing one's current goal at the relatively more abstract, "big-picture" level (e.g., "pursuing academic success") may make it easier to disengage from a low-level goal failure (e.g., a poor test score)—especially when accompanied by the more concrete thought process of identifying alternative routes to the same superordinate goal (e.g., "spend more time studying for the next test;" Brunstein & Gollwitzer, 1996). In this light, training individuals to think both concretely *and* abstractly about a personal goal (e.g., improving their academic performance) may have multiplicative effects on their ability to disengage from an initial academic failure and restore successful goal progress on subsequent academic tasks. This may be especially true for individuals who habitually default to negative rumination when "left to their own devices," given the notoriously unconstructive, hard-to-stop, problem- rather than solution-focused thinking and decision-making associated with that response style (e.g., Watkins & Baracaia, 2002). Specifically, prior findings indicate that asking participants to reflect on "why" they pursue their goals induces a relatively abstract construal mode, whereas asking them to reflect on "how" they pursue their goals induces a relatively concrete construal mode (Fujita & Han, 2009; Maglio & Trope, 2012; Schmeichel & Vohs, 2009). The current study examined whether combining these prompts—i.e., asking participants to reflect on both "why" and "how" they might strive to improve their academic performance—would have synergistic positive effects on participants' ability to disengage from rumination about a failed academic goal, thus potentially inoculating them against rumination's negative effects on subsequent task performance and emotional vulnerability.

Potential moderators of training effects

In addition to testing the effects of construal training on state rumination, cognitive task performance, and emotional vulnerability (indexed by self-reported negative affect during the

post-training test battery), the current study examined several potential theoretically-derived moderators of training effects. We tested two competing hypotheses arising from the "deficit-compensation" and "strength-harnessing" theoretical models, respectively (see Bogenschneider & Olson, 1998). Specifically, we examined whether relatively higher baseline levels of depression, anxiety, and/or trait rumination would either undermine or strengthen the effects of training. On the one hand, given that highly dysphoric, rumination-prone individuals are less likely to spontaneously implement an active, goal-focused strategy following failure, they may gain greater benefit from a structured intervention that explicitly instructs them to do so (consistent with the "deficit-compensation" hypothesis). On the other hand, less dysphoric individuals would presumably find it easier to override their unhelpful thoughts about the goal failure, possibly enhancing their ability to benefit from the instructions to change the way they mentally represent the broader goal (consistent with the "strength-harnessing" hypothesis). To test this question, we examined whether individual differences in baseline trait rumination, depression, or anxiety symptoms would moderate the effects of training on post-failure rumination, cognitive task performance, and/or negative affect during the post-training (versus pre-training baseline) test battery.

Finally, we also tested whether individual differences in working memory would moderate the effects of training, given prior research suggesting that working memory interacts with both rumination (e.g., Davis & Nolen-Hoeksema, 1998; Lissnyder, Koster, & Raedt, 2012) and construal mode (e.g., Schmeichel, Vohs, & Duke, 2011) to predict negative outcomes relevant for the current project (such as poor cognitive performance and depressive mood). On the one hand, abstract construal inductions have previously been shown to improve goal maintenance in the face of distractions (e.g., Schmeichel et al.). Thus, it was possible that individuals with lower

working memory—and thus greater difficulty maintaining a task goal in the face of potentially distracting off-task ruminations—would benefit more from "abstract" than "concrete" construal training. On the other hand, given the relatively higher cognitive load that is likely imposed by abstract versus concrete processing (Watkins, 2008), it may be that individuals with lower working memory would be unable to adopt and sustain an abstract construal mode as effectively as those with higher working memory, and thus would benefit *less* from both combined and abstract (relative to concrete) construal training.

Overview and Hypotheses

This "proof-of-principle" intervention study developed and tested a novel "combined" (how+why) goal training paradigm designed to reduce rumination, improve cognitive task performance, and reduce emotional vulnerability following a laboratory-induced goal failure. Specifically, we compared the effects of "why-only," "how-only," and "combined" (why+how) goal-focused training and a "free-thinking" condition on abstract and concrete construal preference, cognitive performance, rumination, and emotional vulnerability following a laboratory-induced academic failure experience (in the form of false negative feedback on their baseline cognitive test performance). Participants first completed baseline self-report measures of trait rumination, depression and anxiety symptoms, and a computerized working memory task, followed by the first of two cognitive test batteries (consisting of reading comprehension, creativity, and interpersonal problem-solving subtests) while their off-task ruminative thoughts were tracked via a thought-tracking procedure. Following the test battery, participants provided retrospective ratings of their ruminative thoughts during testing, and were then given bogus negative feedback about their performance on the tests (thus inducing a state of goal failure). Participants then completed a standard rumination induction procedure to increase focus on their

failure experience, and were then randomly assigned to one of four training conditions—Why-only (focusing on "why" improving their academic performance is of value for them), How-only (focusing on "how" they might improve their academic performance), Combined (focusing on both "why" and "how" they might improve their academic performance), or a "free thinking" Control condition (during which they simply "sat quietly" and "collected their thoughts" in preparation for the next round of cognitive tests). They then completed a second cognitive test battery that was identical in format to the first, while their task performance and rumination were again evaluated. Subjective distress ratings were also collected at several timepoints during each test battery, providing an index of emotional vulnerability.

Data collected during the initial pilot stage ($N = 35$, seven per condition) was used to refine the training procedures for maximum credibility, feasibility, and variability in off-task rumination measures. Only high trait ruminative participants (based on a previously published cutoff on the Ruminative Responses Scale) were recruited to participate in the pilot study, to ensure that our intervention has adequate feasibility and acceptability with vulnerable individuals who represent our ultimate target sample. A number of minor adjustments were made to the training and study procedures based on preliminary results and oral and written feedback provided by the pilot participants (see Appendix C).

Hypotheses. Given that our "why" and "how" prompts were adapted from existing abstract and concrete goal construal manipulation procedures (e.g., Maglio & Trope, 2012), we hypothesized that the Why-only condition would lead to increased abstract construal preference, the How-only condition would lead to increased concrete construal preference, and the Combined condition would lead to high levels of both abstract *and* concrete construal. Furthermore, in light of the potentially synergistic effects of integrating "why"-focused (abstract)

and how"-focused (concrete) goal construal processes, it was hypothesized that those in the

"Combined" condition would show the least post-failure rumination and negative affect and the

best cognitive task performance at post-training, followed by those in the "Why-only" and

"How-only" conditions (which were expected to confer approximately equal benefits), followed

by those in the control condition. Furthermore, as discussed above, we examined whether

baseline individual differences in rumination, depression and anxiety symptoms, or working

memory would moderate the effects of training on rumination, task performance, and emotional

vulnerability. We had two competing hypotheses for each moderation effect, deriving from the

"strength-utilization" and "deficit-compensation" models, respectively.

In sum, the current research is the first to our knowledge to: 1) develop and test a novel

training paradigm that combines "how" and "why" goal-focused processes within a single

combined training phase; 2) examine the effects of "why-only," "how-only," and "combined"

goal-focused training (versus a "free-thinking" control condition) on post-failure cognitive task

performance, rumination, and emotional vulnerability in a sample with variable trait rumination

levels; and 3) examine the role of baseline individual differences in trait rumination, anxiety and

depressive symptoms, and working memory performance in moderating training effects. Thus, in

addition to advancing theoretical knowledge by testing for a more nuanced role of goal-focused

construal mechanisms in rumination and cognitive performance, the current project introduces a

novel intervention strategy to combat a highly prevalent, transdiagnostic risk marker for

psychopathology.

Method

Sample Size and Power Analysis

To determine what sample size would be sufficient for detecting an effect of how+why

training on each of our outcomes, we compared the effect sizes of traditional abstract versus concrete construal training on a range of outcomes within the published literature, including negative affect, global negative self-beliefs, and interpersonal problem-solving. Given the medium-to-large effect sizes reported in a majority of studies (e.g., Kross, Ayduck, & Michel, 2005; Watkins & Baracaia, 2002; Wong & Moulds, 2012;), we conservatively conducted a power analysis on the assumption of a medium sized effect of training condition (d=.60), and determined that a sample size of 38 participants per cell would yield an 80% probability of detecting differences across conditions. Thus, we chose to recruit approximately 50 participants per cell for the "proof-of-principle" intervention study, allowing for slightly more power to detect moderation effects as well as potentially subtle differences between our three active conditions (including the novel "combined" condition, for which no prior effect sizes are available).

Participants and Recruitment

Participants (N = 263; mean age=18.86, SD=1.27; 71.5% female) were undergraduate students recruited either through flyers posted on campus or through the University's psychology department participant pool. Participants were offered course credit or $20 compensation for this single-session, 2-hour study. Eligibility was determined by a set of prescreening measures (collected either via a brief online survey or as part of the department-wide preselection procedures) that included the 22-item Ruminative Responses Scale (RRS; Nolen-Hoeksema & Morrow, 1991) and the modified Sources of Validation Scale (SVS; Harber, 1995). To ensure we had adequate levels of variability within each condition to test our moderation hypotheses, a stratified sampling strategy was used to recruit approximately equal numbers of participants with high (>60; i.e., at least one standard deviation above the mean), medium (42-59; i.e., within one

standard deviation of the mean), and low (<42; i.e., at least one standard deviation below the

mean) levels of trait rumination on the RRS (based on a previously published college student

mean and standard deviation; Yoon & Joormann, 2012). The RRS was chosen as the primary

screener because of its consistent pattern of association with depression, anxiety, and other

emotional disorders (Aldao & Nolen-Hoeksema, 2010), confirming its utility as a transdiagnostic

risk marker for the development of psychopathology. Additionally, to ensure that participants

were highly invested in their academic goals, only those who gave a rating of "3" (very

important) or "4" (extremely important) to the "academic/intellectual goals" item of the SVS

were invited to participate.

Of the 263 participants who signed informed consent, nine participants opted to stop the

study early (four prior to receiving false negative feedback about their performance on the first

test battery, and five after having received the negative feedback but prior to completing the

second test battery).[1] Two participants' computer data were not collected due to experimenter

error. Finally, 38 participants (14.44% of the sample) reported disbelieving the false negative

feedback, and were thus excluded from analyses (though all analyses were re-run with the full

sample to check whether the results changed substantially).[2] There were no differences in rates

of dropout or disbelief of the feedback across training conditions (both $\chi^2 < 4$, p>.10). The final

---

[1] Two of these participants reported wanting to stop due to physical illness, one explicitly noted that she "suffers from anxiety" and was feeling too distressed to continue, and two others stated that they were simply exhausted and were having difficulty concentrating. Several other participants ended the study early because they were running late to class or another appointment. All consented participants were fully debriefed and given payment or credit for their participation, regardless of whether they had completed all study measures.

[2] Results generally followed the same pattern when the "non-believers" were included in analyses; however, several training effects became less pronounced or were no longer significant, in line with our concern that participants who disbelieved the failure feedback did not stand to benefit as much from training (and technically were not representative of the intervention's "target audience," given they had not experienced the failure context that would presumably put them at heightened risk of ruminating). Thus, we continued to exclude these participants from all reported analyses, though results for the full sample are available from the author on request.

sample consisted of 214 participants: 56 in the Why-Only condition, 53 in the How-Only

condition, 54 in the Combined condition, and 51 in the "Free-Thinking" Control condition.

Intervention Conditions

After completing the baseline cognitive task battery (which was described to participants as

an "academic aptitude test" that strongly predicts future academic and intellectual achievement),

participants received a computer print-out that allegedly displayed their individualized

performance feedback. All participants were told that their performance was in the "low average"

or "below average" range on multiple subscales, relative to other students at the same university

(see sample feedback print-out in Appendix B). After viewing the feedback, all participants were

informed that they would be completing a thinking exercise allegedly designed to help them

"process the feedback and prepare for the next round of tests." Participants were then randomly

assigned to one of the four training conditions: "Why-Only," "How-Only," "Combined," or

"Control."

The instructions and training format for the "Why-Only" and "How-Only" training

paradigms were adapted from an existing, well-validated goal construal training procedure (see

Schmeichel & Vohs, 2009), whereas the "combined" training condition was a novel training

paradigm that incorporated both types of prompts (see Appendix A for a flowchart of each active

training condition). The "Free-Thinking" control condition was matched in duration to the three

active training conditions, but with no specific prompts or instructions to adopt a certain

processing mode. Specifically, participants in each training condition underwent the following

procedures, which were administered via Inquisit software (Inquisit, Version 3..6.0; Millisecond

Software, 2011)**:**

"Why-only" training. Participants in the "why-only" condition were asked to generate two

reasons (i.e., higher-order goals) specifying "why" improving their academic importance might

be of personal value for them. For each of the two higher-order goals they specified, they were

then asked to generate one further reason (i.e., higher-order goal) specifying "why" this new goal

was of personal value for them. For example, if the initial reason they specified was that "I want

to graduate with a good GPA," they were then asked to specify one further reason why

"graduating with a good GPA" is of value for them. Thus, there were four total blocks of "why"

prompts (including the two initial prompts and the follow-up prompt for each). Each block

began with a screen that displayed the appropriate prompt (e.g., "Please specify one reason why

improving your reading comprehension, creative ability, and/or problem-solving ability might be

of personal value for you"). After participants typed in their reason (which was limited to a

couple lines of text), they saw a new screen displaying their response for 20 seconds, during

which they were prompted to "think about the value" of the reason they had specified. After 20

seconds, a new screen displaying the next "why" prompt automatically appeared (e.g., "Please

specify one reason why what you just indicated – '[participant's previous response appears here]'

- is of personal value for you"). The response they inputted was again displayed on the next

screen for 20 seconds, followed by the next "why" prompt.

   "How-only" training. Participants in the "how-only" training condition were asked to

generate two alternative means (i.e., lower-order goals) by which they might improve their

academic performance. For each of the two lower-order goals they specified, they were then

asked to generate one further, more specific means (i.e., lower-order goal) specifying "how" they

might accomplish this goal. For example, if one of the initial means they specified was "read

more carefully," they were then asked to specify one further means by which they might

accomplish the goal of "reading more carefully." Thus, there were four total blocks of "how"

prompts (including the two initial prompts and the follow-up prompt for each), which followed the same timing and format as the "why" prompts in the "Why-only" condition.

"Combined" training. Participants in the "Combined" condition were asked to generate one reason (i.e., higher-level goal) specifying "why" and one specific means (i.e., lower-level goal) specifying "how" they might improve their academic performance. Each of these two initial prompts was succeeded by a follow-up prompt of the same type (i.e., a further "why" or a further "how" prompt, respectively). Thus, participants competed four total blocks, all of which were identical in timing and format to the "Why-Only" and "How-Only" conditions. The order of the "how" versus "why" block pairs was counterbalanced across participants.

Note, we recognized that participants would receive a smaller dose of each type of training ("why-focused" and "how-focused") in the "Combined" condition than in the "Why-only" and "How-only" conditions. However, we chose this design because we felt it was more important to match conditions on total dose, and this design provides a more conservative test of the effectiveness of "combined" training, relative to an alternative design in which the total number of trials (and thus the total dose of training) would be doubled.

"Free Thinking" Control condition. Participants in the "Free Thinking" control condition were instructed to sit quietly for four minutes—matched in approximate length to the training paradigm in the other three conditions—and were told, "In a few minutes you will have the opportunity to complete a second round of the same test battery you just completed. To help you process your test results and prepare for the next round of tests, I'd like you to just sit quietly and collect your thoughts for a few minutes." This control condition deliberately limited the guidance and instruction that were provided to participants, on the expectation that they would engage in whatever level of rumination they are naturally prone to engage in following a setback,

particularly when "left to their own devices." To minimize external distractors, participants were asked to put away their cell phones and other personal belongings during this period.

After all study measures were complete (and just before funnel debriefing), participants in the Control condition were asked to complete a Thinking Period Feedback Form (see Appendix C) that prompted them to describe, in as much detail as they could remember, what their thoughts and feelings were during the 4-minute thinking period. They were also asked to rate the proportion of time they spent ruminating, i.e., "thinking about the meaning, causes, and/or consequences of your performance on the 1st academic test battery," on a 5-point Likert-type scale (where 1="Less than 10%"; 3="25-50%"; 5="More than 90%"). Correlations between this item and the other continuous study measures are displayed in Table 4.

Note, the Control condition did not include specific instructions about the use of "concrete" versus "abstract" construal, nor any explicit encouragement to use goal-affirmation strategies more broadly (though some participants may have been primed into a goal-focused mode by the experimenter's mention that the aim of the thinking period was "to help you process your test results and prepare for the next round of tests"). Thus, this condition controlled for two elements that we expected may be critical mechanisms of change in the active conditions (a shift toward strategic, goal-focused processing and a shift in level of construal), though we realized we would not be able to conclude which one or both of these mechanisms may have accounted for condition differences. As a preliminary, "proof-of-principle" test of the intervention's efficacy, we felt it was more important to control for both elements to maximize the likelihood of condition effects on rumination and associated outcomes, and we expected that this study would pave the way for future research aimed at further isolating the relevant intervention mechanism(s).

Materials

Prescreening measures.

The <u>Ruminative Responses Scale</u> (RRS; Nolen-Hoeksema & Morrow, 1991) is a 22-item self-report scale that measures the tendency to respond to depressed or negative mood by ruminating; that is, by dwelling on the causes and consequences of one's negative emotions or symptoms. Participants rate the extent to which each item describes how they typically respond to sad or depressed mood (e.g., "think about how alone I feel"; "think, 'Why do I always react this way?'"; etc.) on a 4-point Likert scale (from 1="almost never" to 4="almost always"). It was administered as a prescreening measure during a preselection battery, such that a stratified sample of participants with high, medium, and low levels of rumination could be recruited (as described under Participants). The RRS was also tested as a baseline moderator of training effects.

The "academic/intellectual goals" item of the modified <u>Sources of Validation Scale</u> (SVS; adapted from Harber, 1995) was also administered during preselection, to ensure that participants attach significant personal importance to this goal domain (which is most directly relevant for the current study). Participants rated this item on a scale ranging from 0 (Not at all important) to 4 (Extremely important), and only participants who endorsed a rating of 3 (Very Important) or 4 were invited to participate.

Baseline symptom and potential moderator measures.

The <u>Depression Anxiety Stress Scales</u> – Depression and Anxiety (DASS-Depression and DASS-Anxiety; Lovibond & Lovibond, 1995) are seven-item subscales that assess symptoms of depression and anxiety, respectively. This measure has adequate psychometric properties (Antony, Bieling, Cox, Enns, & Swinson, 1998; Clara, Cox, Enns, 2001). Participants rate the extent to which each symptom has applied to them in the past week (e.g., "I felt down-hearted and blue"; "I had a feeling of shakiness"; etc.) on a 4-point Likert scale (from 0="Did not apply to me at all" to 3="Applied to me very much, or most of the time"). These subscales were administered at baseline to examine the potential role of preexisting mood symptoms as moderators of training effects (i.e., do persons high in depression and/or anxiety symptoms show either stronger or weaker effects of training than persons who are relatively low in symptoms).

The <u>Running Memory Span</u> task (RMS; Pollack, Johnson, & Knaff, 1959) is a brief, well-established behavioral measure of working memory capacity (WMC; Broadway & Engle, 2010), and was administered at baseline to test whether individual differences in WMC moderate the effects of training. On each trial of the RMS, participants were instructed to report the last *n* letters (ranging from a span of three to eight) in a "running" sequence of letters being presented one-by-one on the screen. There was one block of trials for each letter span (*n*), with six trials in each block. Three of the trials in each block presented only the letters to be recalled ("targets"), whereas the other three included one, two, or three additional letters ("distractors"), respectively. Participants' WMC scores were based on the total number of letters recalled in the correct serial position, following Broadway and Engle (2010).

Rumination induction task.

After receiving false negative feedback regarding their performance on the first cognitive battery, participants completed a standard rumination induction task in an effort to ensure that they ruminated on their failure experience. Specifically, participants were asked to respond to six emotion-focused, self-focused, and symptom-focused prompts adapted from Nolen-Hoeksema and Morrow's (1990, 1993) classic rumination induction task (e.g., "think about what your feelings might mean"; "think about the physical sensations in your body at this moment"; etc.). Participants' self-generated responses to each prompt were then displayed on the screen for 20 seconds, using the same presentation format as in the later training paradigm. Although these prompts make no explicit reference to failure or negative mood, they have been shown to reliably induce negative mood and impair subsequent cognitive task performance (compared to a non-self-focused distraction condition) in dysphoric, rumination-prone samples (e.g., Lyubomirsky, Kasri, & Zehm, 2003; Nolen-Hoeksema, 1991, 1996).

Outcome measures.

Construal mode. The Modified Behavior Identification Form (MBIF) was adapted from Vallacher and Wegner's (1989) Behavior Identification Form, a well-validated measure assessing the tendency to construe everyday actions (e.g., "Making a list") at relatively higher versus lower levels of abstraction (e.g., "Getting organized" versus "Writing things down"). In the modified version that was developed and validated for the current project (see Appendix B), participants separately rated both the "abstract" and the "concrete" descriptions of ten behaviors on a 5-point Likert scale (from 1 = "Not at all descriptive of the behavior" to 5 = "Extremely descriptive of the behavior"). Thus, two separate subscale scores were derived: an "abstract

construal preference" and a "concrete construal preference" score.[3] These subscales were

administered at baseline (during a preselection survey) and again at post-training to examine

whether changes in preferred construal mode occur as a function of the training conditions.

   State rumination measures. The Thought-Sampling Procedure (adapted from Christoff et al.,

2009; Smallwood et al., 2007) was used to concurrently track the frequency, ruminative content,

and emotional valence of off-task thoughts during both the baseline and post-training cognitive

task batteries by the following two methods: 1) computerized sound probes, played at

unpredictable intervals of 60-90 seconds, that prompt participants to indicate whether or not their

thoughts had been task-focused just before the probe was presented (Yes/No); and 2) via

participants' on-line self-reporting of instances when they become spontaneously aware that their

thoughts are not task-focused.

   For each type of off-task thought episode ("probe-caught" and "self-caught") that they

reported, participants were asked the following prompts about their thoughts (that had occurred

either immediately prior to the sound probe, or prior to their awareness of being off-task): 1)

whether their thoughts had been "self-focused," 2) whether their thoughts had been "self-

evaluative," and 3) how positive or negative were their thoughts, on average (on a scale ranging

from 1 = "Very Negative" to 5 = "Very Positive"). Prior to starting the baseline cognitive battery,

participants received detailed instructions about how to report their off-task thought episodes,

and were given examples of when a "Yes" response is appropriate for each item. In line with the

most widely used definition of rumination, which suggests it includes any form of negative, self-

---

[3] Note, this revised construal measure was originally developed and validated to serve as a manipulation check for the How-only, Why-only, and Combined training conditions in the current study; however, given the modifications we made to the traditional abstract and concrete goal construal paradigms to address several other theoretically derived mechanisms tied to rumination (e.g., goal salience and strategy), we ultimately shifted away from conceptualizing our goal-focused training conditions as a "construal mode" manipulation as such. Rather, we tested construal preference as one potential training mechanism by including it among our other primary outcome measures.

focused thought (e.g., Nolen-Hoeksema et al., 2008), rumination frequency was operationally

defined as the number of off-task thought episodes for which a "Yes" response to *either* the self-

focused and/or self-evaluative question was endorsed, and a "1" (Extremely Negative) or "2"

(Slightly Negative) valence rating was endorsed.

The Short Stress State Questionnaire (SSSQ; Helton, 2004) is a 24-item self-report measure

that retrospectively assesses several dimensions of mental content (including negative self-

focused thought, task motivation, and confidence) experienced by participants during a recently

completed task. It was administered following each test battery to further characterize the nature

and quantity of participants' ruminative thoughts during the cognitive tasks. Each item was rated

on a scale ranging from 1 (Not at all) to 5 (Extremely). Although the items of the SSSQ assess a

range of thought contents closely tied to rumination (e.g., "I felt self-conscious," "I was worried

about what other people think of me," etc.), the measure has traditionally been divided into a

three-factor structure consisting of test-related worry, distress, and task engagement (with the

three subscales showing good psychometric properties and relatively good factor invariance over

time; Helton & Naswall, 2015). While the "worry" factor arguably overlaps quite closely with

our "state rumination" construct, defined in the present study as "negative, self-focused thought,"

several of the "worry" subscale items appear to be more heterogeneous and less clearly tied to

rumination (e.g., "I daydreamed about myself"). Given the SSSQ has not been explicitly

conceptualized as a measure of rumination before, we conducted our own exploratory and

confirmatory factor analyses to derive a "clean" rumination subscale with good factor invariance.

These analyses resulted in a 4-item Rumination / Negative Self-Focus subscale (as well as a

second, 3-item Motivation subscale) with excellent model fit and strong measurement invariance

across test periods (see Appendix E for complete model fit and longitudinal invariance statistics,

as well as factor loadings on the selected items). Of course, given we have only tested the model fit of these alternative SSSQ subscales in a single sample, further replication will be needed to confirm and cross-validate this abbreviated two-factor structure.

Emotional vulnerability measure. The <u>Subjective Units of Distress Scale</u> (SUDS; Wolpe, 1969) is a single-item rating used to index self-reported negative state affect, on a scale ranging from 0 (no distress) to 100 (extreme distress). Participants were asked to rate their current distress level at multiple timepoints throughout the study, including at baseline, post-failure induction (i.e., immediately after receiving their negative feedback), post-rumination-induction, post-training, and after the first two tasks of each cognitive battery (thus providing a repeated measure of emotional vulnerability during testing). Participants also completed a final distress rating at the end of the study, so that relaxation exercises could be administered if their distress was still too elevated (i.e., 20 or more points above their baseline level) to ensure that no one left the study distressed.

Cognitive performance measures. Each cognitive measure was split into two versions that were administered during the baseline and post-training cognitive test batteries, respectively. The two batteries were identical in format and duration, each lasting 24 minutes. Administration order for the two versions of each test was counterbalanced across participants.

The <u>Graduate Record Exam (GRE) Reading Comprehension Test</u> is composed of several short reading passages, each followed by a series of multiple-choice questions assessing passage comprehension. This task was chosen because it assesses an academic performance domain that is particularly relevant to our sample of undergraduate students, and because similar reading comprehension tasks have been shown to elicit variable degrees of off-task mind-wandering that, in turn, reliably predict comprehension test performance (e.g., McVay & Kane, 2012; Schooler,

Reichle, & Halpern, 2004). Moreover, previous research has shown that experimentally induced rumination (versus distraction) led to impaired performance on a sample GRE Reading Comprehension Test in dysphoric participants (Lyubomirsky, Kasri, & Zehm, 2003). In the current study, each cognitive test battery included three sample passages from the GRE, with three or four comprehension questions following each passage (for a total of 10 comprehension questions). One set of passages was administered during the baseline cognitive battery, and another during the post-training battery (in counterbalanced order). Participants were given exactly 10 minutes to complete this test, which is approximately half as long as the time typically allotted on the actual GRE—thus increasing the plausibility of the poor performance feedback that participants would receive after the baseline battery.

The Torrance Tests of Creative Thinking – Unusual Uses Test (TTCT-UUT; Torrance, 1962) is a widely used test for assessing creative reasoning, with good reliability and validity (Kim, 2006), in which participants are asked to name as many possible uses as they can for a mundane object (either a "paper clip" or a "brick" in the current study). This task assesses cognitive abilities such as divergent thinking, flexible shifting of mental set, and verbal fluency, all of which have been shown to improve when individuals are induced into a more "abstract" versus a more "concrete" construal mode (e.g., Foerster, Friedman, & Liberman, 2004). In line with Ciarocco, Vohs, and Baumeister (2010), this measure was selected in part because participants are unlikely to have preexisting knowledge or expectancies of what constitutes high performance on the task, and thus are more likely to believe the poor performance feedback they are given after the baseline battery. Participants were instructed to name "as many *possible* but *creative* uses as you can" for each of the two objects, one during the baseline cognitive task battery and one during the second, post-training task battery (object order was counterbalanced).

Of note, this wording of our task instructions deliberately struck a balance between encouraging *quality* and *quantity* of responses, given 1) past research indicating that this measure more meaningfully assesses creative ability when participants are explicitly instructed to be creative (Silvia et al., 2008), but also 2) the need to impose time pressure to maximize the credibility of our later false negative feedback. Specifically, participants were given ten minutes to complete each version of the task.

Following established scoring procedures (see Silvia et al., 2008), responses to the Unusual Uses Test were used to yield two separate performance indices: a "Creative Fluency" index representing the total number of uses participants generated during each test period, and a "Creative Quality" index representing the average creativity (i.e., originality, associative remoteness, and cleverness) of a participant's responses, as rated by independent coders. The scoring criteria for the "Creative Quality" index were adapted from Silvia et al. (2008) and are presented in Appendix D. To prepare the data for scoring, participants' handwritten responses were transcribed into an electronic spreadsheet and randomized, such that each response was rated individually and ratings were not biased by a given participant's poor handwriting, overall quantity or quality of responses, etc.

To facilitate the efficient, reliable rating of the individual responses collected from participants in the current study, we utilized mTurk's crowdsourcing infrastructure to access a large pool of workers with a strong work record (with filters set to require that each worker has completed at least 5000 prior mTurk work assignments with a >98% approval rate). mTurk volunteers who selected the task were redirected to a spreadsheet file that contained detailed rating instructions, a sample scoring rubric (as shown in Appendix D), and a set of ~200 responses for them to rate. (To avoid needless confusion and cognitive load, the rows were

sorted so that all "paperclip" responses appeared together, as did all "brick" responses; the order in which they appeared was counterbalanced across response sets). Workers who successfully completed the rating task were compensated $1; the task took most workers approximately 15-20 minutes to complete. A minimum of four (and a maximum of five) mTurk workers rated each set of 200 responses (~45 sets total), until a reliability criterion of average-measures $ICC>.60$ was reached for that set (indicating good reliability; see Cicchetti, 1994). Ratings for each response were then averaged across coders, and every participant's response ratings within each test period were aggregated to create their overall "Creative Quality" score.

The <u>Means-Ends Problem-Solving</u> test (MEPS; Platt & Spivack, 1975) is a measure of interpersonal problem-solving that has frequently been used to examine problem-solving impairments tied to dysphoric rumination (e.g., Lyubomirsky & Nolen-Hoeksema, 1995; Watkins & Baracaia, 2002). In each MEPS scenario, participants are presented with a hypothetical problem situation and a favorable outcome, and are asked to "fill in the steps" that would enable an individual to get from the beginning to the end of the story. The scenarios and scoring criteria were adapted from Lyubormirsky and Nolen-Hoeksema (1995), who modified the MEPS for use with an undergraduate student sample. One scenario involved a conflict with a friend, while the other scenario involved a conflict with a professor (see Appendix D for the complete scenarios and task instructions). Participants were given four minutes to provide their solutions to each scenario, following Ruby, Smallwood, Sackur, and Singer (2013). Each individual response was coded by one of three raters who were trained to reliability on the first 50 responses (average-measures $ICC=.84$; single-measures $ICC=.63$, indicating good reliability when each response is coded by a "typical" single rater; see Cicchetti, 1994). Each response was rated for the number of listed solutions that were "model solutions" (based on a list of 10 "ideal"

solutions for each problem scenario, adapted from a list of model solutions that was developed and normed by Lyubomirsky & Nolen-Hoeksema, 1995; see Appendix D). Raters also gave an overall "Effectiveness" rating to each response on a 1-7 Likert-type scale (1="Not at all effective," 7="Highly effective"). These two measures—number of model solutions and overall effectiveness—were averaged to produce a maximally reliable "Problem-Solving Effectiveness" aggregate score.

Funnel debriefing. Following completion of the post-training cognitive battery, participants completed a funnel debriefing interview assessing whether they had any suspicion about the credibility of the negative performance feedback they received, or about the overall purpose of the study. Using an adapted version of a funnel debriefing script (see Appendix D for full script), the experimenter began by querying participants' overarching impressions of the study (e.g., "what do you think the point of the study was?"), and asked progressively more specific, targeted questions (e.g., "Did any part of the study seem strange to you?"; "What about the feedback you received after the first set of academic aptitude tests?" etc.). For any "Yes" response a participant gave, the experimenter asked further follow-up questions to determine if and when the participant began to have suspicions, how the suspicions might have affected his or her responses during the study, etc. Participants' responses were coded "1" if they clearly indicated believing the feedback (70.4% of sample), "2" if they indicated having some suspicion but not disbelieving the feedback outright (12.8% of sample), and "0" if they indicated disbelieving the feedback entirely (14.8% of sample). To preserve as much of the data as possible, only participants whose responses were coded a "0" were excluded from analyses.

Procedure

During informed consent, participants were informed that they would be completing a variety of questionnaires asking about their thoughts and feelings, as well as several rounds of academic aptitude tests on which they may receive some feedback. As a cover story for the purpose of the cognitive tasks and feedback, they were told that the goal of the study is to improve our understanding of a well-validated battery of academic aptitude tests, which together have been shown to predict a wide range of important academic and intellectual abilities. Specifically, they were told that the researchers are interested in learning how sensitive these tests might be to various emotional and psychological factors, and whether immediate feedback on the tests helps improve subsequent test performance.

After informed consent, participants completed a baseline measure of state affect (SUDS-distress), followed by the baseline measure of working memory (Running Span). The remaining baseline measures, including trait rumination (RRS), construal level (MBIF), depression and anxiety symptoms (DASS-Depression and Anxiety), demographic information (including past or current history of psychiatric diagnoses and/or treatment, as well as self-reported SAT scores, which were used as a proxy for baseline academic aptitude), were then administered in randomized order. Participants then completed another SUDS rating, and were instructed on how to report their "self-caught" and "probe-caught" off-task thoughts during the upcoming cognitive task battery, which was described as an "academic aptitude test" that is "highly predictive of future academic and intellectual achievement." For increased plausibility, participants were shown a blank copy of the results print-out they would later receive, and were briefly oriented to the alleged "subscales" on which they were being assessed (e.g., "Analytic reasoning ability" and "Inferential reasoning ability" subscales for the Reading Comprehension Test; see sample results

print-out in Appendix D). They then completed the battery of baseline cognitive tasks, beginning

with the Unusual Uses Test and GRE Reading Comprehension Test (administered in

counterbalanced order), followed by the MEPS problem-solving task.[4] Throughout testing,

participants' off-task thoughts were monitored via the thought-sampling procedure, and their

negative affect was monitored via SUDS ratings administered after completion of each task.

Following completion of the baseline test battery, participants completed the SSSQ about

their thoughts and feelings during testing while the experimenter allegedly left to input and print

their test results. The experimenter returned seven minutes later with the fake results print-out,

which included the participant's unique ID number and current date at the top for increased

plausibility. All participants were informed that their performance was "low average" or "below

average" on most indices relative to other students at the university (as shown in the sample

results print-out in Appendix D). After participants were done reviewing the print-out, another

SUDS rating was administered, to check whether the failure feedback had elicited the expected

rise in negative affect. Participants then completed the rumination induction task, followed by

another SUDS rating to check whether negative affect was maintained. Participants were then

reminded that they would have the opportunity to complete a second round of similar academic

aptitude tests in a few minutes, and were randomly assigned to one of the four training

conditions (Why-only, How-only, Combined, or Control), whose stated goal was to help them

"process the feedback and prepare for the next round of tests." After the four-minute training

task, the MBIF was re-administered to check whether the expected construal level(s) were

induced, and another SUDS-distress rating was collected.

---

[4] Given the risk that the effects of our failure induction and subsequent training might "wear off" over the 24-minute test battery, we chose to always administer the Reading Comprehension and Unusual Uses tasks first, as these measures were expected to elicit more off-task rumination—and thus afford more opportunity for differential training effects to emerge—than the relatively briefer, more engaging MEPS problem-solving task.

Participants then completed the second, post-training test battery while their off-task thoughts and negative affect were again monitored. The first two cognitive tasks (Reading Comprehension and the Unusual Uses Test) were administered in the same order for each participant as occurred during their baseline battery, but counterbalanced across participants (e.g., Participant 1 completed the tasks in AB order both times, while Participant 2 completed the tasks in BA order both times). Afterward, participants again provided retrospective self-report ratings of their thought contents during testing via the SSSQ. Participants in the Control condition then completed the Thinking Period Feedback Form, which assessed their thought contents and levels of failure-related rumination during the "free-thinking" period. At the end of the study, participants completed the funnel debriefing to assess whether they had any suspicion about the credibility of the negative performance feedback they received.

Finally, all participants were debriefed, and completed a final SUDS-distress rating to determine whether their distress was still elevated (i.e., at least 20 points higher than baseline), in which case they were offered breathing relaxation exercises. Participants received course credit or payment for participating.

Results

Data Reduction and Scoring

All measures were scored following the original published scoring procedures. Plots of data and descriptive analyses were conducted to reveal any potential skewness or outliers and to verify that statistical assumptions have been met. One extreme outlier, defined as a value deviating by more than three times the interquartile range from the lower or upper quartile of a variable's distribution, was removed from the baseline Divergent Thinking creativity measure. To reduce positive skew, the DASS-Depression variable and the probe-caught and self-caught

rumination frequency variables for each test period were log-transformed. All other continuous variables were approximately normally distributed.

Descriptive Statistics and Zero-order Correlations

A series of one-way analyses of variance (ANOVAs; with follow-up comparisons as needed) and chi-square tests were conducted to determine whether the four training conditions differed on baseline levels of rumination, depression, anxiety, or working memory, as well as on basic demographic variables (i.e., age, gender, ethnicity and race, or past or current psychiatric treatment) and self-reported SAT scores (which were collected as a proxy for general academic aptitude). No baseline differences were found on any baseline variables (*p*>.10) *except* for working memory performance (as indexed by the Running Span task) and average self-reported SAT scores (both *F*>4.5, p<.01), with participants in the Why-Only condition scoring significantly lower on both measures than most or all of the other groups. Thus these two variables were entered as covariates in all subsequent analyses. Table 1 displays the descriptive statistics for each baseline measure by condition, and Table 2 displays the zero-order correlations between all continuous study measures (collapsed across condition).

Overview of Analytic Strategy

For all analyses involving a hypothesized within-subject effect of time (e.g., change in negative affect from pre- to post-failure induction; change in construal from pre- to post-training; etc.), we used a mixed-effects regression modeling approach. This analytic method has several well-documented advantages over more traditional repeated-measures ANOVA approaches, including improved flexibility in modeling continuous-categorical variable interactions and more robust, unbiased handling of missing data (see Nich & Carroll, 1997). Each model was fitted using the "lme4" package in R (R Core Team, 2013; Bates, Maechler, Bolker, & Walker, 2013).

In addition to the fixed-effect predictors entered in each model (as described below), all

models included a random intercept effect of Subject to control for variation in individual

participants' mean levels. Significant effects of Time were probed and plotted according to the

recommendations of Aiken and West (1991). Namely, simple intercepts and simple slopes were

computed for the outcome variable (y) regressed on Time (x_1), using the beta weights derived

from the regression model. For significant Time x Continuous Moderator interactions, simple

intercepts and slopes of Time were computed separately at low (1 SD below mean) and high (1

SD above mean) values of the moderator (x_2), following Preacher, Curran, and Bauer (2006).

Note, given there were considerable amounts of missing data for our baseline covariate and

moderator measures[5] (primarily due to experimenter and/or computer error), we conducted

Little's (1988) chi-square test to check our assumption that the baseline data were missing

completely at random (MCAR), which the test confirmed they were (χ^2=114.62, df=114, p=.466).

Thus, to preserve full statistical power for our regression analyses, we imputed the missing

baseline data using an expectation-maximization (EM) algorithm, which estimates missing

values in a multivariate dataset based on an iterative maximum likelihood approach (Dempster,

Laird, & Rubin, 1977). This is a widely recommended strategy for data imputation in cases

where the data are MCAR (e.g., Little & Rubin, 1987; Schafer, 1997).

Effects of Failure and Rumination Induction on Negative Affect

To check whether our false negative feedback and subsequent rumination induction were

successful at inducing a negative mood state (as indexed by participants' SUDS-Distress ratings),

we conducted a mixed-effects regression model with SUDS-Distress regressed on Time, coded

---

[5] Specifically, there were 23 cases (8.7% of sample) missing from the SAT Average variable and seven cases (3.2% of sample) missing from the baseline trait rumination, depression, anxiety, and working memory variables, respectively.

as an ordered categorical factor with three levels ("Baseline," "Post-Failure Feedback", and "Post-Rumination Induction") As shown in Figure 1, there was a significant main effect of Time (omnibus $F_{(2,422)}$=149.33, p<.001), with follow-up pairwise contrasts indicating that negative affect increased following the failure feedback as expected. Negative affect remained elevated relative to baseline following the rumination induction, though it unexpectedly decreased relative to immediately after the failure feedback. Given the assumption that highly ruminative participants should be more vulnerable to failure and rumination induction procedures, we also tested a moderation model with SUDS-Distress regressed on Time (coded as above), Trait Rumination (RRS, coded as a continuous mean-centered variable), and their 2-way interaction term. As expected, the Time x RRS interaction term was significant (omnibus $F_{(2,419)}$=16.81, p<.001), such that more highly ruminative participants showed a steeper increase in negative affect from baseline to post-failure induction, with no difference from low ruminative participants in the decrease from pre- to post-rumination induction (see Figure 1). Thus, the overall failure induction procedures were successful at elevating negative mood, especially in highly ruminative participants.

Effects of Training on Primary Outcomes

To examine the effects of each training condition on construal preference, rumination, and associated outcomes, as well as the moderation of these effects by trait rumination, depression, anxiety, and working memory, we ran mixed-effects regression models with each outcome regressed on "Training Condition" (coded as a categorical factor with four levels: Combined, Why-Only, How-Only, Control), "Time" (coded as an ordered categorical factor with two levels: Pre-training, Post-training), and their 2-way interaction term, as well as continuous fixed effects of our two covariates (Average SAT and Running Span, both mean-centered for all regression

analyses). To test our hypotheses that the "Combined" condition would lead to the greatest improvement from pre- to post-training, followed by the "Why-Only" and "How-Only" conditions, followed by the "Control" condition, we tested the six corresponding pairwise contrasts (Combined vs WhyOnly, Combined vs HowOnly, Combined vs Control, WhyOnly vs HowOnly, WhyOnly vs Control, and HowOnly vs Control) using the "multcomp" package in R (Hothorn, Bretz, & Westfall, 2008), with a Tukey correction for multiple comparisons. Significant interactions involving Time and one or more of the Condition contrasts were probed and plotted according to the recommendations of Aiken and West (1991). Namely, simple intercepts and simple slopes were computed for the outcome variable (y) regressed on Time (x_1) within each training condition, using the beta weights derived from the regression model (with each respective condition set as the "reference" level).

For outcomes that involved an additional within-subject effect crossed with Time (e.g., Cognitive Performance, which consisted of four task types administered across both test periods: Reading Comprehension, Creativity, Fluency, and Problem-Solving), a multivariate regression with omnibus F-test statistics was first run to determine whether the effects of training condition differed by the type of measurement (e.g., whether the 3-way Condition x Time x Task Type interaction was significant). If so, then subsequent planned contrast analyses were conducted separately within each measurement type; if not, then all non-significant fixed effects involving the Measurement Type factor were dropped from the model, and a random effect of Measurement Type was included to control for random variation across measurements.

To test our moderation hypotheses for each outcome measure, the planned contrast analyses described above were re-run with each of our four theoretically derived moderator variables— trait rumination (RRS), depression symptoms (DASS-D), anxiety symptoms (DASS-A), and

working memory performance (Running Span)—separately entered as a continuous, mean-centered moderator. For significant 3-way Condition x Time x Moderator interactions, simple intercepts and slopes of Time were computed separately at low (1 SD below mean) and high (1 SD above mean) values of the moderator (x_2), following Preacher, Curran, and Bauer (2006).

Note, only significant 2- or 3-way Training Condition x Time interaction effects (at the alpha=.05 level) are noted in the text, given the focus of the current project. However, regression parameter estimates for each Training Condition x Time interaction are displayed in Tables 5-8 and Figures 2-5 (with more detailed statistics for each moderation analysis provided in Appendix F).

See Table 3 for a broad overview of the training effects obtained for each outcome measure.

Abstract and concrete construal preference. As shown in Table 4, planned pairwise contrasts did not reveal any significant 2-way Training Condition x Time interactions for either the MBIF Concrete or MBIF Abstract subscale, and there were no moderation effects for either subscale (all $p>.05$). Thus, there was no conclusive evidence that our brief adapted versions of the traditional "how/why" training paradigm modified either "concrete" or "abstract" construal preference, suggesting the training conditions may be tapping other mechanisms in addition to abstract/concrete construal as such.

Of note, however, prior to correcting for multiple comparisons, there were significant Training Condition x Time interactions predicting the MBIF Concrete subscale, such that the HowOnly condition led to the expected greater increase in concreteness preference than either the Combined (standardized $b=.34$, $p=.024$) or the WhyOnly ($b=.30$, $p=.042$) condition. Though these uncorrected post-hoc comparisons must be interpreted with caution, they are broadly consistent with past findings from the construal literature, which has traditionally only compared

the effects of two conditions—the why-focused "abstract" and the how-focused "concrete" condition—on construal preference and related outcomes.

Cognitive performance. To examine the effects of training condition on changes in cognitive performance from pre- to post-training, participants' performance on each individual task (Fluency, Creativity, Reading Comprehension, and Problem-Solving) at both baseline and post-training was converted to a z-score, and the four z-scores were entered as repeated measurements of the Cognitive Performance outcome variable. The omnibus multivariate regression model revealed a significant Training Condition x Task Type x Time interaction ($F_{(9,1455)}$=1.92, p=.044), suggesting that the effects of training condition on changes in cognitive performance differ across the four tasks. As such, separate regression models were run to estimate the effects of the Training Condition x Time interaction within each subtest, as well as the moderation of these effects by our hypothesized moderators. Table 5 displays the regression statistics for each planned contrast conducted within each task type (Reading Comprehension, Creative Quality, Creative Fluency, and MEPS Problem-Solving). Note, for ease of interpretation, only significant moderation effects (at the alpha=.05 level) are shown in Table 6; a complete list of parameter estimates and related planned contrast statistics for each moderation model is available in Appendix F.

Of note, the omnibus F-test also revealed a significant main effect of Time ($F_{(1,1465)}$=17.28, p<.001), with separate follow-up regression analyses revealing that there was an overall improvement in performance (with standardized beta-weights ranging from .18-.33) on every cognitive measure except Creative Quality (on which there was no overall change from pre- to post-training; standardized b=.02, p=.807).

Reading comprehension. As shown in Table 5, there were no 2-way Condition x Time

interactions predicting reading comprehension. However, when trait rumination was entered as a moderator, there was a significant 3-way Condition (HowOnly vs Combined) x Time x Trait Rumination interaction ($p=.025$); see Figure 2. Specifically, among individuals with relatively higher (but not lower) levels of trait rumination, those in the Combined condition showed greater improvement in reading comprehension performance, in line with our hypotheses. Note, a broadly similar pattern (larger slopes of increase in the Combined condition relative to the others) was observed across all three of our baseline emotional symptom moderators (RRS, DASS-Depression, and DASS-Anxiety), though most pairwise contrasts did not reach significance when correcting for multiple comparisons. See Appendix F for complete regression statistics.

Creative quality. As shown in Table 5, there were significant 2-way Training Condition (WhyOnly vs Combined) x Time ($p<.001$) and Training Condition (Control vs WhyOnly) Time (p=.002) interactions predicting Creative Quality, such that participants in both the Combined and the Control conditions showed greater improvement in their creativity scores than did participants in the WhyOnly condition (see Figure 3). There were no significant moderation effects at the alpha=.05 level (but see Appendix F for an unexpected pattern of marginally significant moderation effects by both DASS-Anxiety and DASS-Depression).

Creative fluency and problem-solving. No significant 2-way or 3-way Condition x Time interactions emerged for either the Fluency or Problem-Solving subtests (see Table 5).

State rumination.

Off-task rumination frequency. To examine the effects of training condition on changes in probe-caught and self-caught rumination levels from pre- to post-training, the number of endorsed rumination episodes of each type (Probe-Caught and Self-Caught) at both baseline and post-training were entered as repeated measurements of the Off-Task Rumination outcome variable. The omnibus multivariate regression model did not reveal a significant Condition x Time x Episode Type interaction ($F_{(3,564)}$=1.07, p>.05); thus, fixed-effect interactions involving Episode Type were removed from the model, and planned contrasts were run with respect to overall Rumination Frequency. Of note, the omnibus F-test again revealed a significant main effect of Time ($F_{(1,572)}$=6.86, p=.009), with the beta-weight from the follow-up regression model indicating that overall off-task rumination frequency decreased from pre- to post-training (standardized b=-.31, p<.001) As shown in Table 6, however, there were no Condition x Time interactions (all p>.05), suggesting the training conditions did not differentially impact the frequency of off-task rumination episodes reported during the second (versus first) test battery.

Retrospectively reported state rumination. Given that the SSSQ Rumination subscale was administered only once per test period (unlike the off-task rumination frequency measure), we ran one set of planned contrasts on the mixed-effects regression model predicting participants' SSSQ Rumination score. Prior to entering the planned condition contrasts, we tested the main effect of Time on SSSQ Rumination, which did not reach significance (standardized *b*=-.08, p=.124). However, as shown in Table 7, there was a significant Condition (Control vs HowOnly) x Time interaction (*p*=.020), such that, unexpectedly, those in the Control condition reported a greater *decrease* in state rumination from pre- to post-training. Figure 4a displays the simple intercepts and slopes for the four training conditions over time.

When examining depression symptoms (DASS-Depression) as a moderator, there was a

significant Condition (Combined vs. Control) x Time x DASS-Depression interaction (p=.026), such that, among those with relatively higher DASS-Depression scores, the Combined condition led to a greater reduction in retrospectively reported state rumination from pre- to post-training than did the Control condition, as expected. By contrast, for those with relatively lower depression symptoms, only the Control condition led to a significant decrease in state rumination (see Figure 4b). Note, this pattern was also replicated for our DASS-Anxiety moderator, though the corresponding Condition (Combined vs. Control) x Time x DASS-Aanxiety interaction failed to reach significance after correcting for multiple comparisons (see Appendix F).

Emotional vulnerability. Finally, to examine the effects of training on participants' emotional vulnerability during subsequent task performance, their self-reported SUDS-Distress ratings after the first and second subtest within each test battery (i.e., Timepoint 1, Timepoint 2) were entered as repeated measurements. (Of note, none of the results changed when including baseline SUDS-Distress as a covariate.) The omnibus multivariate regression model did not reveal a significant 3-way Training Condition x Time x Timepoint interaction ($F_{(3,620)}$=.64, p>.05); thus, this interaction term was dropped from the model prior to conducting planned contrast analyses. The F-test also did not reveal a main effect of Time ($F_{(1,207)}$=.12, p>.05), suggesting there was no overall change in SUDS-Distress ratings during testing from pre- to post-training.

Table 9 displays the results for the Condition x Time planned contrasts, none of which reached significance. When examining Depression Symptoms (DASS-Depression) as a moderator, however, there were significant 3-way Condition (Combined vs WhyOnly) x Time x DASS-Depression (p=.003), Condition (HowOnly vs WhyOnly) x Time x DASS-Depression (p<.001), and Condition (Control vs WhyOnly) x TestPeriod x DASS-Depression (p<.001)

interactions predicting SUDS-Distress; moreover, the Condition (HowOnly vs WhyOnly) x Time x DASS-Depression interaction was replicated across all four moderator variables (including Trait Rumination, Anxiety Symptoms, and Working Memory; see Table 8). Unexpectedly, the WhyOnly condition led to a greater *decrease* in distress among participants with higher depression symptoms (as well as higher trait rumination or anxiety symptoms), whereas it led to a greater *increase* in distress among participants with *lower* depression symptoms (as well as lower trait rumination or anxiety symptoms) (see Figure 5). Finally, with respect to working memory (indexed by Running Span performance), the WhyOnly (versus HowOnly) condition led to a relatively greater *decrease* in participants with *higher* working memory, whereas it led to a greater *increase* in distress among participants with *lower* working memory (see Figure 6).

Mediation of Training Effects by State Rumination

To test our hypothesis that the Combined condition would lead to the largest reduction in rumination and, subsequently, the greatest performance improvement, followed by the Why-Only and How-Only conditions, followed by the Control condition, we created dummy variables corresponding to each of the six planned pairwise contrasts described above (see "Overview of Analytic Strategy"). Following Hayes and Preacher's (2014) recommendations for testing mediation hypotheses with a multicategorical predictor in the SPSS PROCESS macro, we re-ran the mediation model six times, with one dummy variable entered as the predictor (X) on each run and the other dummy variables entered as covariates. Each mediation model tested the indirect effect of the Condition dummy variable on post-training reading comprehension or creativity (as these were the two performance outcomes that showed significant training effects), through either: 1) frequency of off-task rumination episodes (collapsed across episode types) or 2) retrospectively reported state rumination (via the SSSQ Rumination subscale) during the 2nd test

period. Pre-training (i.e., Test Period 1) measures of both the performance outcome and the mediator were included as covariates to control for baseline differences not due to training condition. Finally, given the finding that trait rumination (RRS) moderated the effects of training on reading comprehension, RRS was included as a moderator of the training condition → reading comprehension pathway (see hypothesized model diagrams in Appendix G).

None of the mediation models we tested were significant (i.e., all 95% confidence intervals included 0), suggesting there may be other mechanisms besides self-reported state rumination— such as heightened task motivation, greater self-confidence, or perhaps improved cognitive flexibility—that might better account for the effects of training on cognitive performance.

Discussion

The overall aim of this project was to develop and test a novel how+why goal-focused intervention paradigm designed to reduce post-failure rumination and associated negative outcomes. Specifically, we tested whether Why-only, How-only, or Combined training, compared to a "Free-thinking" Control condition, would improve cognitive performance or reduce rumination and emotional vulnerability (as indexed by self-reported distress during testing) after an initial laboratory-induced goal failure. When comparing training effects within the overall sample, findings were somewhat mixed across outcomes, and only partially in line with hypotheses: namely, with respect to cognitive performance, both the Combined *and* the "free-thinking" Control condition had more positive effects on "creative quality" performance than did the Why-only condition, with no other condition differences on any of the four performance outcomes. With respect to state rumination, the "free-thinking" Control condition had more positive effects on retrospectively reported state rumination than the How-only condition, with no condition differences on rumination frequency; finally, there were no

condition differences with respect to emotional vulnerability, and no conclusive differences with respect to changes in concrete or abstract construal preference.

Examining moderation of these training effects by individual differences in trait rumination and emotional symptoms, however, helped shed some light on this inconsistent pattern: for individuals with *higher* trait rumination or depression symptoms, the Combined condition led to the greatest improvement in reading comprehension performance and the greatest reduction in retrospectively reported state rumination, consistent with the "deficit-compensation" hypothesis. By contrast, for individuals with *lower* (versus higher) rumination or emotional symptoms, the Control condition was again the most helpful across most outcomes (including reading comprehension performance, retrospectively reported state rumination, and emotional vulnerability). With respect to emotional vulnerability, the Why-only condition unexpectedly led to the greatest reduction in distress among highly ruminative, depressed, and anxious participants (whereas the How-only condition *increased* distress in these individuals). Finally, state rumination did not appear to mediate the effects of training condition on cognitive performance, contrary to hypotheses.

Our findings for nuanced effects of training, depending on both individual differences in symptoms and type of outcome, highlight the importance of matching intervention strategies both to the individual and to the task context. Indeed, the variability in our results across outcomes is not unique to this type of single-session, proof-of-principle intervention trial: a review by De Los Reyes and Kazdin (2006) indicated that even the most rigorously controlled, evidence-based intervention studies have shown extremely inconsistent effects across outcomes within the same sample and on the same outcome across samples, suggesting there is a great deal

of work to be done in further specifying *when* and *for whom* a given treatment strategy is effective.

Theoretical Implications: When and for Whom is How+Why Goal-focused Training Beneficial?

Contrary to a common assumption of past rumination work (e.g., Watkins et al., 2008), the current findings suggest that "why"-focused thinking after a failure may not be inherently maladaptive for high-ruminative individuals, especially when paired with more concrete "how"-focused thinking. Indeed, the fact that the Combined condition was most beneficial for high ruminative and/or depressed individuals suggests that goal-focused "how" and "why" processing, particularly when combined with the exercise of shifting flexibly between them, provided a healthier alternative for those whose "default" mode tends to be negative and unconstructive. This is in line with past research linking habitual rumination to insufficient internalization and clarity of personally valued goals (e.g., Teismann et al., 2014; Thomsen, Tonnesvang, Schnieber, & Olesen, 2011). Pending replication in a larger clinical trial, the current results suggest that even a brief, four-minute training focused on articulating both the reasons and strategies associated with a currently actionable goal may be helpful in remediating these deficits. Specifically, the "why" prompts likely helped participants to internalize the goal of "improving their academic performance" by explicitly connecting it to higher-order goals that already have strong personal value (e.g., "becoming a surgeon" or "supporting a family someday"). Meanwhile the "how" prompts may have aided participants in clarifying the concrete action steps they could take to achieve that goal, thus increasing its specificity and perceived attainability.

Interestingly, there did not appear to be any benefit of the "how-focused" condition alone, or much benefit of the "why-focused" condition alone (except with respect to emotional vulnerability, as discussed below). The benefits came primarily from the synergy of these two types of prompts, which could be a function of enhanced flexibility, or of simply making both dimensions of the current goal—its broader personal value and the specific means for achieving it—salient and accessible. This latter interpretation is consistent with past findings that fantasizing about a desired future outcome (as we likely induced participants to do by considering "why" improving their academic performance was of value) is not always beneficial for later goal-pursuit; indeed, it only helps when accompanied by the contemplation of concrete action-steps that would connect the present to the desired future (referred to as "mental contrasting"; Oettingen, 2012).

As a notable exception to this Combined training advantage, the Why-only condition consistently helped reduce emotional vulnerability for high-ruminative, depressed, and anxious individuals (particularly relative to the How-only condition), whereas the Why-only condition *increased* distress for low-ruminative, low-depressed and anxious participants. Again, in line with the deficit-compensation hypothesis, this finding suggests that habitual ruminators may find reflecting on the personal value of the task goal to be a more positive, self-affirming alternative to the emotion-focused and problem-focused rumination to which they likely normally default. By contrast, for healthy individuals who would normally use their own preferred strategies, such as distraction or planning, to recover from a failure experience, the "forced" exercise of reflecting on the value of their goal may only serve to amplify the perceived importance of the goal failure, thus increasing their distress. Unlike the Why-only condition, the How-only condition unexpectedly *increased* high-ruminative and dysphoric participants' distress, perhaps

because they had greater (actual or perceived) difficulty with generating effective strategies, which would have heightened their frustration and distress during the subsequent test battery. Future research that explicitly assesses the perceived effectiveness of participants' strategies, as well as the extent to which they successfully applied their strategies to the subsequent task, would be helpful in clarifying this finding.

Differential effects of training on cognitive performance measures. Of note, reading comprehension was the only cognitive performance measure on which highly ruminative participants differentially benefited from the Combined condition. By contrast, performance on the "creative quality" measure appeared to benefit equally from the Combined and "free-thinking" Control conditions, regardless of trait rumination levels. Though we can only speculate on the reasons for these discrepant findings across the two measures, pending replication and further research, one possible clue lies in the different cognitive sets demanded by the two tasks. Reading comprehension requires carefully constrained, task-focused attention to the material provided in each passage. The creativity task, by contrast, requires a relatively unconstrained flow of spontaneous, free-associative thought. In line with these different task demands, past studies have shown that greater mind-wandering, or "stimulus-independent thought," is associated with poorer reading comprehension performance (e.g., McVay & Kane, 2012; Schooler et al., 2004), whereas experimentally induced mind-wandering *improves* creativity on "divergent thinking" tasks like the one used in the current study (Baird et al., 2012). Relatedly, dysphoric rumination impairs reading comprehension performance (e.g., Lyubomirsky et al., 2003), whereas rumination in formerly depressed individuals has been *positively* linked to divergent thinking performance (Verhaeghen, Joormann, & Khan, 2005). Thus, it is possible that the kind of structured, goal-focused training we provided in the Combined condition is most

beneficial on tasks that require minimizing mind-wandering and constraining one's attention to the current task, whereas the "free-thinking" Control condition may be as or more beneficial on tasks that call for more unconstrained, spontaneous self-reflection (and perhaps *especially* so for rumination-prone participants).

The lack of training effects for the "fluency" dimension of our creativity measure, while not as we hypothesized, comports with past research suggesting that the "creative fluency" metric is a less meaningful index of creative ability than is "creative quality"—particularly when participants are instructed to "be creative" on the task, which likely induces a quality-over-quantity approach (Silvia et al., 2008).

Our lack of condition effects for the interpersonal problem-solving measure (MEPS) is more puzzling, given past findings indicating that dysphoric rumination should impair performance on this measure (e.g., Lyubomirsky & Nolen-Hoeksema, 1995) and that providing task-focused thinking instructions can temporarily overcome this impairment in depressed individuals (Watkins & Baracaia, 2002). Of note, our results also did not replicate these earlier studies' findings that trait rumination and depression are positively associated with impaired MEPS performance after a rumination induction (given the negligible correlations—in the -.01 to .11 range—between post-failure MEPS performance and any of our baseline rumination/emotional symptom measures; see Table 4). One possible explanation for these null findings is that the effects of our failure induction and subsequent four-minute training paradigm may have already worn off by the time participants completed the problem-solving task, which was always administered after the 10-minute reading comprehension and 10-minute creativity task. Future research that varies the dose of training as well as the order of task presentation is needed to help clarify this lack of effects for the problem-solving measure.

Benefits of "free-thinking" control condition, especially for non-ruminators. The "free thinking" control condition was unexpectedly helpful, especially for individuals with low levels of trait rumination and emotional symptoms. This finding suggests that healthy individuals may be well-equipped to choose whatever emotion regulation strategies work best for them when "left to their own devices" after an initial goal failure. Indeed, a qualitative inspection of Control participants' written responses to the Thinking Period feedback prompt revealed that many of them had spontaneously implemented various healthy emotion regulation strategies during this period—such as reappraising the meaning and importance of their test performance, thinking of strategies for improving their performance on the next round, and strategically distracting themselves. Moreover, Control participants' rating of how much time they spent ruminating on the negative feedback during the "free-thinking" period was positively correlated with trait rumination and emotional symptoms, consistent with our interpretation of the different roles that the Control condition played for ruminative versus non-ruminative individuals.

Given that the Control condition had largely positive effects for the overall sample, however, it appears we did not have a truly inert control condition in our study, which makes it difficult to isolate out the active ingredients of our intervention. In particular, there was no condition in which participants were steered away from engaging in goal-focused processing, and all participants (including those in the Control condition) were told that the goal of training was to help them prepare for the next round of tests. Thus, it might be valuable in a future study to include a control condition whose stated aim is "emotional clarity" or "self-understanding"—an aim that more closely resembles high-ruminative individuals' typical motivations for ruminating (e.g., Vine, Aldao, & Nolen-Hoeksema, 2014)—and contrast its effects to those of the goal-focused training conditions.

No conclusive evidence of training effects on construal preference. We did not find clear training effects on either abstract or concrete construal preference (though there was some evidence for the hypothesized greater increase in "concreteness" preference in the How-only training condition, versus the Combined and Why-only conditions, prior to correcting for multiple comparisons). This lack of clear construal effects is not entirely surprising given that we tailored and modified the traditional goal construal manipulaton in several ways to better suit the present purpose of helping participants reframe their goals in a more constructive, non-ruminative manner. For instance, the traditional abstract and concrete construal manipulations instruct participants to begin with a single behavior (e.g., "maintain good personal relationships") and generate four levels of progressively more abstract "why" or progressively more concrete "how" responses, respectively, for that behavior (e.g., Maglio & Trope, 2012; Schmeichel & Vohs, 2009). By contrast, we included only one follow-up "why" or "how" prompt for each initial response, in order to allow for a greater variety of separate reasons and/or strategies to be generated. Our Combined condition further cut the dose of "why" and "how" prompts (respectively) in half, making it even less likely that the training would generalize to an overall stronger preference for abstract and concrete behavioral construals (independently of the particular goal contents being construed). Future trials that test larger doses of both the "how" and "why" components of training, as well as assess more content-specific changes in preferred construal mode (e.g., increases in one's tendency to construe the current task goal both abstractly and concretely), will be needed to help clarify the potential role of construal processes in how+why training.

No mediation by state rumination or moderation by working memory. State rumination during the second test battery did not mediate training effects on cognitive performance,

suggesting it is necessary to consider other, untested mediators in future research. In light of our

moderation findings, it is also likely that different mechanisms are operative for low versus high

ruminative and dysphoric participants. For instance, among highly ruminative participants, the

"Why" prompts in both the Why-only and Combined training conditions may have increased

motivation by making the value of the wider goal more salient, while the "how" prompts may

have enhanced self-efficacy by providing concrete strategies for participants to implement during

the second test battery. Additionally, the Combined condition may have enhanced general

cognitive flexibility (known to be impaired in high-ruminative, dysphoric participants; e.g.,

Davis & Nolen-Hoeksema, 2000) by providing practice with shifting between two processing

modes.   For low-ruminative participants, on the other hand, the Control condition may have

provided more freedom and autonomy for implementing their own preferred strategies, which

may have tapped more heterogeneous but equally effective mechanisms (e.g., distraction and

reappraisal for some, planning and problem-solving for others, etc.).

Finally, there was no moderation by working memory (except in the case of emotional

vulnerability, though the effect was small and only occurred with respect to a single outcome, so

strong inferences cannot be made). This lack of moderation may have simply been due to a

ceiling effect in working memory performance among our cognitively high-functioning

undergraduate sample (a concern that has been raised in prior working memory research with

undergraduate students; e.g., Onraedt & Koster, 2014). However, it may also be the case that

how+why training was equally beneficial regardless of working memory capacity (especially the

relatively simple, four-minute version of training utilized in this initial proof-of-principle study).

Regardless, future research examining the effects of larger training doses on a more varied range

of outcomes will be crucial for clarifying the role of executive function variables, if any, in moderating or mediating the effects of this novel intervention.

Clinical Implications

Clinically, the current results suggest that, on some key tasks, rumination-prone, emotionally vulnerable individuals can benefit most from an intervention that combines "how"- and "why"-focused processing of a currently actionable goal (such as "improving one's academic performance"), particularly when the task requires shifting out of their habitually more ruminative mode into a more task-focused mode. For instance, when a highly ruminative client has received a poor grade on her last chemistry exam and must now closely study the material to prepare for the next exam, a combined how+why goal-processing strategy is likely to be more advantageous than a more open-ended, "free-thinking" alternative or a choice of just one mode (why OR how) for processing her goal. On the other hand, if this same client needs to write a free-response essay for her creative writing class, the most helpful approach may be to let herself engage in the unconstrained, free-associative mode of self-reflection at which she is already quite adept. By providing our rumination-prone clients with metacognitive knowledge about the circumstances in which their self-reflective thinking style is likely to be impairing, versus circumstances in which it may be a valid source of inspiration and creative insight, we may help them develop a more harmonious relationship with their own self-reflections (rather than squelching the genuine desire for greater clarity and insight that often accompanies rumination; e.g., Vine, Aldao, & Nolen-Hoeksema, 2014).

This new intervention strategy could interface well with existing treatment approaches like Acceptance and Commitment Therapy (ACT), mindfulness, and motivational interviewing, given their shared emphasis on bolstering approach motivation by focusing attention on valued

personal goals. Notably, while these existing interventions typically require considerable investment of resources and therapist involvement, the current intervention strategy could be implemented quickly and easily in a self-guided, automated format, making it a potentially more affordable and widely accessible alternative to costlier interventions. With this automated format, there is also potential for further strengthening the intervention by increasing the dose and ultimately customizing the delivery of training: for instance, borrowing from some of the established principles of effective cognitive training (e.g., Wickens, Hutchins, & Cumming, 2013), training could be made "adaptive" by allowing participants to progress through increasingly difficult levels of training that require them to generate more reasons and strategies, or to alternate between reasons and strategies—i.e., between the "why" and the "how"—at more frequent or variable rates.

Limitations and Conclusion

The present findings should be considered in light of several limitations. First, it is important to consider that not all answers to the questions of "why" and "how" are created equal, in that goals can be framed in more or less adaptive ways (respectively) on several key dimensions—e.g., *promotion* ("make good things happen") versus *prevention* ("keep bad things from happening") (Crowe & Higgins, 1997); *learning* (master, grow) versus *performance* (prove, impress) (Grant & Dweck, 2003); and *autonomous* ("I choose to do this because it matters to me") versus *controlled* ("I have to do this because my parents/professors/society tell me to") (Deci & Ryan, 2000). Though we attempted to elicit relatively more promotion-focused, learning-oriented, and autonomous responses from participants through the wording of our "why" and "how" prompts, we did not assess where participants' responses actually fell on these dimensions, so this will be crucial to measure and more explicitly manipulate in future research.

By the same token, it would be valuable to code and explicitly categorize participants' written responses to each prompt in a future qualitative analysis, though this would likely require a larger sample to ensure adequate power.

Second, despite our efforts to increase variability in off-task rumination during each test period following our pilot phase, both by lengthening the battery and increasing the number of sound probes in each task, the base rates and variability in rumination episodes were still quite low; this limited our ability to examine potentially important differences in rumination frequency across tasks and episode types. Future research can help clarify the role of off-task rumination as a possible mechanism by administering longer, less intellectually engaging tasks that afford more opportunities for off-task mind-wandering and rumination (such as the Sustained Attention to Response Task, which has been shown to elicit high levels of mind-wandering; see Robertson, Manly, Andrade, Baddeley, & Yiend, 1997).

Third, in an effort to match the wording of the "how" and "why" prompts on as many dimensions as possible, we allowed participants to respond to the "how" prompt with strategies for improving their performance "on this next set of tests and others like them" (rather than restricting responses to "this next set of tests"). As a result, some participants generated strategies that would not be actionable within the timeframe of the study (e.g., "spend more time at the library"), which may have limited the training's immediate utility and had the unintended effect of increasing frustration and negative affect during the second test battery. Future research should examine the effectiveness of "how"-focused training when participants are encouraged to generate actionable strategies within the goal domain being assessed, and perhaps provided with some guidance in the form of sample strategies from which to choose. Likewise, it would be intriguing in future longitudinal work to test whether why+how training leads to later, more

gradual improvement in cognitive performance and associated outcomes, as would be expected following repeated opportunities to enact and revise one's strategies and actualize the anticipated value of one's goal reengagement.

Finally, the use of an analogue sample of undergraduate students with varying self-reported rumination levels makes it difficult to generalize our results to a diagnosed clinical population. Of note, however, the mean trait rumination (RRS, $M=64.5$), DASS-Depression ($M=8.14$), and DASS-Anxiety ($M=7.07$) levels of participants in the top quartile of each respective measure's distribution were comparable to previously published means in clinically diagnosed samples (e.g., DASS scales: Brown, Chorpita, Korotitsch, & Barlow, 1997; RRS: Watkins & Baracaia, 2002), suggesting that the current sample provided a suitable range and the "high" symptom subsamples were a good analogue.

Finally, given that state rumination did not emerge as a mediator of training effects, and the overall sample (regardless of trait rumination levels) improved on several outcomes in both the Combined and "free-thinking" conditions, it may be more accurate to conceptualize our intervention as promoting more adaptive post-failure coping responses, rather than reducing rumination as such. This broader conceptualization does not diminish the treatment's relevance for highly ruminative and dysphoric samples, given these samples are particularly vulnerable to maladaptive post-failure coping (e.g., Lyubomirsky & Nolen-Hoeksema, 1995; Watkins & Baracaia, 2002) and thus would be especially likely to benefit from an intervention that provides alternative response strategies. This re-conceptualization will be helpful in further refining the intervention and identifying potential mechanisms to target and test in future studies.

Despite these limitations, this is the first study to our knowledge to develop and test a novel combined (how+why) goal construal intervention that shows promise for cheaply and efficiently

reducing rumination and associated negative outcomes. In terms of its theoretical contribution, this project also bridges several previously disparate frameworks from social, cognitive, and clinical psychology to provide novel insight into the role of goal-focused processing in adaptive versus maladaptive self-reflection. Pending the results of future, longer intervention trials that show more robust effects across outcomes, this new treatment paradigm holds potential for addressing a highly prevalent, pernicious mental health problem—simply by training rumination-prone individuals to reflect on both the "why" and the "how" of their valued goals.

References

Aiken, L. S., & West, S. G. (1991). Multiple regression: Testing and interpreting interactions. Newbury Park, London, Sage.

Airaksinen, E., Larsson, M., Lundberg, I., and Forsell, Y. (2004). Cognitive functions in depressive disorders: evidence from a population-based study. *Psychol. Med., 34,* 83–91.

Aldao, A., & Nolen-Hoeksema, S. (2010). Specificity of cognitive emotion regulation strategies: A transdiagnostic examination. *Behaviour Research and Therapy, 48*(10), 974-983. doi: 1.1016/j.brat.201.06.002

Amabile, T. M. (1996). *Creativity in context.* Boulder, CO: Westview Press.

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the depression anxiety stress scales in clinical groups and a community sample. *Psychol Assess, 10*(2), 176-181. doi: 1.1037/1040-359.1.2.176

Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W., Franklin, M. S., & Schooler, J. W. (2012). Inspired by distraction: Mind wandering facilitates creative incubation. *Psychological Science, 23,* 1117-1122.

Bogenschneider, K. & Olson, J. (Eds). (1998). Building resiliency and reducing risk: What youth need from families and communities to succeed. Retrieved from: www.familyimpactseminars.org/doc.asp?d=s_wifis10exec.pdf

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods, 42,* 563-57. doi: 1.3758/BRM.42.2.563

Brockett, A. T., LaMarca, E. A., & Gould, E. (2015). Physical Exercise Enhances

Cognitive Flexibility as Well as Astrocytic and Synaptic Markers in the Medial Prefrontal Cortex. *PLoS ONE, 10*(5): e0124859. doi:1.1371/journal.pone.0124859

Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behavior Research and Therapy, 35*(1), 79-89.

Brunstein, J. C., & Gollwitzer, P. M. (1996). Effects of failure on subsequent performance: The importance of self-defining goals. *Journal of Personality and Social Psychology, 70*(2), 395-407. doi: 1.1037/0022-3514.7.2.395

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 106*(21), 8719-8724. doi: 1.1073/pnas.0900234106

Ciarocco, N. J., Vohs, K. D., & Baumeister, R. F. (2010). Some good news about rumination: task-focused thinking after failure facilitates performance improvement. *Journal of Social and Clinical Psychology, 29,* 1057-1073.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–29.

Clara, I. P., Cox, B. J., & Enns, M. W. (2001). Confirmatory factor analysis of the depression-anxiety-stress scales in depressed and anxious patients. *Journal of Psychopathology and Behavioral Assessment, 23*(1), 61-67. doi: 1.1023/A:1011095624717

Davis, R. N., & Nolen-Hoeksema, S. (2000). Cognitive inflexibility among ruminators

and nonruminators. *Cognitive Therapy and Research*, *24*(6), 699–711.

doi:1.1023/A:1005591412406

De Los Reyes, A., & Kazdin, A. E. (2006). Conceptualizing changes in behavior in intervention

research: The Range of Possible Changes model. *Psychological Review*, *113*(3), 554–583.

doi:10.1037/0033-295X.113.3.554

Deci, E. L., & Ryan, R. M. (2000). The ''what'' and ''why'' of goal pursuit: Human needs and

the self-determination of behavior. *Psychological Inquiry, 11,* 227–268.

Dempster, A.P., Laird, N.M., & Rubin, D. B. (1977). "Maximum Likelihood from

Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society, Series

B*, *39*(1), 1–38.

Donaldson, C., & Lam, D. (2004). Rumination, mood and social problem-solving in

major depression. *Psychological Medicine, 34*(7), 1309-1318. doi:

1.1017/S0033291704001904

Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist*, 41, 1040–

1048.

Farrant, B. M., Fletcher, J., & Maybery, M. T. (2014). Cognitive flexibility, theory of

mind, and hyperactivity/inattention, *Child Development Research, 2014*, Article ID

741543. doi:1.1155/2014/741543

Förster, J., Friedman, R. S., & Liberman, N. (2004). Temporal construal effects on

abstract and concrete thinking: Consequences for insight and creative cognition. *Journal

of Personality and Social Psychology, 87*(2), 177-189. doi: 1.1037/0022-3514.87.2.177

Fujita, K., & Han, H. A. (2009). Moving beyond deliberative control of impulses: The

effect of construal levels on evaluative associations in self-control conflicts.
Psychological Science, 20(7), 799-804. doi: 1.1111/j.1467-928.2009.02372.x

Goldin, P. R., Lee, I., Ziv, M., Jazaieri, H., Heimberg, R. G., & Gross, J. J. (2014). Trajectories of change in emotion regulation and social anxiety during cognitive-behavioral therapy for social anxiety disorder. *Behaviour Research and Therapy, 56,* 7–15. doi:10.1016/j.brat.2014.02.005

Grant, H., & Dweck, C. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology, 85,* 541–553.

Harber, K. (1995). *Sources of validation scale*. Unpublished scale.

Helton, W. S. (2004). Validation of a Short Stress State Questionnaire. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(11), 1238-1242.

Helton, W. S., & Näswall, K. (2015). Short Stress State Questionnaire: Factor structure and state change assessment. *European Journal of Psychological Assessment, 31*(1), 20-3. http://dx.doi.org/1.1027/1015-5759/a000200

Hood, K. (2007). Rumination and distraction as predictors of response and relapse following cognitive therapy or pharmacotherapy for major depression. (68), *ProQuest Information & Learning*, US.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346-363.

Hu, L.T. and Bentler, P.M. (1999), Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives, *Structural Equation Modeling, 6*(1), 1-55.

Huffziger, S., & Kuehner, C. (2009). Rumination, distraction, and mindful self-focus in

depressed patients. *Behaviour Research and Therapy, 47*(3), 224-23. doi: 1.1016/j.brat.2008.12.005

Kenward, M.G., & Roger, J.H. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics, 53*(3), 983–997.

Kim, K. H. (2006). Can we trust reactivity tests? Review of the torrance tests of creative thinking (TTCT). *Creative Research Journal, 18*, 3-14.

Kline P (1993). *The Handbook of Psychological Testing.* Routledge, London.

Knäuper, B., Cannell, C. F., Schwarz, N., Bruce, M. L., & Kessler, R. C. (1999). Improving accuracy of major depression age-of-onset reports in the US national comorbidity survey. *Int J Methods Psychiatr Res, 8*(1), 39-48. doi: 1.1002/mpr.55

Koole, S. L., & Fockenberg, D. A. (2011). Implicit emotion regulation under demanding conditions: The moderating role of action versus state orientation. *Cognition and Emotion,* First published on: 23 February 2011 (iFirst)*.* DOI:1.1080/02699931.201.544891

Koole, S. L., Smeets, K., van Knippenberg, A., & Dijksterhuis, A. (1999). The cessation of rumination through self-affirmation. *Personality Processes and Individual Differences, 77*, 111-125.

Kross, E., Ayduk, O., & Mischel, W. (2005). When asking "why" does not hurt: Distinguishing rumination from reflective processing of negative emotions. *Psychological Science, 16*(9), 709-715. doi: 1.1111/j.1467-928.2005.0160.x

Kuhl, J., & Helle, P. (1986). Motivational and volitional determinants of depression: The degenerated-intention hypothesis. *Journal of Abnormal Psychology, 95*, 247-251.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2015). lmerTest: Tests in Linear

Mixed Effects Models. *R package version 2.0-25*. http://CRAN.R

project.org/package=lmerTest

Lissnyder, E. D., Koster, E. H. W., De Raedt, R. (2012). Emotional interference in working

memory is related to rumination. *Cognitive Therapy and Research, 36,* 348-357.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with

missing values. *Journal of the American Statistical Association, 83,* 1198-1202.

Little, T. D. (2013). *Longitudinal structural equation modeling (methodology in the social

sciences)*. The Guilford Press.

Little, R.J.A., & Rubin, D.B. (1987). Statistical Analysis with Missing Data. New York: John

Wiley & Sons.

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the depression anxiety stress

scales* (2nd ed.). Sydney: Psychology Foundation.

Lyubomirsky, S., Kasri, F., & Zehm, K. (2003). Dysphoric rumination impairs

concentration on academic tasks. *Cognitive Therapy and Research, 27*, 309-33.

Lyubomirsky, S., & Nolen-Hoeksema, S. (1995). Effects of self-focused rumination on

negative thinking and interpersonal problem solving. *Journal of Personality and Social

Psychology, 69*(1), 176-19. doi: 1.1037/0022-3514.69.1.176

Ma, S. H., & Teasdale, J. D. (2004). Mindfulness-based cognitive therapy for depression:

Replication and exploration of differential relapse prevention effects. *Journal of

Consulting and Clinical Psychology, 72*(1), 31-4. doi: 1.1037/0022-006X.72.1.31

Maglio, S. J., & Trope, Y. (2012). Disembodiment: Abstract construal attenuates the

influence of contextual bodily state in judgment. *Journal of Experimental Psychology:

General, 141*(2), 211-216. doi: 1.1037/a0024520

McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict

variation in reading comprehension? On the influence of mind wandering and executive

attention. *Journal of Experimental Psychology: General, 141*(2), 302-32. doi:

1.1037/a0025250

Michl, L. C., McLaughlin, K. A., Shepherd, K., & Nolen-Hoeksema, S. (2013).

Rumination as a mechanism linking stressful life events to symptoms of depression and

anxiety: Longitudinal evidence in early adolescents and adults. *Journal of Abnormal

Psychology, 122*(2), 339-352. doi: 1.1037/a0031994

Morrow, J., & Nolen-Hoeksema, S. (1990). Effects of responses to depression on the

remediation of depressive affect. *Journal of Personality and Social Psychology, 58*(3),

519-527. doi: 1.1037/0022-3514.58.3.519

Ng, F., Trauer, T., Dodd, S., Callaly, T., Campbell, S., & Berk, M. (2007). The validity of

the 21-item version of the depression anxiety stress scales as a routine clinical outcome

measure. *Acta Neuropsychiatrica, 19*(5), 304-31. doi: 1.1111/j.1601-5215.2007.00217.x

Nolen-Hoeksema, S. (1987). Sex differences in unipolar depression: Evidence and

theory. *Psychological Bulletin, 101*(2), 259-282. doi: 1.1037/0033-2909.101.2.259

Nolen-Hoeksema, S. (1991). Responses to depression and their effects on the duration of

depressive episodes. *Journal of Abnormal Psychology, 100*(4), 569-582. doi:

1.1037/0021-843X.10.4.569

Nolen-Hoeksema, S., & Morrow, J. (1991). A prospective study of depression and

posttraumatic stress symptoms after a natural disaster: The 1989 loma prieta earthquake.

Journal of Personality and Social Psychology, 61(1), 115-121. doi: 1.1037/0022-

3514.61.1.115

Nolen-Hoeksema, S., & Watkins, E. R. (2011). A heuristic for developing transdiagnostic

models of psychopathology: Explaining multifinality and divergent trajectories.

Perspectives on Psychological Science, 6(6), 589-609. doi: 1.1177/1745691611419672

Onraedt, T., & Koster, E. H. W. (2014). Training Working Memory to Reduce Rumination.

PLoS ONE, *9*(3): e90632. doi:10.1371/journal.pone.0090632

Platt, J. J., & Spivack, G. (1975). *Manual for the Means-Ends Problem-Solving (MEPS):

A measure of interpersonal problem-solving skill.* Philadelphia: Hahnemann Medical

College and Hospital.

Pollack, I., Johnson, L. B., & Knaff, P. R. (1959). Running memory span. *Journal of

Experimental Psychology*, *57*, 137-146.

Pornprasertmanit, S., Miller, P., Schoemann, A., & Rosseel, Y. (2014). semTools: Useful tools

for structural equation modeling. *R package version 0.4-6.* http://CRAN.R-

project.org/package=semTools

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for

assessing and comparing indirect effects in multiple mediator models. *Behavior Research

Methods, 40*(3), 879-891. doi: 1.3758/BRM.4.3.879

Reinecke, A., Waldenmaier, L., Cooper, M. J., & Harmer, C. J. (2013). Changes in automatic

threat processing precede and predict clinical changes with exposure-based cognitive-

behavior therapy for Panic Disorder. *Biological Psychiatry, 73*(11), 1064-1070.

Rickels, K., & Schweizer, E. (1990). The clinical course and long-term management of

generalized anxiety disorder. *J Clin Psychopharmacol, 10*(3 Suppl), 101S-110S.

Rimes, K. A., & Watkins, E. (2005). The effects of self-focused rumination on global

negative self-judgements in depression. *Behaviour Research and Therapy, 43*(12), 1673-

1681. doi: 1.1016/j.brat.2004.12.002

Riso, L. P., du Toit, P. L., Blandino, J. A., Penna, S., Dacey, S., Duin, J. S., Pacoe, E. M., Grant, M. M., & Ulmer, C. S. (2003). Cognitive aspects of chronic depression. *Journal of Abnormal Psychology, 112*(1), 72-8. doi: 1.1037/0021-843X.112.1.72

Roberts, J. E., Gilboa, E., & Gotlib, I. H. (1998). Ruminative response style and vulnerability to episodes of dysphoria: Gender, neuroticism, and episode duration. *Cognitive Therapy and Research, 22*(4), 401-423. doi: 1.1023/A:1018713313894

Robertson, I. H., Manly, T., Andrade, J., Baddeley BT, & Yiend, J. (1997). Oops: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia, 35,* 747–758.

Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data*. New York: Chapman & Hall.

Schmeichel, B. J., & Vohs, K. (2009). Self-affirmation and self-control: Affirming core values counteracts ego depletion. *Journal of Personality and Social Psychology, 96*(4), 770-782. doi: 1.1037/a0014635

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., Martinez, J., L., & Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts, 2,* 68–85.

Smallwood, J., McSpadden, M., & Schooler, J. W. (2007). The lights are on but no one's home: Meta-awareness and the decoupling of attention when the mind wanders. *Psychonomic Bulletin & Review, 14*(3), 527-533. doi: 1.3758/BF03194102

Teachman, B. A., Marker, C. D., & Smith-Janik, S. B. (2008). Automatic associations and panic

disorder: Trajectories of change over the course of treatment. *Journal of Consulting and Clinical Psychology,* 76(6): 988–1002.

Teasdale, J. D., Moore, R. G., Hayhurst, H., Pope, M., Williams, S., & Segal, Z. V. (2002). Metacognitive awareness and prevention of relapse in depression: Empirical evidence. *Journal of Consulting and Clinical Psychology, 70*(2), 275-287. doi: 1.1037/0022-006X.7.2.275

Teismann, T., Het, S., Grillenberger, M., Willutzki, U., & Wolf, O. T. (2014). Writing about life goals: Effects on rumination, mood and the cortisol awakening response. *The Journal of Health Psychology*, 19(11), 1410-19. doi: 10.1177/1359105313490774

Torrance, E. P. (1966). *The torrance tests of creative thinking-norms-technical manual research edition-verbal tests, forms a and b-figural tests, forms a and b*. Princeton, NJ: Personnel Press.

Unsworth, N., & McMillan, B. D. (2012). Mind wandering and reading comprehension: examining the roles of working memory capacity, interest, motivation, and topic experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication. doi: 1.1037/a0029669

Updegraff, J. A., & Suh, E. M. (2007). Happiness is a warm abstract thought: Self-construal abstractness and subjective well-being. *The Journal of Positive Psychology, 2,* 18-28.

Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology, 57*(4), 660-671. doi: 1.1037/0022-3514.57.4.660

Verhaeghen, P., Joorman, J., & Khan, R. (2005). Why we sing the blues: the relation between

self-reflective rumination, mood, and creativity. *Emotion, 5*(2), 226-32.

Vine, V., Aldao, A., & Nolen-Hoeksema, S. (2014). Chasing clarity: Rumination as a strategy for making sense of emotions. *Journal of Experimental Psychopathology, 5*(3), 229-243.

Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. *Psychological Bulletin, 134*(2), 163-206. doi: 1.1037/0033-2909.134.2.163

Watkins, E. R. (2011). Dysregulation in level of goal and action identification across psychological disorders. *Clinical Psychology Review, 31*(2), 260-278. doi: 1.1016/j.cpr.201.05.004

Watkins, E., & Baracaia, S. (2002). Rumination and social problem-solving in depression. *Behaviour Research and Therapy, 40*(10), 1179-1189. doi: 1.1016/S0005-7967(01)00098-5

Watkins, E. R., Scott, J., Wingrove, J., Rimes, K., Bathurst, N., Steiner, H., Kennell-Web, S., Moulds, M., & Malliaris, Y. (2007). Rumination-focused cognitive behaviour therapy for residual depression: A case series. *Behaviour Research and Therapy, 45*(9), 2144-2154. doi: 1.1016/j.brat.2006.09.018

Watkins, E. R., Taylor, R. S., Byng, R., Baeyens, C., Read, R., Pearson, K., & Watson, L. (2012). Guided self-help concreteness training as an intervention for major depression in primary care: A phase ii randomized controlled trial. *Psychological Medicine, 42*(7), 1359-1371. doi: 1.1017/S0033291711002480

Watkins, E. R., & Teasdale, J. D. (2004). Adaptive and maladaptive self-focus in depression. *Journal of Affective Disorders, 82*(1), 1-8. doi: 1.1016/j.jad.2003.1.006

Watson, D., & Clark, L. A. (1994). *Manual for the positive and negative affect schedule (expanded form)*. Iowa: University of Iowa.

Wegner, D. M., & Vallacher, R. R. (1987). The trouble with action. *Social Cognition,*
    5(3), 179-19. doi: 1.1521/soco.1987.5.3.179

Wickens, C. D., Hutchins, S., Carolan, T., Cumming, J. (2013). Effectiveness of part-task
    training and increasing-difficulty training strategies: a meta-analysis approach. *Human*
    Factors: The Journal of the Human Factors and Ergonomics Society, *55*(2), 461-70.

Wolpe, J. (1969). *The practice of behavior therapy*. New York, NY: Pergamon Press.

Wong, Q. J. J., & Moulds, M. L. (2012). Processing mode during repetitive thinking in
    socially anxious individuals: Evidence for a maladaptive experiential mode. *Journal of*
    Behavior Therapy and Experimental Psychiatry, 43(4), 1064-1073. doi:
    1.1016/j.jbtep.2012.05.002

World Health Organization. (2001). The world health report 2001: Mental health: New
    understanding, new hope. Fact sheet: The world health report (pp. 1-4).

World Health Organization. (2008). *Global burden of disease: 2004 update*. Geneva:
    WHO Press.

Yoon, K. L., & Joormann, J. (2012). Is timing everything? Sequential effects of
    rumination and distraction on interpersonal problem solving. *Cognitive Therapy and*
    Research, 36(3), 165-172. doi: 1.1007/s10608-010-9330-2

Table 1. Descriptive statistics for baseline measures within each training condition.

	Why-Only	How-Only	Combined	Control
	M(*SD*)	*M*(*SD*)	*M*(*SD*)	*M*(*SD*)
Gender	8.4% Female	69.8% Female	74.1% Female	68.6% Female
Age	18.78(1.26)	18.65(.93)	19.07(1.68)	18.82(1.14)
Trait Rumination (RRS)	42.57(14.48)	44.71(16.01)	43.47(14.38)	47.51(15.27)
DASS-Depression	1.07(2.93)	1.37(4.14)	1.01(3.53)	1.81(3.82)
DASS-Anxiety	9.74(2.58)	9.97(3.16)	1.35(3.14)	9.86(2.94)
Working Memory (Running Span)	17.37(4.90)	21.48(6.24)	21.13(6.41)	19.47(6.96)
SAT Average	637.83(62.78)	676.46(68.32)	68.53(54.89)	656.62(62.69)
Baseline SUDS	13.87(15.57)	8.69(1.67)	12.18(16.34)	8.12(1.33)

Note. RRS = Ruminative Responses Scale; DASS=Depression Anxiety Stress Scales; SAT = Scholastic Aptitude Test; SUDS=Subjective Units of Distress Scale.

Table 2. Zero-order correlations between study measures, collapsed across training condition (for full intervention trial, Phase 2b).

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. RRS	.58***	.71***	-.01	.28*	.11†	.04	.01	.00	.11	.14*	-.06	.11†	.33***	.31***	.34***	.36***	.36***	.39***
2. DASS-A	—	.71***	.06	.44***	.08	.05	.01	-.15*	-.04	.11	-.11†	-.01	.25***	.28***	.28***	.30***	.25***	.26***
3. DASS-D		—	-.06	.26	.03	-.03	-.01	-.07	.02	.09	-.11	.08	.34***	.28***	.36***	.41***	.34***	.35***
4. Running Span			—	.23†	.15*	.11†	.15*	.13*	.13†	.11	.16*	.10	-.07	-.03	-.05	-.12†	-.23***	-.21***
5. Time Spent Ruminating (Control only; *N*=51)				—	.24†	.24†	.05	-.07	.07	.24	.08	.16	.30	.07	.21	.23†	.39***	.47***
6. Fluency – Test Period 1					—	.75***	.16**	.19***	.21***	.31***	.23***	.06	-.13†	-.07	-.18***	-.11†	-.14*	-.14*
7. Fluency – Test Period 2						—	.14*	.12	.23***	.26***	.20***	.11	-.12	-.05	-.15*	-.09	-.08	-.07
8. Reading Comprehension – Test Period 1							—	.30***	.08	.07	.19***	.10	.03	.04	-.15*	-.15*	-.16*	-.16*
9. Reading Comprehension – Test Period 2								—	.10	.09	.11†	.13*	-.06	-.01	-.15*	-.15*	-.08	-.07
1. Creative Quality – Test Period 1									—	.40***	.06	.09	-.01	.04	.04	-.03	.05	-.05
11. Creative Quality - Test Period 2										—	-.04	.05	.04	-.07	-.08	-.07	.10	.01
12. MEPS Problem-Solving – Test Period 1											—	.37***	.03	.11	-.09	-.07	-.13*	-.04
13. MEPS Problem-Solving – Test Period 2												—	.07	.12†	-.03	.00	.07	.15*
14. Off-Task Rumination Frequency – Test Period 1													—	.57***	.36***	.32***	.41***	.32***
15. Off-Task Rumination Frequency – Test Period 2														—	.25***	.35***	.30***	.37***
16. Mean SUDS - Test Period 1															—	.81***	.41***	.35***
17. Mean SUDS - Test Period 2																—	.38***	.45***
18. SSSQ State Rumination – Test Period 1																	—	.73***
19. SSSQ State Rumination - Test Period 2																		—

Note. RRS=Ruminative Responses Scale; DASS-A=Depression Anxiety Stress Scales – Anxiety subscale; DASS-D= Depression Anxiety Stress Scales – Depression subscale; MEPS=Means-Ends Problem-Solving task; SUDS=Subjective Units of Distress Scale (0-100); SSSQ = Short State Stress Questionnaire. Test Period 1=baseline, pre-training test battery; Test Period 2=post-training test battery. Off-task Rumination Frequency is collapsed across Probe-caught and Self-caught episodes within each test period, and Mean SUDS is collapsed across Timepoint 1 (post-1st-subtest) and Timepoint 2 (post-2nd-subtest) within each test period (given thee was no significant Condition x Episode Type x Test Period or Condition x Timepoint x Test Period interaction, respectively—see Appendix D, Tables 2 and 3).

Table 3. Overview of primary findings.

	MBIF Abstr	MBIF Concr	GRE Reading Comp	Creative Quality	Creative Fluency	Problem-Solving	Off-task Rumination	SSSQ Rumination	Emotional Vulnerability (SUDS)
Training x Time effect?	No	No (if uncorrected, **Yes: How>Why= Combo)**[1]	No	**Yes:** Combo=Ctrl> Why	No	No	No	**Yes:** Combo=Ctrl< How,Ctrl<Why	No
Moderation effect(s)?	No	No	**Yes** (RRS):	No	No	No	No	**Yes** (DASS-D):	**Yes** (DASS-D, DASS-A, RRS, Working Mem)
High Sxs			*Combo>How= Why,Ctrl> How*					*Combo>How*	*Why< Combo=How=Ctrl*
Low Sxs			*No diff's*					*Ctrl>Combo= Why= How*	*Combo=Ctrl=How < Why*

Note. MBIF = Modified Behavior Identification Scale; Abstr = Abstractness Preference subscale; Concr = Concreteness Preference subscale; GRE Reading Comp = Graduate Record Exam – Reading Comprehension subtest; SSSQ = Short State Stress Questionnaire; SUDS = Subjective Units of Distress Scale; DASS-A = Depression Anxiety Stress Scales – Anxiety subscale; DASS-D = Depression Anxiety Stress Scales – Depression subscale. Sxs = symptoms. Combo = Combined condition; Ctrl = "Free-thinking" Control condition; How = How-only condition; Why = Why-only condition. Only significant 2-way (Training Condition x Time) and 3-way (Training Condition x Time x Moderator) effects are shown (at the $p<.05$ level); see Tables 5-9 and Appendix F for full results, including marginally significant effects.

[1]As noted in the text, we did not find training effects on concrete construal preference when correcting for pairwise comparisons; however, when examining the individual, uncorrected pairwise contrasts (thus more closely matching the dichotomous abstract/concrete construal manipulations that are typically compared in the traditional construal literature), we observed the expected effects of the HowOnly (versus WhyOnly and Combined) training condition on the MBIF Concrete subscale, as shown.

Table 4. Planned contrasts to test hypothesized effects of Training Condition for MBIF Abstract and Concrete subscales.

Abstractness preference (MBIF-Abstract)	*Unstandardized b*	*Standardized b*	*SE*	*z*	*p*
Condition (Why-Only vs. Combined) x Time	.09	.12	.14	.85	.735
Condition (How-Only vs. Combined) x Time	.14	.19	.14	1.33	.411
Condition (Control vs. Combined) x Time	.10	.14	.15	.97	.651
Condition (How-Only vs. Why-Only) x Time	-.12	.07	.14	.48	.934
Condition (Control vs. Why-Only) x Time	.07	.02	.15	.14	.998
Condition (How-Only vs. Control) x Time	.03	.05	.14	.34	.975
Concreteness preference (MBIF-Concrete)	*Unstandardized b*	*Standardized b*	*SE*	*z*	*p*
Condition (Why-Only vs. Combined) x Time	.03	.04	.15	.25	.990
Condition (How-Only vs. Combined) x Time[†]	.24	.34	.15	2.27	.068
Condition (Control vs. Combined) x Time	.10	.14	.15	.92	.691
Condition (How-Only vs. Why-Only) x Time	.21	.30	.15	2.04	.115
Condition (Control vs. Why-Only) x Time	.07	.10	.15	.68	.844
Condition (How-Only vs. Control) x Time	.14	.20	.15	1.33	.414

Note. MBIF = Modified Behavior Identification Scale; Abstr = Abstractness Preference subscale; Concr = Concreteness Preference subscale.

Table 5. Planned contrasts to test hypothesized effects of Condition x Test Period for each Cognitive Performance outcome (with only significant moderation effects displayed; see Appendix E for all moderation statistics).

Reading Comprehension	*Unstandardized b*	*Standardized b*	*SE*	*z*	*p*
Condition (Why-Only vs. Combined) x Time	-.22	-.14	.23	-.60	.880
Condition (How-Only vs. Combined) x Time[†]	-.79	-.49	.23	-2.14	.091
Condition (Control vs. Combined) x Time	-.17	-.10	.23	-.45	.944
Condition (How-Only vs. Why-Only) x Time	-.57	-.35	.22	-1.57	.288
Condition (Control vs. Why-Only) x Time	.05	.03	.23	.15	.998
Condition (How-Only vs. Control) x Time	-.38	-.38	.23	-1.69	.230
Trait Rumination (RRS) as moderator					
Condition (Why-Only vs. Combined) x Time x RRS[†]	-.90	-.55	.24	-2.32	.059
Condition (How-Only vs. Combined) x Time x RRS[*]	**-.97**	**-.60**	**.23**	**-2.65**	**.025**
Condition (Control vs. Combined) x Time x RRS[†]	-.80	-.50	.24	-2.10	.096
Condition (How-Only vs. Why-Only) x Time x RRS	-.07	-.05	.22	-.21	.994
Condition (Control vs. Why-Only) x Time x RRS	.09	.06	.23	.25	.989
Condition (How-Only vs. Control) x Time x RRS	-.17	-.10	.22	-.48	.936
Creativity – Divergent Thinking	*Unstandardized b*	*Standardized b*	*SE*	*z*	*p*
Condition (Why-Only vs. Combined) x Time[***]	**-.20**	**-.78**	**.20**	**-3.81**	**< .001**
Condition (How-Only vs. Combined) x Time	-.10	-.40	.21	-1.97	.132
Condition (Control vs. Combined) x Time	-.01	-.05	.21	-.23	.991
Condition (How-Only vs. Why-Only) x Time	.09	.37	.20	1.84	.175
Condition (Control vs. Why-Only) x Time[**]	**.19**	**.73**	**.21**	**3.54**	**.002**
Condition (How-Only vs. Control) x Time	-.09	-.36	.21	-1.72	.216
Creative Fluency	*Unstandardized b*	*Standardized b*	*SE*	*z*	*p*
Condition (Why-Only vs. Combined) x Time	-.42	-.05	.14	-.38	.967
Condition (How-Only vs. Combined) x Time	.19	.02	.14	.17	.997

	Unstandardized b	Standardized b	SE	z	p
Condition (Control vs. Combined) x Time	-1.91	-.24	.14	-1.68	.234
Condition (How-Only vs. Why-Only) x Time	.61	.08	.14	.56	.906
Condition (Control vs. Why-Only) x Time	-1.50	-.19	.14	-1.33	.417
Condition (How-Only vs. Control) x Time	2.10	.27	.14	1.87	.160
MEPS Problem-Solving	*Unstandardized b*	*Standardized b*	*SE*	*z*	*p*
Condition (Why-Only vs. Combined) x Time	-.01	.00	.21	-.02	.999
Condition (How-Only vs. Combined) x Time	.08	.03	.21	.15	.997
Condition (Control vs. Combined) x Time	-.70	-.29	.22	-1.34	.410
Condition (How-Only vs. Why-Only) x Time	.01	.04	.21	.18	.996
Condition (Control vs. Why-Only) x Time	.09	-.28	.21	-1.33	.417
Condition (How-Only vs. Control) x Time	.77	.32	.21	1.50	.316

†p<.10; *p<.05; **p<.01; ***p<.001

Note. MEPS = Means-Ends Problem Solving test; RRS = Ruminative Responses Scale. All continuous predictors and moderators were standardized for these analyses. Unstandardized betas represent the slope of change in units of the outcome variable (e.g., Reading Comprehension Accuracy) for every 1 SD increase in the predictor. Standard errors (SEs) correspond to *standardized* beta estimates.

Table 6. Planned contrasts to test hypothesized effects of Condition x Test Period for Off-Task Rumination Frequency (via thought tracking), collapsed across Episode Type (i.e., probe-caught and self-caught).

Off-Task Rumination Frequency (via thought-tracking)	*Unstandardized b*	*Standardized b*	*SE*	*z*	*p*
Condition (Why-Only vs. Combined) x Time	.08	.12	.15	.79	.774
Condition (How-Only vs. Combined) x Time	-.02	-.06	.15	-.40	.960
Condition (Control vs. Combined) x Time	-.11	-.21	.15	-1.38	.383
Condition (How-Only vs. Why-Only) x Time	-.11	-.18	.15	-1.21	.500
Condition (Control vs. Why-Only) x Time[†]	-.19	-.33	.15	-2.20	.067
Condition (How-Only vs. Control) x Time	.08	.15	.15	.99	.635

[†]$p < .10$

Note. All continuous predictors and moderators were standardized for these analyses. Unstandardized betas represent the slope of change in off-task rumination episodes for every 1 SD increase in the predictor. Standard errors (SEs) correspond to *standardized* beta estimates.

Table 7. Planned contrasts to test hypothesized effects of Condition x Test Period for Retrospective State Rumination.

Retrospectively reported State Rumination (via SSSQ Rumination subscale)	Unstandardized b	Standardized b	SE	z	p
Condition (Why-Only vs. Combined) x Time	.20	.20	.15	1.35	.403
Condition (How-Only vs. Combined) x Time	.30	.30	.15	2.01	.123
Condition (Control vs. Combined) x Time	-.11	-.11	.15	-.73	.812
Condition (How-Only vs. Why-Only) x Time	.10	.10	.15	.67	.851
Condition (Control vs. Why-Only) x Time	-.31	-.31	.15	-2.08	.105
Condition (How-Only vs. Control) x Time[*]	**.41**	**.41**	**.15**	**2.74**	**.020**
Depression symptoms (DASS-D) as moderator					
Condition (Why-Only vs. Combined) x Time x DASS-D	.16	.16	.16	.99	.638
Condition (How-Only vs. Combined) x Time x DASS-D	.17	.17	.15	1.12	.547
Condition (Control vs. Combined) x Time x DASS-D[*]	**.41**	**.41**	**.16**	**2.65**	**.026**
Condition (How-Only vs. Why-Only) x Time x DASS-D	.01	.01	.15	.04	.999
Condition (Control vs. Why-Only) x Time x DASS-D	.25	.25	.16	1.57	.273
Condition (How-Only vs. Control) x Time x DASS-D	-.25	-.25	.15	-1.68	.234

[*]$p<.05$
Note. SSSQ = Short State Stress Questionnaire; DASS-D = Depression Anxiety Stress Scales - Depression subscale. All continuous predictors and moderators were standardized for these analyses. Unstandardized betas represent the slope of increase in SSSQ Rumination subscale score for every 1 SD increase in the predictor. Standard errors (SEs) correspond to *standardized* beta estimates.

Table 8. Planned contrasts to test hypothesized effects of Condition x Test Period for Negative Affect during Testing (SUDS-Distress), collapsed across Timepoint (i.e., post-subtest-1 and post-subtest-2).

Negative Affect during Testing (SUDS-Distress)	Unstandardized b	Standardized b	SE	z	p
Condition (Why-Only vs. Combined) x Time	1.81	.10	.09	1.10	.562
Condition (How-Only vs. Combined) x Time	2.72	.15	.09	1.66	.242
Condition (Control vs. Combined) x Time	-.73	-.04	.09	-.44	.947
Condition (How-Only vs. Why-Only) x Time	.91	.05	.09	.56	.903
Condition (Control vs. Why-Only) x Time	-2.55	-.14	.09	-1.56	.292
Condition (How-Only vs. Control) x Time[†]	3.45	.19	.09	2.12	.095
Trait rumination (RRS) as moderator					
Condition (Why-Only vs. Combined) x Time x RRS	-2.89	-.16	.10	-1.65	.237
Condition (How-Only vs. Combined) x Time x RRS	.98	.05	.09	.59	.879
Condition (Control vs. Combined) x Time x RRS	.62	.03	.09	.36	.968
Condition (How-Only vs. Why-Only) x Time x RRS[*]	**3.87**	**.21**	**.09**	**2.41**	**.048**
Condition (Control vs. Why-Only) x Time x RRS	3.51	.19	.09	2.10	.100
Condition (How-Only vs. Control) x Time x RRS[†]	.37	.02	.09	.23	.992
Depression symptoms (DASS-D) as moderator					
Condition (Why-Only vs. Combined) x Time x DASS-D[**]	**-6.25**	**-.34**	**.10**	**-3.32**	**.003**
Condition (How-Only vs. Combined) x Time x DASS-D	.69	.04	.09	.41	.955
Condition (Control vs. Combined) x Time x DASS-D	.41	.02	.09	.24	.990
Condition (How-Only vs. Why-Only) x Time x DASS-D[***]	**6.93**	**.38**	**.10**	**3.95**	**<.001**
Condition (Control vs. Why-Only) x Time x DASS-D[***]	**6.66**	**.36**	**.10**	**3.71**	**<.001**
Condition (How-Only vs. Control) x Time x DASS-D	.27	.01	.09	.17	.997
Anxiety symptoms (DASS-A) as moderator					
Condition (Why-Only vs. Combined) x Time x DASS-A	-2.17	-.17	.10	-1.59	.265
Condition (How-Only vs. Combined) x Time x DASS-A	2.06	.16	.09	1.68	.226

Condition (Control vs. Combined) x Time x DASS-A	.60	.05	.10	.47	.933
Condition (How-Only vs. Why-Only) x Time x DASS-A**	**4.23**	**.32**	**.10**	**3.37**	**.003**
Condition (Control vs. Why-Only) x Time x DASS-A[†]	2.77	.21	.10	2.13	.089
Condition (How-Only vs. Control) x Time x DASS-A	1.46	.11	.09	1.26	.467
Working memory (RunningSpan) as moderator					
Condition (Why-Only vs. Combined) x Time x RunningSpan	-2.31	-.18	.10	-1.71	.227
Condition (How-Only vs. Combined) x Time x RunningSpan	.89	.07	.09	.77	.797
Condition (Control vs. Combined) x Time x RunningSpan	-1.11	-.09	.09	-.99	.648
Condition (How-Only vs. Why-Only) x Time x RunningSpan*	**3.20**	**.25**	**.10**	**2.36**	**.048**
Condition (Control vs. Why-Only) x Time x RunningSpan	1.20	.09	.10	.91	.652
Condition (How-Only vs. Control) x Time x RunningSpan	2.01	.15	.09	1.78	.205

[†]$p<.10$; *$p<.05$; **$p<.01$; ***$p<.001$

Note. SUDS = Subjective Units of Distress Scale; RRS = Ruminative Responses Scale; DASS-D = Depression Anxiety Stress Scales - Depression subscale; DASS-A = Depression Anxiety Stress Scales - Anxiety subscale. All continuous predictors and moderators were standardized for these analyses. Unstandardized betas represent the slope of change in SUDS-Distress (on a 0-100 scale) for every 1 SD increase in the predictor. Standard errors (SEs) correspond to *standardized* beta estimates.

Figure 1. Simple slopes of change in negative affect (SUDS) from baseline to post-failure and post-rumination induction, plotted for low (-1 SD) and high (+1 SD) levels of trait rumination (RRS).
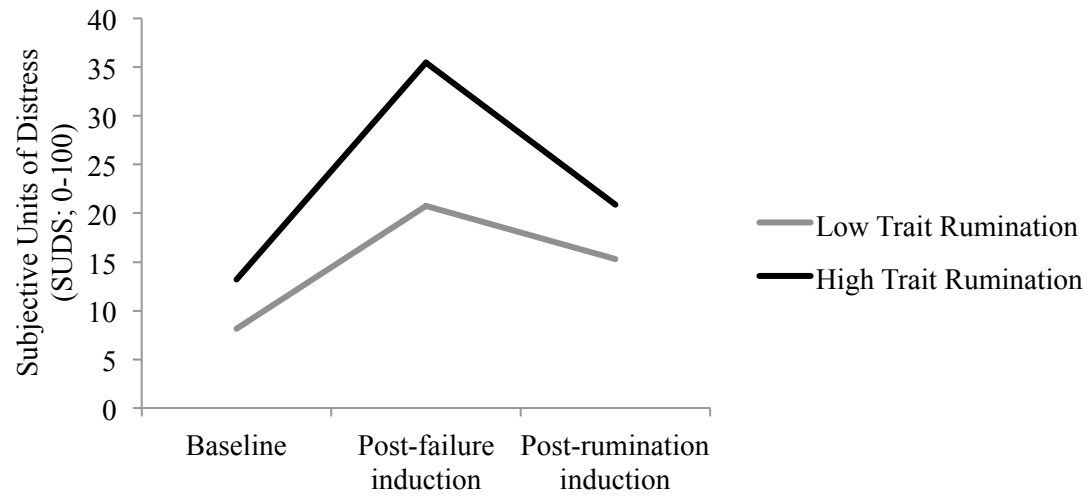
Note. Low and High Trait Rumination reflect model-predicted values at -1 and +1SD from the mean on the Ruminative Responses Scale (RRS).

Figure 2. Simple slopes for the Test period x Condition x Trait Rumination interaction predicting Reading Comprehension.
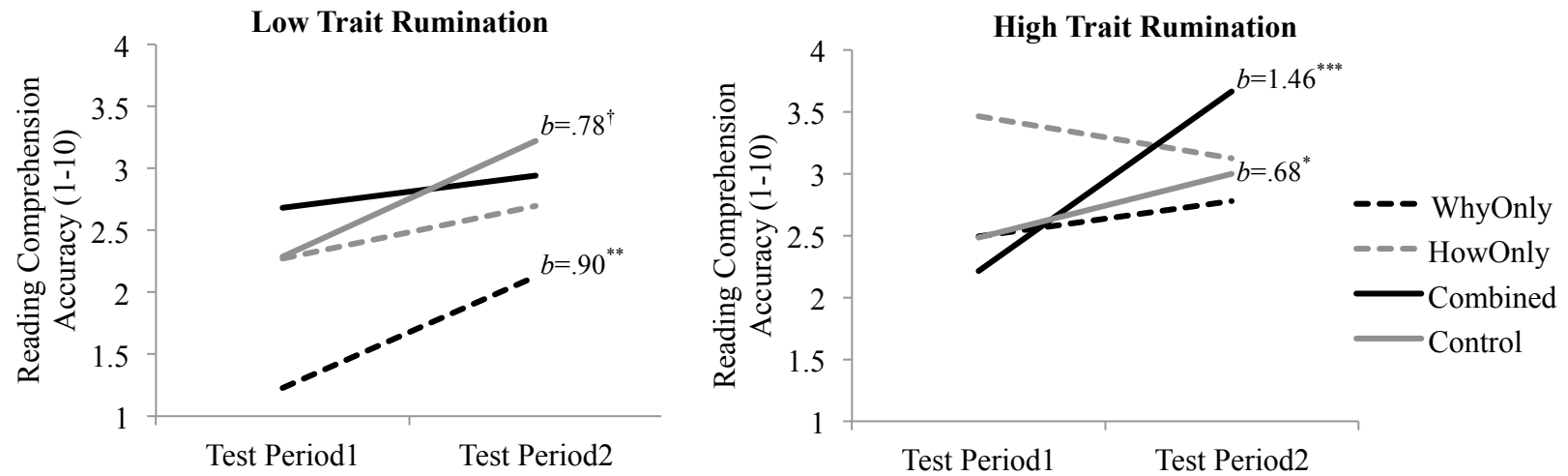
†p<.10; *p<.05; **p<.01; ***p<.001

Note. Low and High Trait Rumination reflect model-predicted values at -1 and +1SD from the mean of the Ruminative Responses Scale. Unstandardized betas (*b*) are shown only for significant or marginally significant simple slopes of Test Period, and reflect the estimated change in Reading Comprehension Accuracy (# of questions answered correctly out of 10) from pre- to post-training.

Figure 3. Simple slopes for the Test Period x Condition interaction predicting Creative Quality.
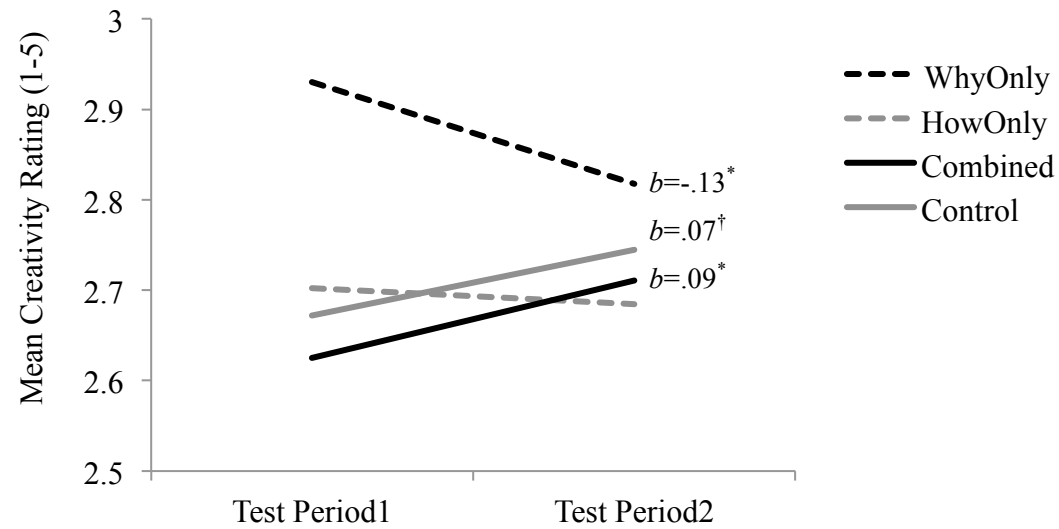
†p<.10; *p<.05
Note. Unstandardized betas (*b*) are shown only for significant or marginally significant simple slopes of Test Period, and reflect the estimated change in Mean Creativity Rating (on a 1-5 scale, where 1=Not at all creative and 5=Very creative) from pre- to post-training.

Figure 4a. Simple slopes for the Test Period x Condition interaction predicting Retrospective State Rumination (SSSQ Rumination subscale).
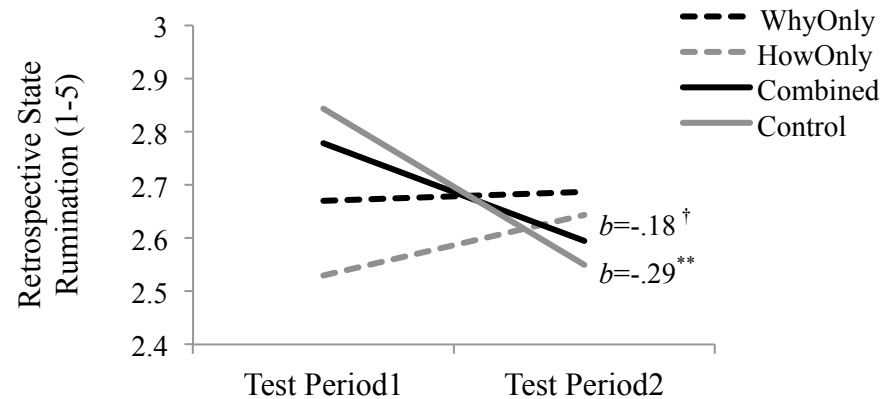
Figure 4b. Simple slopes for the Test Period x Condition x Depression Symptoms interaction predicting Retrospective State Rumination. (SSSQ Rumination subscale).
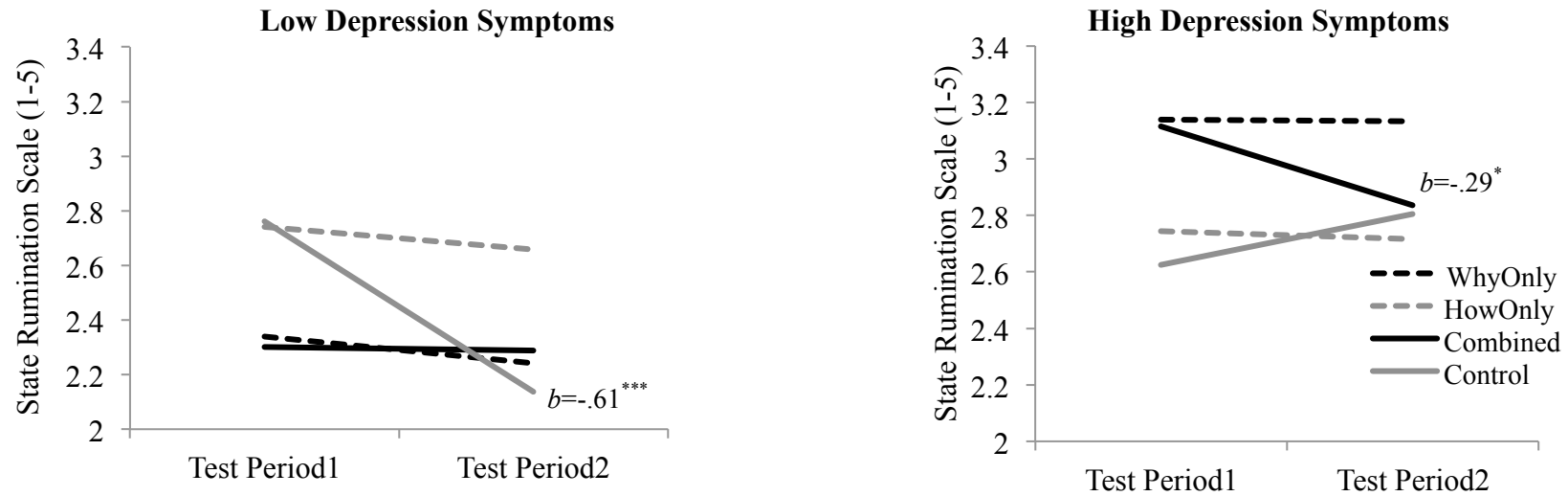
$^†p<.10$; $^*p<.05$; $^{**}p<.01$; $^{***}p<.001$

Note. Low and High Depression Symptoms reflect model-predicted values at -1 and +1SD from the mean of the Depression Anxiety Stress Scales – Depression (DASS-D) subscale. Unstandardized betas (b) are shown only for significant or marginally significant simple slopes of Test Period, and reflect the estimated change in State Rumination (on a 1-5 scale, where 1=Not at all and 5=Extremely) from pre- to post-training.

Figure 5a. Simple slopes for the Test Period x Condition x Depression Symptoms interaction predicting Emotional Vulnerability (i.e., mean SUDS-Distress rating during testing).
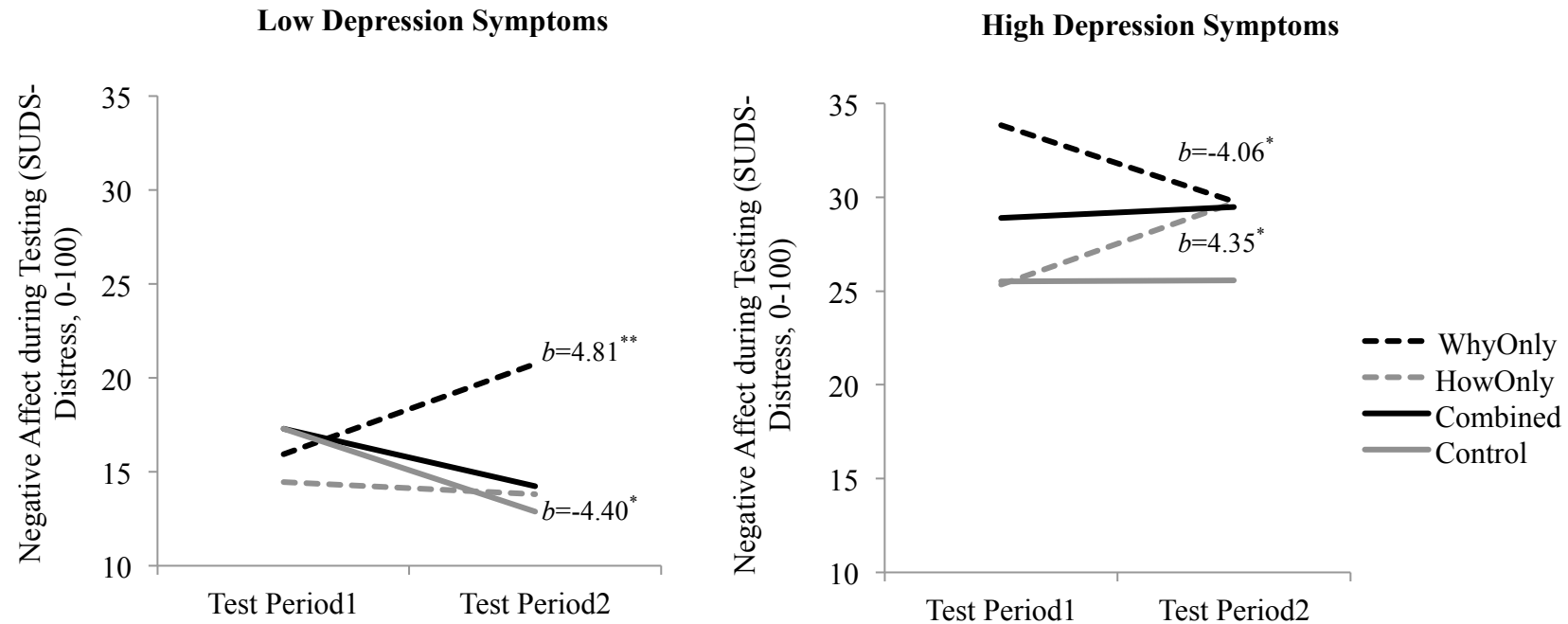
*p<.05; **p<.01

Note. Low and High Depression Symptoms reflect model-predicted values at -1 and +1SD from the mean of the Depression Anxiety Stress Scales – Depression (DASS-D) subscale. Unstandardized betas (*b*) are shown only for significant or marginally significant simple slopes of Test Period, and reflect the estimated change in mean SUDS-Distress rating (on a 0-100 scale, where 0=No distress and 100=Extreme distress) from pre- to post-training. The simple slopes for Low versus High Anxiety Symptoms (DASS-Anxiety) and Low versus High Trait Rumination (Ruminative Responses Scale) follow a similar pattern to the one displayed above.

Figure 5b. Simple slopes for the Test Period x Condition x Working Memory interaction predicting Negative Affect during Testing (SUDS-Distress).
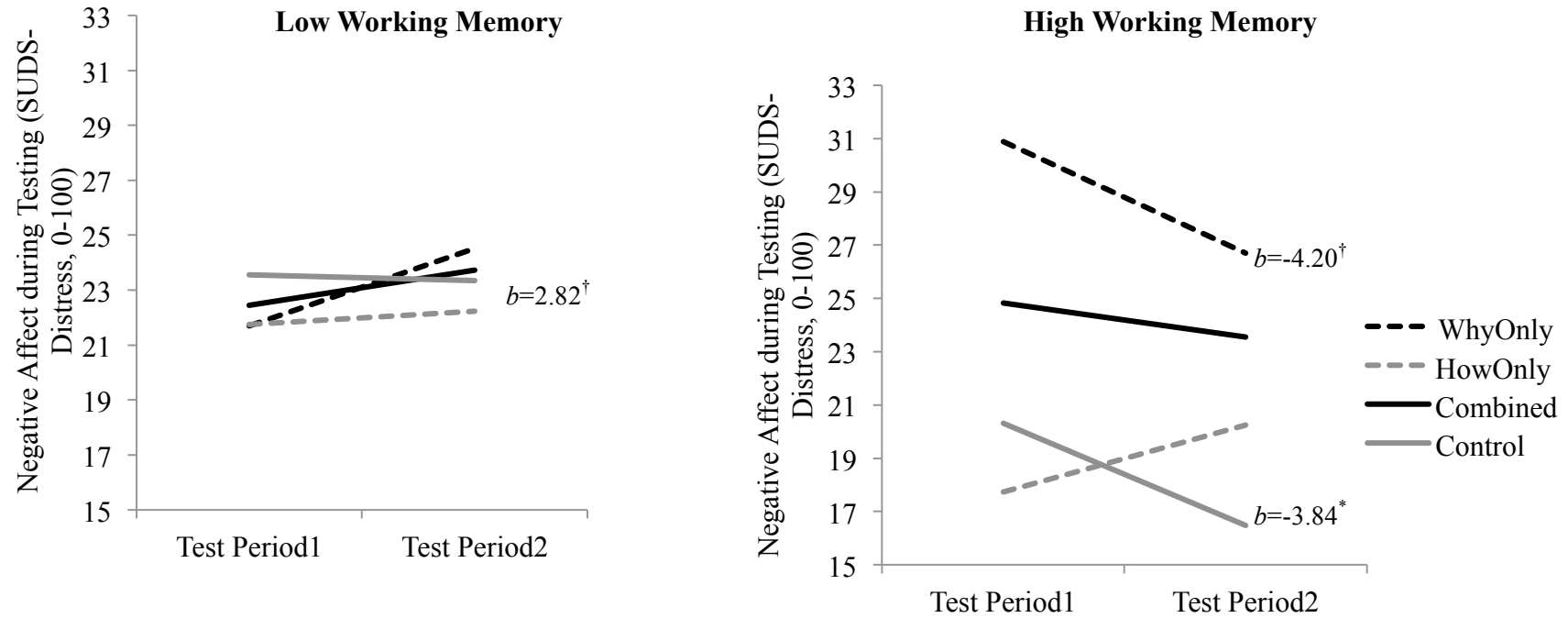
†p<.10; *p<.05
Note. Low and High Working Memory reflect model-predicted values at -1 and +1SD from the mean of participants' Running Span score. Unstandardized betas (b) are shown only for significant or marginally significant simple slopes of Time (i.e., significant increases or decreases in SUDS from Test Period1 to Test Period2).

Appendix A
Scripts and Flowcharts of Each Training Condition

"Why" prompt: Please list ONE REASON why improving your reading comprehension, creative ability, and/or problem-solving performance on this next round of tasks (and others like them) might be of personal value for you.
- Follow-up prompt: "For the reason you specified above [participant's previous response displayed], please list one further reason why THIS is of personal value for you."

"How" prompt: Please list ONE STRATEGY you could use to improve your reading comprehension… on this next round of tasks (and others like them).
- Follow-up prompt: "Please list one MORE SPECIFIC STRATEGY you could use to accomplish the above [participant's previous response displayed]. (If this is as specific as it gets, then please come up with another, equally specific strategy)."

"Free-thinking" Control instructions: To help you process your test results and prepare for the next round of tests, I'd like you to just sit quietly and collect your thoughts for a few minutes. I will let you know when it's time to go on to the second test battery. Go ahead.

Why-only condition   **How-only condition**
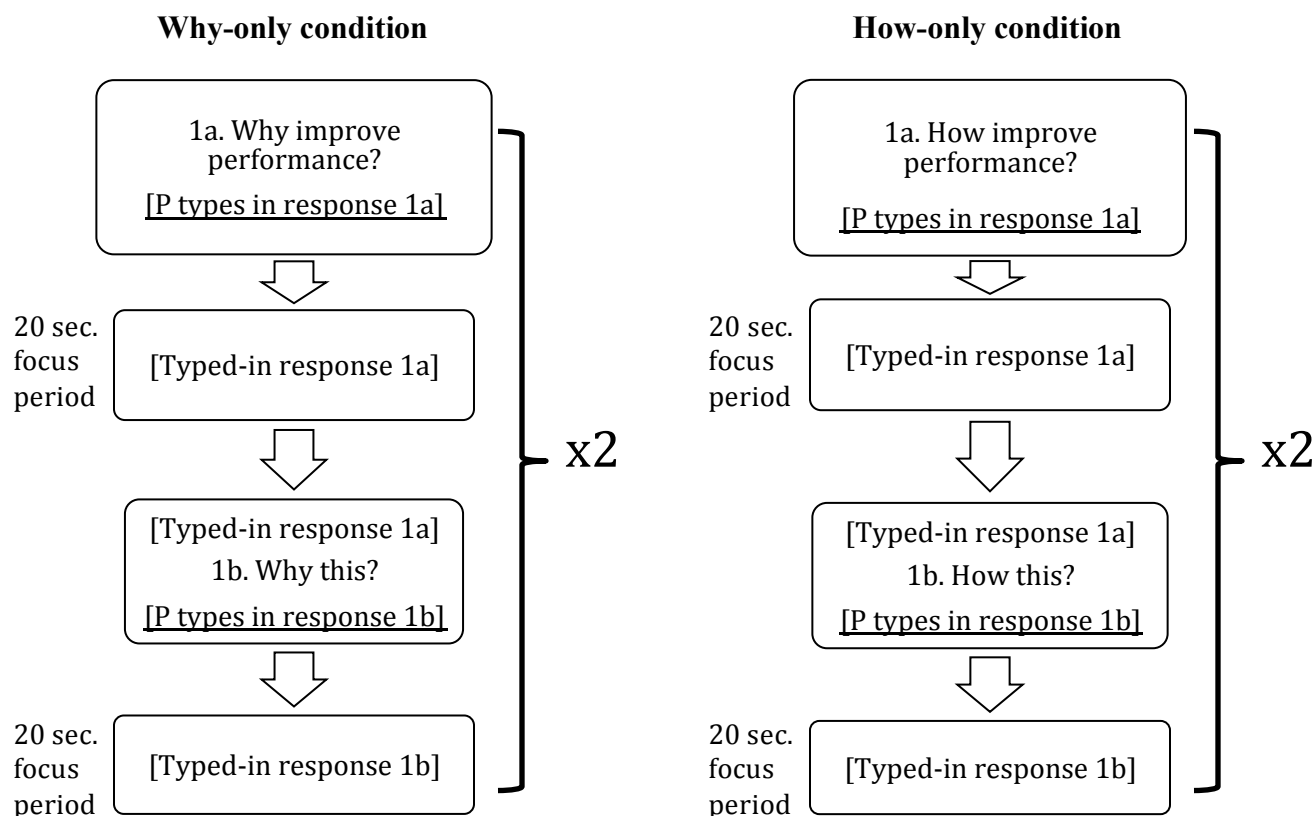
Combined condition (order of "Why" & "How" counterbalanced)

1a. Why improve performance? [P types in response 1a]	1a. How improve performance? [P types in response 1a]
20 sec. focus period — [Typed-in response 1a]	20 sec. focus period — [Typed-in response 1a]
[Typed-in response 1a] 1b. Why this? [P types in response 1b]	[Typed-in response 1a] 1b. How this? [P types in response 1b]
20 sec. focus period — [Typed-in response 1b]	20 sec. focus period — [Typed-in response 1b]

x1   x1

Appendix B
Modified Behavior Identification Scale (MBIF) Validation Phase

In order to examine the effects of training on changes in abstract and concrete construal

preference, we first developed and validated a revised, abbreviated Likert-scale version of an

existing forced-choice "abstract" versus "concrete" construal measure (the Behavior

Identification Form [BIF]; Vallacher & Wegner, 1989). Unlike the original BIF, which yielded a

single "abstractness" composite score (on the assumption that "abstract" and "concrete"

construal are opposite endpoints along a single dimension), the revised measure allows for

separate, dissociable "abstract" and "concrete" construal level scores to be computed—thus

potentially providing a more appropriate and informative check for the novel "combined"

(why+how) construal training condition in our intervention study.

Methods

Participants and Recruitment

For the first, preliminary validation phase, participants ($N = 137$) were recruited through the

Mechanical Turk (mTurk) website. All mTurk visitors who were 18 or older and fluent in

English were eligible to participate; to avoid participant burden and minimize dropout, no further

demographic questions were asked during this phase. For the second, confirmatory validation

phase, undergraduate students (N=426; mean age = 19.00 years, SD=1.15; 63.3% female) were

recruited through the psychology department participant pool, thus providing a sample that more

closely resembled the undergraduate student sample we would be recruiting for the intervention

trial.

Materials

The original <u>Behavior Identification Form</u> (BIF; Vallacher & Wegner, 1989) is a well-

validated measure assessing the tendency to construe everyday actions at relatively higher or

lower levels of abstraction. The measure lists 25 common behaviors (e.g., "Making a list"), each of which is followed by two descriptions—a relatively abstract, higher-level construal (e.g., "Getting organized") and a relatively concrete, lower-level construal (e.g., "Writing things down"). For each item, participants select the description that sounds preferable to them. A total score is computed by summing the total number of higher-level descriptions selected by the respondent; higher scores indicate a tendency toward more high-level (i.e., abstract) construal.

The Modified Behavior Identification Form (MBIF) is a newly revised version of the BIF, which was developed and validated during this phase of the project. Specifically, the measure's format was modified so that participants would separately rate both the "abstract" and the "concrete" description of each behavior on a 5-point Likert scale (from 1 = "Not at all descriptive of the behavior" to 5 = "Extremely descriptive of the behavior"). Thus, two separate sum scores are derived: an "abstract construal preference" and a "concrete construal preference" score. This novel format allowed us to test the theoretical assumption that "abstract" and "concrete" construal are not merely opposite ends of a single continuum, but rather may represent potentially independent dimensions (such that one could be equally high on both abstract and concrete processing, for example). This pilot phase was conducted so that a validated, abbreviated version of this measure could be administered pre- and post-construal training in the main intervention trial, thus providing a more conceptually precise construal manipulation check for the "combined" condition (which was expected to induce higher levels of both abstract and concrete construal).

The Construal Check Form (Fujita et al., 2006) is an 8-item construal measure identical in format to the BIF, but consisting of a different set of everyday behaviors (e.g., "Sweeping the floor") and corresponding descriptions (e.g., Abstract—"Being clean"; Concrete—"Moving a broom"). Together with the original BIF, this additional construal measure was administered to

help establish the convergent validity of the MBIF.

The Ruminative Responses Scale (RRS; Nolen-Hoeksema & Morrow, 1991) is a 22-item self-report scale that measures the tendency to respond to depressed or negative mood by ruminating (see description in main text). It was administered in this preliminary phase to help establish the discriminant validity of the MBIF, as well as to examine the previously untested assumption that trait rumination should correlate with the use of more abstract and/or less concrete construal mode (e.g., Watkins et al., 2011).

The Rumination-Reflection Questionnaire – Reflection subscale (RRQ-Ref; Trapnell & Campbell, 1999) is a 12-item self-report scale that measures the tendency to engage in introspective thought and self-reflection, which is thought to be relatively adaptive and shows only modest correlations with depressive rumination (e.g., Thomsen, Tonnesvang, Schnieber, & Olesen, 2011). It was administered to help establish the discriminant validity of the BIF, as well as to distinguish between the potentially dissociable roles of maladaptive rumination versus more adaptive self-reflection in predicting preferred construal level.

The Depression Anxiety Stress Scales (Lovibond & Lovibond, 1995) is a self-report measure consisting of three seven-item subscales, which assess symptoms of depression, anxiety, and stress, respectively (see descriptions of DASS-Depression and DASS-Anxiety in main text). These subscales were administered to help establish the discriminant validity of the MBIF.

The Positive and Negative Affect Schedule (PANAS; Watson & Clark, 1994) is a widely used self-report measure of positive and negative affect. Participants rate various adjectives (10 positive and 10 negative) for how closely they match their general emotional state. The PANAS has shown good reliability and validity (Watson & Clark) and was administered to help establish the discriminant validity of the BIF.

Note, given prior findings that relatively more abstract construal levels are modestly

correlated (in the range of r=.24-.28) with higher subjective well-being and positive affect (Freitas, Clark, Kim, & Levy, 2009; Updegraff & Suh, 2007), some correlation was expected between the MBIF indices and the mood-related DASS and PANAS measures. Discriminant validity was established, however, based on the relatively smaller magnitude of these correlations compared to those between the MBIF and previously published measures of construal (i.e., the original BIF and Construal Check Form).

Procedure

For the first, preliminary validation phase, participants who clicked on the mTurk study link were redirected to the informed consent page of the validation survey, administered via Qualtrics software. After completing informed consent, participants were administered the BIF, MBIF, and Construal Check Form in randomized order, followed by the battery of rumination, reflection, and discriminant validity measures (RRS, RRQ-Reflection, DASS-Depression and Anxiety, and PANAS), also in randomized order. The full study took approximately 10 minutes to complete. All participants were then debriefed and assigned an anonymous study code, which they entered in mTurk to be reimbursed $1 for participating.

For the second, confirmatory validation phase, students provided responses to the modified, abbreviated MBIF, as well as the RRS, DASS-Depression and Anxiety, and PANAS, as part of a department-wide online preselection survey. The full survey took approximately 30 minutes to complete, and students received course credit for participating.

<div align="center">

Results

</div>

Data Preparation and Descriptive Statistics

All previously published measures were scored following the original published scoring procedures. DASS-Depression scores were log-transformed in order to reduce positive skew. One extreme outlier, defined as a value deviating by more than three times the interquartile range

from the lower or upper quartile of a variable's distribution, was removed from the PANAS-Negative Affect subscale. After following these data-cleaning procedures, all continuous measures were normally distributed.

Exploratory Factor Analysis and Preliminary Scale Validation

To explore the MBIF's factor structure and determine which of the 50 total items (consisting of 25 pairs of "abstract" and "concrete" ratings) to retain in the final, abbreviated measure, an exploratory factor analysis (EFA) was conducted using the "psych" software package in R. Based on an initial scree test conducted on the full 50-item correlation matrix (using the "eigenvalue > 1" rule as the primary criterion; Kaiser, 1960), an EFA model with a two-factor solution was run using maximum likelihood estimation and a "promax" rotation (which allows for non-zero correlations between factors). Items with factor loadings below .5 on the appropriate factor (or with high loadings on both factors) were removed, resulting in 20 rating items—one "abstract" and one "concrete" description rating for each of 10 listed behaviors—that were retained for the abbreviated measure. Factor loadings, item-total correlations, and descriptive statistics for each item in the newly abbreviated measure are displayed in Table B1 (below). Each of the two factors ("abstract" and "concrete") accounted for 23% of the total variance, such that the two factors together explained 46%. A further scree test conducted on the abbreviated set of items corroborated that this two-factor solution was the best fit for the data. Thus, the "abstract" and "concrete" indices of the MBIF were examined separately in subsequent analyses.

To test the internal consistency of the revised and modified BIF, Cronbach's alphas were computed for each subscale. The standardized alphas for both the "abstract" and "concrete" subscale were .89, indicating excellent reliability (well above the $\alpha \geq .70$ criterion recommended by Kline, 1999).

As a preliminary investigation of the measure's convergent validity, correlations between the MBIF subscales and the original BIF and Construal Check Form were examined. Given that higher scores on the latter measures reflect a relatively greater tendency toward abstract construal, the "abstract" subscale of the MBIF was expected to correlate positively with these measures, whereas the "concrete" subscale was expected to correlate negatively. Indeed, moderate to strong correlations were found in the expected directions (see Table B2).

Further, to examine the discriminant validity of the MBIF index score(s), their correlations with the RRQ-Rumination, RRQ-Reflection, DASS-Anxiety and Depression, and PANAS Positive and Negative Affect subscales were computed. As expected, these correlations were relatively lower in magnitude than their correlations with the two previously published measures of construal mode (see Table B2 below). Of note, the abstract subscale showed small positive correlations with positive affect and reflection while the concrete subscale showed a small negative correlation with rumination, each of which was in line with past research and theory. In sum, the abbreviated MBIF subscale derived from the EFA appeared to have acceptable discriminant and convergent validity based on the initial validation sample.

Confirmatory Factor Analysis

To cross-validate the MBIF with a separate, undergraduate student sample, a confirmatory factor analysis (CFA) was conducted (using the "sem" package in R) on just the 10 items retained from the earlier EFA, to test the hypothesis that the 2-factor solution with "abstract" and "concrete" descriptions loading onto two distinct factors is a good fit for the data. The model indeed showed adequate fit, with RMSEA = .07, SRMR = .07, and CFI = .89.

Table B1. Factor loadings, item-total correlations, and descriptive statistics for Modified Behavior Identification Form items retained based on exploratory factor analysis (EFA).

Behavior	"Concrete" factor item	Loading	Item-Total Corr	*M*	*SD*	"Abstract" factor item	Loading	Item-Total Corr	*M*	*SD*
Picking an apple	Pulling an apple off a branch	.71	.71	3.70	1.10	Getting something to eat	.62	.68	3.70	1.01
Measuring a room for carpeting	Using a yard stick	.70	.71	3.90	.90	Getting ready to remodel	.74	.75	3.60	1.05
Cleaning the house	Vacuuming the floor	.70	.73	3.60	1.10	Showing one's cleanliness	.65	.68	3.40	1.11
Painting a room	Applying brush strokes	.58	.68	3.60	1.00	Making the room look fresh	.69	.74	3.50	1.05
Caring for houseplants	Watering plants	.74	.74	3.50	1.10	Making the room look nice	.68	.72	3.80	1.04
Locking a door	Putting a key in the lock	.69	.74	3.50	1.10	Securing the house	.70	.74	3.70	1.07
Climbing a tree	Holding on to branches	.70	.74	3.50	1.00	Getting a good view	.77	.74	3.30	1.10
Filling out a personality test	Answering questions	.57	.66	3.20	1.10	Revealing what you're like	.66	.72	3.60	1.07
Resisting temptation	Saying "no"	.65	.71	3.60	1.10	Showing moral courage	.55	.60	3.50	1.12
Pushing a doorbell	Moving a finger	.68	.70	3.70	1.00	Seeing if someone's home	.63	.69	3.70	.98

Note. M=Mean; *SD*=Standard Deviation; Item-Total Corr = correlation between individual item and overall factor.

Table B2. Correlations between each MBIF subscale, established construal measures, and discriminant validity measures in preliminary validation sample.

	2	3	4	5	6	7	8	9	10
1. MBIF Abstract Scale	**-.09**	**.55*****	**.57*****	**.18***	**-.14**	**-.07**	**-.02**	**-.15†**	**.21***
2. MBIF Concrete Scale	____	**-.45*****	**-.64*****	**-.09**	**.00**	**-.06**	**-.18***	**-.04**	**-.01**
3. Construal Check Form		____	.82***	.25***	-.02	-.02	.09	-.07	.17*
4. BIF Abstractness Scale			____	.18*	-.14	-.07	.05	-.07	.18*
5. RRQ-Reflection				____	.09	.11	.30***	.16†	.23***
6. DASS-Depression					____	.69***	.65***	.57***	-.35***
7. DASS-Anxiety						____	.66***	.64***	-.12
8. Ruminative Responses Scale (RRS)							____	.63***	-.15†
9. PANAS Negative Affect								____	-.29***
1. PANAS Positive Affect									____

†p<.10; *p<.05; **p<.01; ***p<.001

Note. MBIF=Modified Behavior Identification Form; BIF=Behavior Identification Form (original dichotomous scale, with higher values reflecting greater abstract construal); RRQ=Rumination Reflection Questionnaire; DASS=Depression Anxiety Stress Scales; PANAS = Positive and Negative Affect Schedule (Trait version).

Appendix C
Pilot Trial: Methods, Results, and Modifications

PARTICIPANTS AND RECRUITMENT.

Participants (*N*=35; mean age=19.68, *SD*=2.84; 68.6% female) were undergraduate students recruited either through flyers posted on campus or through the University's psychology department participant pool. As for the full intervention trial, participants were offered course credit or $20 compensation for the study. Eligibility was determined by the same set of prescreening measures (collected either via a brief online survey or as part of the department-wide preselection procedures): the 22-item Ruminative Responses Scale (RRS; Nolen-Hoeksema & Morrow, 1991) and the modified Sources of Validation Scale (SVS; Harber, 1995). As a preliminary test of the intervention's feasibility and acceptability in a vulnerable, high-ruminative sample, we invited only high-ruminative respondents to participate in this initial pilot phase, as determined by a score of 60 or higher on the RRS (given this cutoff reflects a score of at least one standard deviation above the previously published college student mean; Yoon & Joormann, 2012). Additionally, we used the same cutoff on the SVS as for the full intervention trial (i.e., only those who rated the "academic/intellectual goals" item as "very important" or "extremely important" were invited to participate).

To select an optimal control condition for the full intervention trial, we piloted two alternative control conditions during the initial pilot phase: a "Free-Thinking" control condition (see "Intervention Conditions" section in main text) and a "Guided" control condition (see below). Thus, the final pilot sample included a total of seven participants in each of the five conditions: Why-Only, How-Only, Combined, and each of the two Control conditions. Participants' mean RRS score was 67.37 (*SD*=8.19), which was comparable to a previously published mean of 65.0 in a clinically diagnosed depressed sample (Watkins & Baracaia, 2002).

PILOT MEASURES AND PROCEDURES.

"Guided" Control condition.

In this alternative control condition, participants responded to a set of computer prompts that were matched in format to the active training conditions, but with no specific instruction to adopt a particular construal style: instead, participants were simply asked to specify "a thought that is coming to mind right now." For each prompt, participants typed their response into a textbox (limited to a couple lines of text) and then saw their response displayed on the screen for a fixed interval, during which time they were instructed to think about their response and what it means to them. Thus, the format, timing, and number of prompts presented in this Control condition were matched with the active training conditions.

Feasibility and Satisfaction Survey

This intervention credibility and satisfaction survey (adapted from MacGregor, Hayward, Peck, Wilkes, 2009) was administered to the 35 participants who enrolled in the initial "pilot" phase. Responses to this survey were analyzed both quantitatively and qualitatively after the first 35 participants were run, and minor changes were made to each construal training paradigm based on the survey feedback prior to starting the full intervention trial.

[For Combined, Why-only, How-only, and Guided Control conditions only]:

1. Do you think the thinking exercise you did was helpful in improving your performance on the 2nd set of academic aptitude tests?

No, it seemed to make it worse
No, it didn't really help
Yes, it helped somewhat
Yes, it helped a great deal

If "No": why not? _____

If "Yes": how was it helpful? _____

2. Do you think the thinking exercise you did helped reduce any negative feelings you felt

after completing the 1st set of academic aptitude tests?

No, it seemed to make them worse
No, it didn't really help
Yes, it helped somewhat
Yes, it helped a great deal

3. Do you think the thinking exercise helped reduce how much time you spent dwelling or mulling over your performance on the 1st set of academic aptitude tests?

No, it seemed to make it worse
No, it didn't really help
Yes, it helped somewhat
Yes, it helped a great deal

4. Do you think the thinking exercise helped increase your concentration on the 2nd set of academic aptitude tasks?

No, it seemed to make it worse
No, it didn't really help
Yes, it helped somewhat
Yes, it helped a great deal

5. Do you think the thinking exercise helped increase your motivation for the 2nd set of academic aptitude tasks?

No, it seemed to make it worse
No, it didn't really help
Yes, it helped somewhat
Yes, it helped a great deal

6. You were asked to spend 30 seconds thinking about each of the responses you typed in for the thinking exercise. Do you think that this was the right amount of time?

No, it was way too long
No, it was a bit too long
Yes, 30 seconds was the right amount
No, it was a bit too short
No, it was way too short

If no, how long do you think would have been the right amount of time? _____

[For Why-only and Combined conditions only:]

7. You were asked to provide 4 different reasons why it's important to you to improve your academic performance. Was this too many, too few, or just right?

Too many; it was hard to come up with that many good reasons.
Just right; I was able to come up with 4 good reasons.
Too few; I could've come up with more good reasons.

If "too many" or "too few," how many reasons do you think would be ideal? _____

8. For each of the 4 reasons you provided, you were then asked to name 1 further reason "why" the specified reason was important to you. Do you think this additional "why" question was a helpful component of the thinking exercise?

No, it made the exercise worse
No, it didn't really help
Yes, it helped somewhat
Yes, it helped a great deal

If "Yes," how was it helpful? _____

If "No," why not? _____

[For How-only and Combined conditions only:]

1. You were asked to provide 4 specific strategies by which you try to improve your academic performance. Was this too many, too few, or just right?

Too many; it was hard to come up with that many specific strategies.
Just right; I was able to come up with 4 specific strategies.
Too few; I could've come up with more specific strategies.

If "too many" or "too few," how many strategies do you think would be ideal? _____

2. For each of the 4 strategies you provided, you were then asked to name 1 further, more specific strategy by which you try to implement that strategy. Do you think this follow-up strategy question was a helpful component of the thinking exercise?

No, it made the exercise worse
No, it didn't really help
Yes, it helped somewhat
Yes, it helped a great deal

If "Yes," how was it helpful? _____

If "No," why not? _____

[For all conditions:]

What did you find most helpful about the exercise?

_____

What did you find least helpful about the exercise?

_____

_____

What would you recommend changing about the exercise?

_____

_____

What would you recommend keeping the same?

_____

_____

Any additional comments about your experience with the exercise that you'd like to share?

_____

PRELIMINARY RESULTS AND STUDY MODIFICATIONS.

Modifications to training paradigm. To refine the training paradigm based on preliminary results and feedback, participants' typed responses to each training prompt, as well as their written and oral responses to the credibility and satisfaction survey, were analyzed both quantitatively and qualitatively. For questions with a multiple-choice response format, relative frequencies of each response level ("No, it seemed to make it worse"; "No, it didn't really help"; etc.) were tallied and graphed for visual inspection. For open-ended questions, responses were read and recurring themes noted. Based on the combined results, the following refinements were made to the construal training procedures prior to the full intervention trial: the total number of blocks in each training condition was shortened from twelve (four reasons and/or strategies with two follow-up prompts for each) to four (two reasons and/or strategies with one follow-up prompt for each), and the focus period following each response was reduced from 30 to 20 seconds; and the wording of the "why" prompts was modified to tap "personal value" rather than "importance," to shift the emphasis from the importance of participants' recent setback to the

value of continued improvement. Appendix A displays the updated wording and format of each training condition.

Selecting a control condition. Given that the pilot phase included two candidate control conditions ("Guided" and "Free Thinking") that were expected to confer certain relative advantages and disadvantages, we plotted and visually compared the patterns of change from pre- to post-training in each condition. This preliminary comparison indicated that the slope of change was generally flatter (for approximately five out of the seven outcomes examined) in the "Free Thinking" condition, which led us to select it as the control condition for the full intervention trial.

Variability in rumination. First, to examine whether the cognitive performance measures elicited adequate levels and variability in off-task rumination, we evaluated the number of total rumination episodes (self-caught or probe-caught "off-task" thought episodes endorsed as self-focused and/or self-evaluative and negative in valence) within each baseline cognitive task for each participant. In a prior study (Unsworth & McMillan, 2012) using a similar version of the current thought-sampling procedure (but in which only probe-caught episodes were monitored), the frequency of off-task thought episodes reported across three total thought probes (M=1.28, SD=.93) showed significant negative correlations with reading comprehension performance as well as self-reported task motivation and interest. Given the relevance of these constructs to our study, we aimed to establish similar mean levels and variances of off-task rumination for each of our cognitive batteries. As shown in Table C1 below, our pilot sample showed comparable rates of both off-task mind-wandering (broadly defined as any off-task thought episode) and off-task rumination, both of which showed some of the same expected patterns of correlation with performance, mood, and motivation measures. Nonetheless, in an effort to further increase the variability in rumination (and particularly in probe-caught episodes) during each test battery, we

added one probe to each of the 10-minute cognitive tests and two probes to the 4-minute

problem-solving task (which previously had not included any sound probes), such that

participants now responded to 14 sound probes per test battery instead of eight.

Table C1. Descriptive statistics and correlations between mind-wandering / off-task rumination episodes, trait rumination, and select outcome measures (in pilot sample).

		M	*SD*	RRS	Creative Fluency		Reading Comprehension		SSSQ Rumination		SSSQ Motivation	
					Pre	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*
ind-wandering: *Test Battery1*	Probe-caught	1.18	1.18	-.20	-.04	-.09	-.14	-.33†	-.21	-.11	-.43*	-.43*
	Self-caught	5.38	3.30	-.14	.02	.03	-.15	-.19	.07	.16	-.12	-.13
Mind-wandering: *Test Battery2*	Probe-caught	1.59	1.71	-.25	.09	-.15	-.14	-.19	-.26	-.30†	-.43*	-.41*
	Self-caught	4.74	3.87	-.23	.18	-.06	.00	-.21	-.13	-.14	-.33†	-.43*
Off-Task Rumination: *Test Battery1*	Probe-caught	.39	.66	-.07	-.00	.04	-.16	-.00	.09	.19	-.04	.03
	Self-caught	1.94	2.42	.20	-.12	-.05	.01	-.18	.54**	.54**	-.07	-.03
Off-Task Rumination: *Test Battery 2*	Probe-caught	.47	.86	-.04	.06	-.21	-.22	-.44**	-.04	-.21	-.35*	-.27
	Self-caught	1.57	2.16	-.04	.15	-.14	.01	-.18	.29†	.15	-.21	-.26

†p<.10; *p<.05; **p<.01; ***p<.001
Note. SSSQ = Short State Stress Questionnaire; Pre = pre-training baseline test period; Post = post-training test period.

Enhancing believability of false feedback. Given the relatively high number of participants

who reported disbelieving our false negative feedback (9 out of the 35 pilot participants, i.e.,

27%), we reviewed participants' responses and feedback during the Funnel Debriefing interview

and made several modifications intended to increase the feedback's plausibility. Specifically, we:

1) increased the duration for which the experimenter was out of the room allegedly "entering and

scoring" participants' responses from five to seven minutes; 2) modified the percentiles, standard

scores, and verbal classifications (e.g., "low average") of participants' alleged test result printout

to be more realistically balanced and less extreme (see updated version in Appendix C); and 3)

added an official "writing test" component (the MEPS problem-solving task, which had

previously been administered as part of a post-test questionnaire battery and was not introduced

as a "test" on which participants would be evaluated), which the experimenter explained by

stating: "your response will be scored later on by two trained research assistants. Since that test takes longer to score, you won't receive feedback on it right away as you will for the other two tests." Overall, these efforts to increase the plausibility of our feedback appear to have been at least partially successful, as the full intervention trial included 14.4% nonbelievers—a substantial improvement over the original 27% in our pilot trial.

Appendix D
Study Forms and Measures

Vallacher & Wegner's (1989) original *Behavior Identification Form*

Any behavior can be described in many ways. For example, one person might describe a behavior as "writing a paper," while another person might describe the same behavior as "pushing keys on the keyboard." Yet another person might describe it as "expressing thoughts." This form focuses on your personal preferences for how a number of different behaviors should be described. Below you will find several behaviors listed. After each behavior will be two different ways in which the behavior might be identified.

For example:
1. Attending class
   a   sitting in a chair
   b   looking at a teacher

Your task is to choose the identification, a or b, that best describes the behavior for you. Be sure to respond to every item. Please select only one alternative for each pair. Remember, select the description that you personally believe is more appropriate for each pair.

1. Making a list
   a   Getting organized*
   b   Writing things down
2. Reading
   a   Following lines of print
   b   Gaining knowledge*
3. Joining the Army
   a   Helping the Nation's defense*
   b   Signing up
4. Washing clothes
   a   Removing odors from clothes*
   b   Putting clothes into the machine
5. Picking an apple
   a   Getting something to eat*
   b   Pulling an apple off a branch
6. Chopping down a tree
   a   Wielding an axe
   b   Getting firewood*

7. Measuring a room for carpeting
   a   Getting ready to remodel*
   b   Using a yard stick
8. Cleaning the house
   a   Showing one's cleanliness*
   b   Vacuuming the floor
9. Painting a room
   a   Applying brush strokes
   b   Making the room look fresh*
10. Paying the rent
   a   Maintaining a place to live*
   b   Writing a check
11. Caring for houseplants
   a   Watering plants
   b   Making the room look nice*
12. Locking a door
   a   Putting a key in the lock
   b   Securing the house*

13. Voting
   a   Influencing the election*
   b   Marking a ballot
14. Climbing a tree
   a   Getting a good view*
   b   Holding on to branches
15. Filling out a personality test
   a   Answering questions
   b   Revealing what you're like*
16. Toothbrushing
   a   Preventing tooth decay*
   b   Moving a brush around in one's mouth
17. Taking a test
   a   Answering questions
   b   Showing one's knowledge*
18. Greeting someone
   a   Saying hello
   b   Showing friendliness*
19. Resisting temptation
   a   Saying "no"
   b   Showing moral courage*

20. Eating
   a   Getting nutrition*
   b   Chewing and swallowing
21. Growing a garden
   a   Planting seeds
   b   Getting fresh vegetables*
22. Traveling by car
   a   Following a map
   b   Seeing countryside*
23. Having a cavity filled
   a   Protecting your teeth*
   b   Going to the dentist
24. Talking to a child
   a   Teaching a child something*
   b   Using simple words
25. Pushing a doorbell
   a   Moving a finger
   b   Seeing if someone's home*

* Higher-level (abstract) alternative.

Total score is the sum of higher-level alternative choices.

Modified Behavior Identification Form

Any behavior can be described in many ways. For example, one person might describe a behavior as "writing a paper," while another person might describe the same behavior as "pushing keys on the keyboard." Yet another person might describe it as "expressing thoughts." This form focuses on your personal preferences for how a number of different behaviors should be described. Below you will find several behaviors listed. After each behavior will be two different ways in which the behavior might be identified.

For example:
1. Attending class
A: sitting in a chair
B: looking at a teacher

Your task is to rate how well you feel that each identification describes the behavior for you. Please make each rating based on the following 1-5 scale:
1 = Not at all descriptive of the behavior
2 = Not very descriptive of the behavior
3 = Slightly descriptive of the behavior
4 = Very descriptive of the behavior
5 = Extremely descriptive of the behavior

1. Eating

a.   Getting nutrition
1 = Not at all descriptive of the behavior
2 = Not very descriptive of the behavior
3 = Slightly descriptive of the behavior
4 = Very descriptive of the behavior
5 = Extremely descriptive of the behavior

b. Chewing and swallowing
1 = Not at all descriptive of the behavior
2 = Not very descriptive of the behavior
3 = Slightly descriptive of the behavior
4 = Very descriptive of the behavior
5 = Extremely descriptive of the behavior

2. Greeting someone

a.   Saying hello
1 = Not at all descriptive of the behavior
2 = Not very descriptive of the behavior
3 = Slightly descriptive of the behavior
4 = Very descriptive of the behavior
5 = Extremely descriptive of the behavior

b.   Showing friendliness
1 = Not at all descriptive of the behavior
2 = Not very descriptive of the behavior
3 = Slightly descriptive of the behavior
4 = Very descriptive of the behavior
5 = Extremely descriptive of the behavior

3. Growing a garden

a.   Planting seeds
1 = Not at all descriptive of the behavior
2 = Not very descriptive of the behavior
3 = Slightly descriptive of the behavior
4 = Very descriptive of the behavior
5 = Extremely descriptive of the behavior

b.   Getting fresh vegetables
1 = Not at all descriptive of the behavior
2 = Not very descriptive of the behavior
3 = Slightly descriptive of the behavior
4 = Very descriptive of the behavior
5 = Extremely descriptive of the behavior

4. Having a cavity filled

a.   Protecting your teeth
1 = Not at all descriptive of the behavior
2 = Not very descriptive of the behavior
3 = Slightly descriptive of the behavior
4 = Very descriptive of the behavior
5 = Extremely descriptive of the behavior

b.   Going to the dentist
1 = Not at all descriptive of the behavior
2 = Not very descriptive of the behavior
3 = Slightly descriptive of the behavior
4 = Very descriptive of the behavior
5 = Extremely descriptive of the behavior

5. Locking a door

a. Putting a key in the lock 1 = Not at all descriptive of the behavior 2 = Not very descriptive of the behavior 3 = Slightly descriptive of the behavior 4 = Very descriptive of the behavior 5 = Extremely descriptive of the behavior	**b. Securing the house** 1 = Not at all descriptive of the behavior 2 = Not very descriptive of the behavior 3 = Slightly descriptive of the behavior 4 = Very descriptive of the behavior 5 = Extremely descriptive of the behavior

6. Making a list

a. Getting organized 1 = Not at all descriptive of the behavior 2 = Not very descriptive of the behavior 3 = Slightly descriptive of the behavior 4 = Very descriptive of the behavior 5 = Extremely descriptive of the behavior	**b. Writing things down** 1 = Not at all descriptive of the behavior 2 = Not very descriptive of the behavior 3 = Slightly descriptive of the behavior 4 = Very descriptive of the behavior 5 = Extremely descriptive of the behavior

7. Painting a room

a. Applying brush strokes 1 = Not at all descriptive of the behavior 2 = Not very descriptive of the behavior 3 = Slightly descriptive of the behavior 4 = Very descriptive of the behavior 5 = Extremely descriptive of the behavior	**b. Making the room look fresh** 1 = Not at all descriptive of the behavior 2 = Not very descriptive of the behavior 3 = Slightly descriptive of the behavior 4 = Very descriptive of the behavior 5 = Extremely descriptive of the behavior

8. Paying the rent

a. Maintaining a place to live 1 = Not at all descriptive of the behavior 2 = Not very descriptive of the behavior 3 = Slightly descriptive of the behavior 4 = Very descriptive of the behavior 5 = Extremely descriptive of the behavior	**b. Writing a check** 1 = Not at all descriptive of the behavior 2 = Not very descriptive of the behavior 3 = Slightly descriptive of the behavior 4 = Very descriptive of the behavior 5 = Extremely descriptive of the behavior

9. Resisting temptation

a. Saying "no" 1 = Not at all descriptive of the behavior 2 = Not very descriptive of the behavior 3 = Slightly descriptive of the behavior 4 = Very descriptive of the behavior 5 = Extremely descriptive of the behavior	**b. Showing moral courage** 1 = Not at all descriptive of the behavior 2 = Not very descriptive of the behavior 3 = Slightly descriptive of the behavior 4 = Very descriptive of the behavior 5 = Extremely descriptive of the behavior

10. Washing clothes

a. Removing odors from clothes 1 = Not at all descriptive of the behavior 2 = Not very descriptive of the behavior 3 = Slightly descriptive of the behavior 4 = Very descriptive of the behavior 5 = Extremely descriptive of the behavior	**b. Putting clothes into the machine** 1 = Not at all descriptive of the behavior 2 = Not very descriptive of the behavior 3 = Slightly descriptive of the behavior 4 = Very descriptive of the behavior 5 = Extremely descriptive of the behavior

Fujita et al's (2006) *Construal Check Form*

1. Sweeping the floor
      a. Moving a broom
      b. Being clean
2. Attending a family reunion
      a. Going to a picnic
      b. Respecting tradition
3. Skydiving
      a. Jumping out of an airplane
      b. Demonstrating one's daringness
4. Making an expensive purchase
      a. Swiping a credit card
      b. Doing something for one's pleasure
5. Staying home to study
      a. Reviewing one's notes
      b. Exerting self-discipline
6. Recycling
      a. Bagging paper, glass, and cans
      b. Caring for the environment
7. Teaching
      a. Talking to students
      b. Having authority
8. Meeting new people
      a. Small talk and shaking hands
      b. Enhancing one's social network

Ruminative Responses Scale

People think and do many different things when they feel depressed. Please read each of the items below and indicate whether you almost never, sometimes, often, or almost always think or do each one when you feel down, sad, or depressed. Please indicate what you generally do, not what you think you should do.

1 = almost never   2 = sometimes   3 = often   4 = almost always

1. think about how alone you feel
2. think "I won't be able to do my job if I don't snap out of this"
3. think about your feelings of fatigue and achiness
4. think about how hard it is to concentrate
5. think "What am I doing to deserve this?"
6. think about how passive and unmotivated you feel.
7. analyze recent events to try to understand why you are depressed
8. think about how you don't seem to feel anything anymore
9. think "Why can't I get going?"
10. think "Why do I always react this way?"
11. go away by yourself and think about why you feel this way
12. write down what you are thinking about and analyze it
13. think about a recent situation, wishing it had gone better
14. think "I won't be able to concentrate if I keep feeling this way."
15. think "Why do I have problems other people don't have?"
16. think "Why can't I handle things better?"
17. think about how sad you feel.
18. think about all your shortcomings, failings, faults, mistakes
19. think about how you don't feel up to doing anything
20. analyze your personality to try to understand why you are depressed
21. go someplace alone to think about your feelings
22. think about how angry you are with yourself

RRQ - Reflection Scale

For each of the statements below, please indicate your level of agreement or disagreement. Use the scale as shown below.

	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree
1. I love exploring my "inner" self.	○	○	○	○	○
2. I often love to look at my life in philosophical ways.	○	○	○	○	○
3. I love to meditate on the nature and meaning of things.	○	○	○	○	○
4. I don't really care for introspective or self-reflective thinking.	○	○	○	○	○
5. My attitudes and feelings about things fascinate me.	○	○	○	○	○
6. I love analyzing why I do things.	○	○	○	○	○
7. I don't care much for self-analysis.	○	○	○	○	○
8. I'm not really a meditative type of person.	○	○	○	○	○
9. Philosophical or abstract thinking doesn't appeal to me that much.	○	○	○	○	○
10. Contemplating myself isn't my idea of fun.	○	○	○	○	○
11. People often say I'm a "deep," introspective type of person.	○	○	○	○	○
12. I'm very self-inquisitive by nature.	○	○	○	○	○

Depression Anxiety Stress Scales (DASS21; A = Anxiety, D = Depression)

Please read each statement and circle a number 0, 1, 2 or 3 that indicates how much the statement applied to you *over the past week*. There are no right or wrong answers. Do not spend too much time on any statement.

0 Did not apply to me at all
1 Applied to me to some degree, or some of the time
2 Applied to me to a considerable degree, or a good part of time
3 Applied to me very much, or most of the time

1	I found it hard to wind down (S)	0	1	2	3
2	I was aware of dryness of my mouth (A)	0	1	2	3
3	I couldn't seem to experience any positive feeling at all (D)	0	1	2	3
4	I experienced breathing difficulty (eg, excessively rapid breathing, breathlessness in the absence of physical exertion) (A)	0	1	2	3
5	I found it difficult to work up the initiative to do things (D)	0	1	2	3
7	I experienced trembling (eg, in the hands) (A)	0	1	2	3
9	I was worried about situations in which I might panic and make a fool of myself (A)	0	1	2	3
10	I felt that I had nothing to look forward to (D)	0	1	2	3
13	I felt down-hearted and blue (D)	0	1	2	3
15	I felt I was close to panic (A)	0	1	2	3
16	I was unable to become enthusiastic about anything (D)	0	1	2	3
17	I felt I wasn't worth much as a person (D)	0	1	2	3
19	I was aware of the action of my heart in the absence of physical exertion (eg, sense of heart rate increase, heart missing a beat) (A)	0	1	2	3
20	I felt scared without any good reason (A)	0	1	2	3
21	I felt that life was meaningless (D)	0	1	2	3

Positive and Negative Affect Schedule (PANAS)

This scale consists of a number of words and phrases that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you feel this way in general. Use the following scale to record your answers:

1	2	3	4	5
Very slightly or not at all	A little	Moderately	Quite a bit	Extremely

_____afraid _____guilty _____scared _____ashamed _____upset

_____jittery _____nervous _____irritable _____hostile _____distressed

_____active _____alert _____attentive _____determined _____enthusiastic

_____excited _____inspired _____interested _____proud _____strong

Informed Consent Agreement

Please read this consent agreement carefully before you decide to participate in the study.

Purpose of the research study: The purpose of the study is to examine how various emotional and thinking styles are related to a person's intellectual or academic performance, and how this performance is influenced by feedback.

What you will do in the study: You will be asked to complete several questionnaires about your mood, personality, and thinking style, and you will be asked to complete a number of academic tasks on which you may receive feedback. You will also complete a number of brief exercises that promote different thinking styles. Throughout the experiment, you can skip any question or item that makes you uncomfortable, and you can stop the study at any time without penalty.

Time required: The study will require about 2 hours of your time.

Risks: There is a possibility that you may experience some temporary distress or anxiety when completing the tasks or answering the questionnaires. You are free to skip any portion of the study that you wish or to stop the study at any time.

Benefits: There are no direct benefits to you for participating in this research study. The study may increase our understanding of the different cognitive or emotional processes that contribute to academic performance, which could help us to identify strategies for improving academic performance and overall psychological functioning in the future.

Confidentiality: The information that you give in the study will be handled confidentially. The only exceptions to this guideline are if we learn of possible child or elderly abuse or danger to self or others. Your information will be assigned a code number. The list connecting your name to this code will be kept in a locked file. When the study is completed and the data have been analyzed, this list will be destroyed. Your name will not be used in any report.

Voluntary participation: Your participation in the study is completely voluntary.

Right to withdraw from the study: You have the right to withdraw from the study at any time without penalty.

How to withdraw from the study: If you want to withdraw from the study, tell the experimenter you want to stop. There is no penalty for withdrawing. You will still receive full credit or payment for the study session. If you would like to withdraw after your materials have been submitted, please contact the principal investigator, Gena Gorlin, by phone at (434) 243-7646 or by e-mail at eig7qr@virginia.edu. You will be debriefed if you withdraw from the study, and you will have the option to remove your data.

Revision date: 11/01/11
Page 1

Post-Debrief Consent Form for Deception Studies
Project Title: Psychological Factors in Academic Performance

During the experiment, you were asked to complete a battery of academic aptitude tests, and you were told that these tests predict future academic and intellectual achievement. You were also given feedback suggesting that your performance on these tests was relatively poor. In fact, none of this was true. **The test feedback we gave you in this study condition was completely fake, and the tests you took were purposefully designed to be practically impossible to complete in the allotted time.** You were given this false feedback in order to induce failure-related rumination, so that we could then test whether the thinking exercise you completed afterward could help alleviate some of the negative effects of ruminating on this perceived failure.

Because you were deceived, you now have the right to refuse to allow your test and questionnaire responses to be used and to ask that they be destroyed immediately. If you do so, there is no penalty. You will still receive full credit/payment for participating.

_____ I give permission for my test and questionnaire responses to be used in the analysis for this experiment.

_____ I do NOT give my permission for my responses to be used in the analysis for this experiment. Please withdraw them from the study and destroy them immediately.

Signature: _____ Date: _____

You will receive a copy of this form for your records.

Revision Date: 11/01/11

IRB-SBS Office Use Only		
Protocol #	2014-0003-00	
Approved	from: 1/22/14	to: 1/21/15
SBS Staff		

Instructions: Each passage below is followed by questions about its content. A passage, select the best answer to each question by circling the letter next to you Answer all questions following a passage on the basis of what the passage state:

Passage 1.

The tale of Piltdown Man, the most infamous forgery in the contentious detective story of the origins of mankind, began in 1912. On December 18 that year Charles Dawson, a well-known amateur British archaeologist, and Arthur Smith Woodward, of
5 the British Museum of Natural History, announced the discovery of some amazing human fossils. The remains comprised nine pieces of skull, a broken jaw with two teeth in place, a few stone tools, and some animal bones, all of which had been discovered on a farm near Piltdown Common in Sussex.

10 When pieced together the skull looked distinctly human. Although Piltdown Man, as the hominid became known, had unusually thick bones, the brain case was large and rounded. There was no sign of prominent brow ridges or other apelike features. However, the shape of the jaw bone resembled that of an ape. The only human
15 characteristic of this jaw was the wear on the two molars, which were ground down flat, as is frequently true of hominids who eat tough or abrasive foods, such as seeds. In other words the creature had the jaw of an ape and the skull of *Homo sapiens.*

The primitive stone tools found with these remains suggested a
20 remote age for Piltdown Man, perhaps the Early Pleistocene or even the Late Pliocene. (In 1912 experts thought the Pliocene lasted from 1 million to 600 000 years ago. Scientists now date it to between 5 million and 1.7 million years ago.) This date was also supported by some animal bones found with Piltdown Man.

25 To most scientists of the time, Piltdown Man fulfilled a prediction made by the pioneering evolutionist Charles Darwin, who had believed that humans and the apes could be connected genetically through a still undiscovered creature. Most significantly, it was half-human in precisely the feature
30 that was then accepted as the most important difference between humans and the apes - the brain. At this time there was little fossil evidence to contradict the idea that the brain was among the first of the human features to evolve.

As time went on, however, *Homo erectus* fossils were found in
35 Java and China, while in South Africa the australopithecines were being discovered. All these fossils had human-like jaws and teeth and relatively small brains in contrast to Piltdown Man's large cranium and apelike jaw. The large brain simply

did not fit with the rest of the fossil evidence. By 1948
40 scientists knew that bones buried in the earth gradually
absorb fluorine. The older a bone, the more fluorine it
contains. When the Piltdown materials were tested for fluorine,
the skull and jaw fragments turned out to be much younger
than the Early Pleistocene animal bones with which the skull
45 had been found.

Scientists were now very suspicious. In 1953 all the Piltdown
material was tested for its authenticity. Not only was the
recent age of the jaw and skull confirmed, but the jaw proved
to be that of a modern orangutan, with the teeth filed down
50 in a quite obvious manner to imitate wear on human teeth.
But the forger had not stopped there. A bone tool found
with the remains had been made in recent times with a steel
knife, which leaves different marks than does a stone flake
or axe. The tools, as well as the animal bones, had been
55 taken from different archaeological sites.

Once the forgery was exposed by modern scientific analysis
the mystery was no longer where Piltdown Man came in human
evolution but who was responsible for the hoax, and why?
Although Dawson, the discoverer of most of the Piltdown
60 material, is frequently singled out as the person responsible
for this practical joke, there is no definite proof and the
question is far from settled.

1. The Piltdown skull seemed distinctly human because it had *(Select ALL answer choices that apply)*
A. a large brain
B. thick bones
C. brow ridges

2. The scientists of the time made which ONE of the following mistakes:
A. believed that fossil discoveries would reveal much about human origins
B. had preconceived ideas about what features an early hominid should have
C. followed the ideas of Darwin in the face of counterevidence
D. incorrectly judged the size of the brain
E. failed to examine other fossil evidence available at the time

3. The animal bones found buried with the Piltdown Man were all of the following **except**
A. shown to be genuinely Pleistocene
B. more recent than first thought
C. unconnected with the human remains
D. deliberately planted at the site
E. not originally from the Piltdown site

Passage 2:

It is exceedingly difficult to make people realize that an evil
is an evil. For instance, we seize a man and deliberately do
him a malicious injury: say, imprison him for years. One would
not suppose that it needed any exceptional clearness of wit to
5 recognize in this an act of diabolical cruelty. But in England
such a recognition provokes a stare of surprise, followed by an
explanation that the outrage is punishment or justice or
something else that is all right, or perhaps by a heated attempt
to argue that we should all be robbed and murdered in our beds
10 if such senseless villainies as sentences of imprisonment were
not committed daily. It is useless to argue that even if this
were true, which it is not, the alternative to adding crimes
ofour own to the crimes from which we suffer is not helpless
submission. Chickenpox is an evil; but if I were to declare
15 that we must either submit to it or else repress it by seizing
everyone who suffers from it and punishing them by inoculation
with smallpox, I should be laughed at; for though nobody could
deny that the result would be to prevent chickenpox to some
extent by making people avoid it much more carefully, and to
20 effect a further apparent prevention by making them conceal
it very anxiously, yet people would have sense enough to see
that the deliberate propagation of smallpox was a creation of
evil, and must therefore be ruled out in favor of purely humane
and hygienic measures. Yet in the precisely parallel case of a
25 man breaking into my house and stealing my diamonds I am
expected as a matter of course to steal ten years of his life.
If he tries to defeat that monstrous retaliation by shooting
me, my survivors hang him. The net result suggested by the
police statistics is that we inflict atrocious injuries on the
30 burglars we catch in order to make the rest take effectual
precautions against detection; so that instead of saving our
diamonds from burglary we only greatly decrease our chances
of ever getting them back, and increase our chances of being
shot by the robber.

35 But the thoughtless wickedness with which we scatter
sentences of imprisonment is as nothing compared to the
stupid levity with which we tolerate poverty as if it were
either a wholesome tonic for lazy people or else a virtue to
be embraced as St. Francis embraced it. If a man is indolent,
40 let him be poor. If he is drunken, let him be poor. If he is
not a gentleman, let him be poor. If he is addicted to the
fine arts or to pure science instead of to trade and finance,
let him be poor. If he chooses to spend his wages on his beer

and his family instead of saving it up for his old age, let
45 him be poor. Let nothing be done for "the undeserving": let
him be poor. Serve him right! Also -- somewhat inconsistently
-- blessed are the poor!

1. The passage is most probably intended to
A. serve as an introduction to a more detailed discussion of poverty
B. censure imprisonment as a punitive measure
C. analyze the possible repercussions of social evils
D. continue a prior discussion of strong measures against social evils
E. make people recognize social evils in the face of deliberate obfuscation

2. It can be inferred from the passage that the author would agree with all the following **except**
A. most people don't realize that by punishing offenders they are surrendering themselves to the vicious cycle of crime and punishment
B. sentences of imprisonment have little success in reducing the crime rate in society
C. it would be ridiculous to inoculate people suffering from chicken pox with small pox
D. if criminals were not strongly punished for their misdeeds there would be no law and order in society
E. tolerating poverty is at least as bad as inflicting punishments on criminals

3. The author's argument about imprisonment would be most weakened by showing that
A. imprisonment is not widely regarded as an act of cruelty
B. chicken pox and burglary are not analogous evils
C. imprisonment does not cause malicious injury
D. sentences of imprisonment are given increasingly rarely
E. a burglar who commits murder in self defense would not be hanged

4. The author apparently believes that people at the time he wrote the passage were
A. inclined to consider poverty a social evil
B. anxious to take the right steps to ensure an orderly society
C. too ready to judge other people unfairly
D. inconsistent in their attitude to poverty
E. in favor of unusually harsh punishment of all offenders

Passage 3:

The principle of selection solved the riddle as to how what
was purposive could conceivably be brought about without the
intervention of a directing power, the riddle which animate
nature presents to our intelligence at every turn, and in
5 face of which the mind of a Kant could find no way out, for
he regarded a solution of it as not to be hoped for. For,
even if we were to assume an evolutionary force that is
continually transforming the most primitive and the simplest
forms of life into ever higher forms, and the homogeneity of
10 primitive times into the infinite variety of the present, we
should still be unable to infer from this alone how each of
the numberless forms adapted to particular conditions of life
should have appeared precisely at the right moment in the
history of the earth to which their adaptations were
15 appropriate, and precisely at the proper place in which all
the conditions of life to which they were adapted occurred:
the humming-birds at the same time as the flowers; the
trichina at the same time as the pig; the bark-coloured moth
at the same time as the oak, and the wasp-like moth at the
20 same time as the wasp which protects it. Without processes
of selection we should be obliged to assume a
"pre-established harmony" after the famous Leibnitzian model,
by means of which the clock of the evolution of organisms
is so regulated as to strike in exact synchronism with that
25 of the history of the earth!

All forms of life are strictly adapted to the conditions
of their life, and can persist under these conditions alone.
There must therefore be an intrinsic connection between the
conditions and the structural adaptations of the organism,
30 and, since the conditions of life cannot be determined by
the animal itself, the adaptations must be called forth by
the conditions. The selection theory teaches us how this
is conceivable, since it enables us to understand that there
is a continual production of what is non-purposive as well
35 as of what is purposive, but the purposive alone survives,
while the non-purposive perishes in the very act of arising.
This is the old wisdom taught long ago by Empedocles.

Reading Comprehension – Test A

1. It can be inferred that the author believes that the "Leibnitzian model" (line 22) is
A. ingenious and worthy of serious consideration
B. untenable by all rational people
C. an acceptable solution to Kant's dilemma
D. unworthy of further consideration
E. an alternative that might still be valid

2. The author's primary purpose in this extract is to
A. suggest that a particular theory explains otherwise puzzling phenomena
B. describe the details of the selection theory for a lay audience
C. justify a particularly controversial model of the origins of life
D. persuade the reader that Empedocles was right
E. prove that selection is the only possible way of looking at evolutionary biology

3. The examples in lines 17 - 19 are intended to
A. reinforce the author's point that is difficult to explain adaptation
B. show that adaptations must take place only at specific times and in specific places
C. give specific illustration of organisms that are particularly well-adapted to their conditions
D. show organisms that have evolved synchronously in a predestined manner
E. demonstrate that intelligent design is needed for purposive evolution

Reading Comprehension – Test B

Instructions: Each passage below is followed by questions about its content. After reading a passage, select the best answer to each question by circling the letter next to your response. Answer all questions following a passage on the basis of what the passage states or implies.

Passage 1.

   The first and most important rule of legitimate or popular
  government, that is to say, of government whose object is the
  good of the people, is therefore, as I have observed, to follow
  in everything the general will. But to follow this will it is
5 necessary to know it, and above all to distinguish it from the
  particular will, beginning with one's self: this distinction is
  always very difficult to make, and only the most sublime virtue
  can afford sufficient illumination for it. As, in order to will,
  it is necessary to be free, a difficulty no less great than the
10 former arises — that of preserving at once the public liberty
  and the authority of government. Look into the motives which have
  induced men, once united by their common needs in a general
  society, to unite themselves still more intimately by means of
  civil societies: you will find no other motive than that of
15 assuring the property, life and liberty of each member by the
  protection of all. But can men be forced to defend the liberty
  of any one among them, without trespassing on that of others?
  And how can they provide for the public needs, without alienating
  the individual property of those who are forced to contribute to
20 them? With whatever sophistry all this may be covered over, it is
  certain that if any constraint can be laid on my will, I am no
  longer free, and that I am no longer master of my own property,
  if any one else can lay a hand on it. This difficulty, which would
  have seemed insurmountable, has been removed, like the first, by
25 the most sublime of all human institutions, or rather by a divine
  inspiration, which teaches mankind to imitate here below the
  unchangeable decrees of the Deity. By what inconceivable art has a
  means been found of making men free by making them subject; of
  using in the service of the State the properties, the persons and
30 even the lives of all its members, without constraining and without
  consulting them; of confining their will by their own admission; of
  overcoming their refusal by that consent, and forcing them to punish
  themselves, when they act against their own will? How can it be that
  all should obey, yet nobody take upon him to command, and that all
35 should serve, and yet have no masters, but be the more free, as, in
  apparent subjection, each loses no part of his liberty but what might
  be hurtful to that of another? These wonders are the work of law. It
  is to law alone that men owe justice and liberty. It is this salutary
  organ of the will of all which establishes, in civil right, the
40 natural equality between men. It is this celestial voice which
  dictates to each citizen the precepts of public reason, and teaches

him to act according to the rules of his own judgment, and not to
behave inconsistently with himself. It is with this voice alone that
political rulers should speak when they command; for no sooner does
45 one man, setting aside the law, claim to subject another to his
private will, than he departs from the state of civil society, and
confronts him face to face in the pure state of nature, in which
obedience is prescribed solely by necessity.

1. The paradox in line 28 is resolved according to the author when an individual
A. submits to the rule of law and thus is at liberty to do anything that does not harm another person
B. behaves according to the natural rights of man and not according to imposed rules
C. agrees to follow the rule of law even when it is against his best interests
D. belongs to a society which guarantees individual liberty at all times
E. follows the will of the majority

2. The author's attitude to law in this passage is best conveyed as
A. respect for its inalienable authority
B. extolling its importance as a human institution
C. resignation to the need for its imposition on the majority
D. acceptance of its restrictions
E. praise for its divine origin

3. The author would agree with all of the following **except**
A. government must maintain its authority without unduly compromising personal liberty
B. individual freedom is threatened in the absence of law
C. justice cannot be ensured in the absence of law
D. political leaders should use the law as their guide to correct leadership
E. the law recognizes that all men are capable of recognizing what is in the general interest

Passage 2.

In regard to propaganda the early advocates of universal
literacy and a free press envisaged only two possibilities:
the propaganda might be true, or it might be false. They did
not foresee what in fact has happened, above all in our
5 Western capitalist democracies - the development of a vast
mass communications industry, concerned in the main neither
with the true nor the false, but with the unreal, the more
or less totally irrelevant. In a word, they failed to take
into account man's almost infinite appetite for distractions.

10 In the past most people never got a chance of fully
satisfying this appetite. They might long for distractions,
but the distractions were not provided. Christmas came but
once a year, feasts were "solemn and rare," there were few
readers and very little to read, and the nearest approach
15 to a neighborhood movie theater was the parish church, where
the performances, though infrequent, were somewhat monotonous.
For conditions even remotely comparable to those now prevailing
we must return to imperial Rome, where the populace was kept
in good humor by frequent, gratuitous doses of many kinds of
20 entertainment - from poetical dramas to gladiatorial fights,
from recitations of Virgil to all-out boxing, from concerts
to military reviews and public executions. But even in Rome
there was nothing like the non-stop distraction now provided
by newspapers and magazines, by radio, television and the
25 cinema. In Brave New World non-stop distractions of the most
fascinating nature (the feelies, orgy-porgy, centrifugal
bumblepuppy) are deliberately used as instruments of policy,
for the purpose of preventing people from paying too much
attention to the realities of the social and political
30 situation. The other world of religion is different from
the other world of entertainment; but they resemble one
another in being most decidedly "not of this world." Both
are distractions and, if lived in too continuously, both
can become, in Marx's phrase, "the opium of the people"
35 and so a threat to freedom. Only the vigilant can maintain
their liberties, and only those who are constantly and
intelligently on the spot can hope to govern themselves
effectively by democratic procedures. A society, most of
whose members spend a great part of their time, not on the
40 spot, not here and now and in the calculable future, but
somewhere else, in the irrelevant other worlds of sport and
soap opera, of mythology and metaphysical fantasy, will find
it hard to resist the encroachments of those who would
manipulate and control it.

45   In their propaganda today's dictators rely for the most
part on repetition, suppression and rationalization - the
repetition of catchwords which they wish to be accepted as
true, the suppression of facts which they wish to be ignored,
the arousal and rationalization of passions which may be

50 used in the interests of the Party or the State. As the art
and science of manipulation come to be better understood,
the dictators of the future will doubtless learn to combine
these techniques with the non-stop distractions which, in
the West, are now threatening to drown in a sea of

55 irrelevance the rational propaganda essential to the
maintenance of individual liberty and the survival of
democratic institutions.

1. The author would be most likely to agree that propaganda
A. can serve a vital function in democracy
B. is concerned mainly with the irrelevant
C. is now combined with entertainment
D. is universally recognized as a danger
E. needs constant vigilance to avoid

2. The "early advocates of universal literacy" (line 1) are mentioned as
A. advocates of propaganda
B. opponents of an idea that the author thinks is correct
C. proponents of an idea that the author wishes to counter
D. people who made wrong predictions about freedom of the press
E. social commentators unaware of man's appetite for distractions

3. The author refers to "Brave New World" as a fictional example of a society in which
A. non-stop distractions are the main instrument of government policy
B. people are totally unaware of political realities
C. entertainment is used to keep people from full awareness of social realities
D. entertainment resembles religion in its effects on the masses
E. non-stop entertainment is provided as it was in Rome

4. By "intelligently on the spot" (line 37) the author apparently means
A. alert to the dangers of propaganda
B. in a particular society at a particular time
C. in a specific time and place
D. conscious of political and social realities
E. deeply aware of current trends

Passage 3.

Nadezhda Krupskaya, in her little book on Lenin, relates that
towards the end of his life Lenin went to see a dramatized
version of The Cricket on the Hearth, and found Dickens's
'middle-class sentimentality' so intolerable that he walked
5   out in the middle of a scene. It is worth noticing that the
dislike of Dickens implied in this anecdote is something
unusual. Plenty of people have found him unreadable, but very
few seem to have felt any hostility towards the general spirit
of his work. Some years later, Bechhofer Roberts published a
10 full-length attack on Dickens in the form of a novel, but it
was a merely personal attack, concerned for the most part with
Dickens's treatment of his wife. It dealt with incidents which
not one in a thousand of Dickens's readers would ever hear
about, and which no more invalidates his work than the
15 **second-best bed** invalidates Hamlet. All that the book really
demonstrated was that a writer's literary personality has little
or nothing to do with his private character.

It is quite possible that in private life Dickens was just the
kind of insensitive egoist that Bechhofer Roberts makes him
20 appear. But in his published work there is implied a
personality quite different from this, a personality which has
won him far more friends than enemies. It might well have been
otherwise, for even if Dickens was a bourgeois, he was certainly
a subversive writer, a radical, one might truthfully say a rebel.
25 Everyone who has read widely in his work has felt this. Gissing,
for instance, the best of the writers on Dickens, was anything
but a radical himself, and he disapproved of this strain in
Dickens and wished it were not there, but it never occurred to
him to deny it. In Oliver Twist, Hard Times, Bleak House,
30 Little Dorrit, Dickens attacked English institutions with a
ferocity that has never since been approached. Yet he managed
to do it without making himself hated, and, more than this,
the very people he attacked have swallowed him so completely
that he has become a national institution himself. In its
35 attitude towards Dickens the English public has always been a
little like the elephant which feels a blow with a walking-stick
as a delightful tickling. Before I was ten years old I was
having Dickens ladled down my throat by schoolmasters in whom
even at that age I could see a strong resemblance to Mr. Creakle,
40 and one knows without needing to be told that lawyers delight
in Sergeant Buzfuz and that Little Dorrit is a favourite in the
Home Office. Dickens seems to have succeeded in attacking
everybody and antagonizing nobody.

1. It can be inferred that the 'second best bed' (highlighted in the first paragraph) refers to something that

A. could not be considered unpleasant in the personal life of Shakespeare (the author of Hamlet)

B. is unwarranted in the plot of Hamlet

C. most readers would approve of if they were aware of it

D. is irrelevant in a discussion of the personality of Shakespeare

E. has no place in an evaluation of the literary merit of the works of Shakespeare

2. Select the sentence that specifically illustrates what the author means in saying "the very people he attacked have swallowed him so completely."

A. "Plenty of people have found him unreadable, but very few seem to have felt any hostility towards the general spirit of his work."

B. "In Oliver Twist, Hard Times, Bleak House, Little Dorrit, Dickens attacked English institutions with a ferocity that has never since been approached."

C. "In its attitude towards Dickens the English public has always been a little like the elephant which feels a blow with a walking-stick
as a delightful tickling."

D. "Before I was ten years old I was having Dickens ladled down my throat by schoolmasters in whom even at that age I could see a strong resemblance to Mr. Creakle, and one knows without needing to be told that lawyers delight in Sergeant Buzfuz and that Little Dorrit is a favourite in the Home Office."

3. The author apparently believes that Gissing *(Select ALL answer choices that apply)*:

A. thought that Dickens was a subversive writer

B. disapproved of a certain aspect of Dickens' writing

C. produced good critical writing on Dickens

Thought Tracking Sheet, *Creative Fluency A*

SOUND

Off-task?	Self-evaluative?	Self-focused?	How positive or negative? 1 = "Extremely negative"; 5 = "Extremely positive"
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5

SELF

Off-task? ✔	Self-evaluative?	Self-focused?	How positive or negative? 1 = "Extremely negative"; 5 = "Extremely positive"
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5

Thought Tracking Sheet, *Reading Comprehension A*

SOUND

Off-task?	Self-evaluative?	Self-focused?	How positive or negative? 1 = "Extremely negative"; 5 = "Extremely positive"
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5
Y / N	Y / N	Y / N	1 2 3 4 5

SELF

Off-task? ✔	Self-evaluative?	Self-focused?	How positive or negative? 1 = "Extremely negative"; 5 = "Extremely positive"
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5
	Y / N	Y / N	1 2 3 4 5

Creative Fluency – Test A

Please list as many *creative* but *possible* uses as you can for a **paper clip.**

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	

Instructions for Judging Creativity
Creativity can be viewed as having three facets. Creative responses will generally be high on all three, although being low on one of them does not disqualify a response from getting a high rating. We will use a 1 to 5 scale:

| 1 | 2 | 3 | 4 | 5 |

Not at all creative   Highly creative

1. Uncommon. Creative ideas are uncommon: they will occur infrequently in our sample. Any response that is given by a lot of people is common, by definition. Unique responses will tend to be creative responses, although a response given only once needn't be judged as creative. For example, a random or inappropriate response would be uncommon but not creative.

2. Remote. Creative ideas are remotely linked to everyday objects and ideas. For example, creative uses for a brick are ―far from common, everyday, normal uses for a brick, and creative instances of things that are round are ―far from common round objects. Responses that stray from obvious ideas will tend to be creative, whereas responses close to obvious ideas will tend to be uncreative.

3. Clever. Creative ideas are often clever: they strike people as insightful, ironic, humorous, fitting, or smart. Responses that are clever will tend to be creative responses. Keep in mind that cleverness can compensate for the other facets. For example, a common use cleverly expressed could receive a high score.

For your reference, below is a rubric with some sample responses that would receive each rating (this list is not exhaustive!). PLEASE LOOK OVER THIS RUBRIC AND CONSULT IT WHILE MAKING YOUR RATINGS.

1-5 Rating Rubric (with sample responses):					
	"1" responses	**"2" responses**	**"3" responses**	**"4" responses**	**"5" responses**
For "brick":	Build a house/building	Build a road/castle/bridge	Build a desk	Use brick as a stamp	Crack it, put water in crack & freeze to demonstrate weathering
	Make a wall	Break window/hurt someone	Base for a table	Brick shoes	Knock a chunk off & present the rest as modern interpretist art
		Prop something open	Hold a car wheel from rolling	Pet brick	
		Make something sink	Weight for working out		
For "paperclip":	Clip papers together	Jewelry (earring, bracelet, etc)	Toy / building tool	Glasses accessory	To pick up your roomate's underwear
		Poke a hole in something	Scrape out dirt from small space	Push reset button on electrical device	Marshmallow stick for smores (giant paperclip)
		Chain	Word art	make a spring	Punch a pinhole to view a solar eclipse
		Weapon / hurt someone	Christmas ornament hanger	A crown for a Barbie	
		Pick a lock	Use them to make 3D shapes	Used in a magnetism experiment	

Means-Ends Problem-Solving (MEPS) Task – Instructions (adapted from Lyubomirsky & Nolen-Hoeksema, 1995)

Description of the task:
Instructions (for both scenarios):
"We are often challenged by different kinds of social problems in our daily life. Below is a hypothetical problem situation; please list the steps you could take to reach the happy ending, given how the situation starts."

Note: Participants were given 4 minutes to complete their response, so some participants' responses will not be complete (in which case their "effectiveness" rating may suffer accordingly).

"Friend" scenario:
"You notice that one of your friends seems to be avoiding you. You really like and enjoy spending time with this person, and want him or her to like you. The situation ends when he or she likes you again. Please begin the story when you notice your friend avoiding you."

"Professor" scenario:
"You are being treated unfairly by one of your professors. The class is an important one for your major, and you very much want to do well in it. The situation ends when you and the professor are getting along again, and you no longer feel you are being treated unfairly. Please list the steps you could take to reach the happy ending, starting from when you notice the professor is being unfair to you."

INSTRUCTIONS FOR ASSESSING MODEL SOLUTIONS AND PROBLEM-SOLVING EFFECTIVENESS

(These were the instructions given to the coders)

For each story response, please follow the steps outlined below.

1. Read the WHOLE story response.

2. Count the number of solutions that belong to one of the "model" solution categories below. (Remember that these should all be solutions to the problem, not events or circumstances that occur AFTER the resolution.) Enter this number in the "Num_Model_Solutions" column.

3. If the response contains potentially effective solutions not listed on the category sheet, type these in the "Notes" column.

5. Read through all of the model solution categories ONCE AGAIN to make sure that you have included all that apply.

6. Now make a global rating about the effectiveness of the WHOLE story response. Make sure to read the whole story response again and think about how effective is the entire set of solution(s) and/or strateg(ies) offered. Effective solutions or strategies (or means) are ones that

are appropriate and useful in enabling one to reach the goal or resolution of the problem or to overcome an obstacle(s). Note that what is effective may vary with the situation presented.

Please use the model solutions as a guide to making this rating. For example, consider a story response that includes all or almost all of the model solutions as a 7. Consider a story response that has few of the model solutions as a 1.

You may take into account how many model categories (M's) are included in the subject's category list, but do not use this as the sole indicator of story response effectiveness. Rate the effectiveness of the story response as a WHOLE.

Please use the following scale when making your rating:

1   2   3   4   5   6   7

MODEL RESPONSE CATEGORIES FOR "FRIEND"

1. (M) First think about possible reasons for your friend's avoiding you, what you may have done to upset or anger your friend.

2. (M) Prepare for the interaction. (For example, think about what you should do OR think about what to say or how to approach the issue OR anticipate possible reactions OR try to be calm or cool.)

3. (M) First make an effort to see the friend more, make friendly advances toward him or her. (This is not to confront but simply to see if the friend is really avoiding you.)

4. (M) Go see the friend in person (may call first).

5. (M) Approach the issue with the friend. Ask what is wrong, what the problem is. (Do this nicely and/or tactfully. For example, do not say "Why are you avoiding me?". Instead, say something like "Is anything wrong?" or "We haven't spent much time together lately" or "Did it upset you when I...?", etc.)

6. (M) Openly discuss the problem. Explain how and why you did whatever you did.

7. (M) Apologize and/or say that whatever you did was not intentional (if the latter is appropriate).

8. (M) Discuss how to correct the problem. Say you will change your behavior OR come up with a way to compromise.

9. (M) Say something to reaffirm the friendship (e.g., say how much you value the friendship, OR say how much it means to you to spend time with the friend OR let the friend know how much you like him or her).

10. (M) Make it up to the friend by doing something nice for him or her. (This only counts as a model response if the underlying problem has been dealt with.)

Examples of NON-model responses:

1. Do not do anything for a while at first (e.g., wait for friend to do something).

2. First avoid friend too OR act mean or insensitive.

3. Do something to distract yourself from the problem (e.g., drink, listen to music, jog, go out).

4. Ask someone else first if and why they think the friend is avoiding you (e.g., mutual friend, roommate, etc.)

5. Seek advice from someone else first about what to say or do (e.g., friend, parent, etc.).

6. Call the friend (i.e., discuss the problem on the phone).

7. When talking to the friend, act cool or distant, OR complain, OR blame friend, OR criticize, OR argue, OR refute friend's complaint (in a negative way, without explaining), OR lie to friend, OR confront friend NOT in a nice way.

MODEL RESPONSE CATEGORIES FOR "PROFESSOR"

1. (M) First think about why your professor may be treating you this way, what may be wrong. (For example, think about whether you are having a conceptual problem or whether you did not hand something in or whether you are not working hard enough.) Note: *This may include checking with other students in the class to test your assumption that you are indeed being treated unfairly.*

2. (M) Prepare for the interaction (e.g., think about what you're going to say, gather specific examples to support your concern about being treated unfairly, etc.).

3. (M) Meet the professor in person.

4. (M) Share your feelings and concerns with the professor, in a respectful and non-confrontational manner.

5. (M) Ask professor what the problem is, what you may be doing wrong.

6. (M) Ask the professor what you can do to improve your performance in the class. (For example, ask how you should approach the material, OR ask if you can get extra help, OR if you can do an extra paper or homework assignment, OR if you can get another chance at a test.)

7. (M) Discuss how to correct the problem. Say you will change your behavior OR come up with a way to compromise, if appropriate.

8. (M) Say something to reaffirm your genuine interest in the class (e.g., say how much you value the subject matter, etc.).

9. (M) If necessary, change your behavior to correct the problem (e.g., study harder, cut out extra-curricular and/or social actitivites, get a tutor, see TA's, do extra work, etc.)

10. (M) If talking to the professor and taking steps to change your behavior does not improve the situation, then seek guidance from an appropriate academic/administrative figure (e.g., an academic advisor or dean). (This only counts as a model response if most or all of the above solutions have been attempted first.)

Examples of NON-model solutions:

6. Ruminate about the meanings and consequences of how the professor is treating you. (For example, mention self-doubts, self-pity, negative expectations for the future, negative self-evaluations (e.g., think about changing your career or major because you do not think that you can make it).

3. Mention negative affect (e.g., worry, get upset, cry, panic, feel frustrated, say obscenities, etc.).

4. Talk to someone not associated with the class (e.g., parents, friends, s.o.).

9. Use self-talk and/or self-boosting (e.g., tell yourself you can do it or that you are determined to pass OR think that you have confidence in yourself, etc.), OR pray, OR try NOT to panic or get upset. (That is, change something psychological; change your thinking rather than your behavior.)

10. Consider other options (e.g., dropping the class OR changing your major OR getting a summer job, etc.).

11. Reject help OR cheat OR try to manipulate the professor or TA(s).

Short Stress State Questionnaire (SSSQ; Helton, 2004)

General Instructions
This questionnaire is concerned with your feelings and thoughts while you were performing the task. Please answer every question, even if you find it difficult. Answer, as honestly as you can, what is true of you. Please do not choose a reply just because it seems like the 'right thing to say'. Your answers will be kept entirely confidential. Also, be sure to answer according to how you felt **WHILE PERFORMING THE TASK**. Don't just put down how you usually feel. You should try and work quite quickly: there is no need to think very hard about the answers. The first answer you think of is usually the best.

Please indicate how well each word describes how you felt DURING THE TASK.

Not at all = 1   A little bit = 2   Somewhat = 3   Very much = 4   Extremely = 5

1. Dissatisfied	1	2	3	4	5
2. Alert	1	2	3	4	5
3. Depressed	1	2	3	4	5
4. Sad	1	2	3	4	5
5. Active	1	2	3	4	5
6. Impatient	1	2	3	4	5
7. Annoyed	1	2	3	4	5
8. Angry	1	2	3	4	5
9. Irritated	1	2	3	4	5
1. Grouchy	1	2	3	4	5

Please indicate how true each statement was of your thoughts **WHILE PERFORMING THE TASK**.

Not at all = 1   A little bit = 2   Somewhat = 3   Very much = 4   Extremely = 5

11. I was committed to attaining my performance goals — 1 2 3 4 5

12. I wanted to succeed on the task — 1 2 3 4 5

13. I was motivated to do the task — 1 2 3 4 5

14. I tried to figure myself out. — 1 2 3 4 5

15. I reflected about myself. — 1 2 3 4 5

16. I daydreamed about myself. — 1 2 3 4 5

17. I felt confident about my abilities. — 1 2 3 4 5

18. I felt self-conscious. — 1 2 3 4 5

19. I was worried about what other people think of me. — 1 2 3 4 5

2. I felt concerned about the impression I was making. — 1 2 3 4 5

21. I performed proficiently on this task. — 1 2 3 4 5

22. Generally, I felt in control of things. — 1 2 3 4 5

23. I thought about how others have done on this task. — 1 2 3 4 5

24. I thought about how I would feel if I were told how I performed. — 1 2 3 4 5

Sample Negative Feedback Print-out

Examinee ID #: 667
Test Date: 11/3/2014

CAI Aptitude Test Scores*:

Reading Comprehension Test

Subscale	Scaled Score**	Percentile Rank	Qualitative Description
Analytic reasoning ability	11	53%	Average
Inferential reasoning ability	7	31%	Below Average
Comprehension Total	**8**	**42%**	Low Average

Creative Ability Test

Subscale	Scaled Score	Percentile Rank	Qualitative Description
Fluency	9	48%	Average
Plausibility	8	41%	Low Average
Uniqueness	6	29%	Below Average
Creative Ability Total	**7**	**36%**	Below Average

Problem Solving Test

Subscale	Scaled Score	Percentile Rank	Qualitative Description
Clarity			
Causal Coherence			
Effectiveness			
Problem Solving Total			

*Scores are normed against a sample of 358 undergraduate students at *University of Virginia.*
**Subtest Scaled Scores have a mean of 10, with scores ranging from 8-12 falling in the Average range (Sattler, 2012).

Rumination Induction Task

Instructions: Please type your initial response to each of the items presented on the following screens, while using your imagination and concentration to think about the causes, meanings, and consequences of the item. You will be given several moments to concentrate on each of the items once you have made your response, so please use that time to try to make sense of and understand the issues raised by the prompt and by your response to it.

Prompts (presented in random order):

What are your feelings right now and why are you feeling this way?

What are the physical sensations in your body right now?

Why do you react to things the way you do?

What kind of person are you?

What do your emotions say about you?

What is your current level of energy?

Subjective Units of Distress (SUDS) Rating Sheet

Participant ID# _____

Date _____

Emotion Ratings. Instructions:

At several points throughout this study, the experimenter will ask you to rate the level of emotional distress (i.e., feelings of sadness and/or anxiety) you are currently experiencing on a 0-100 scale, where **0 is absolutely no distress, and 100 is extremely distressed.** Do you have any questions about how to use that scale? (If so, please ask the experimenter now.)

Emotion rating 1.
> Using the scale from 0-100, how distressed are you feeling right now? _____

Emotion Rating 2.
> Using the scale from 0-100, how distressed are you feeling right now? _____

Emotion Rating 3.
> Using the scale from 0-100, how distressed are you feeling right now? _____

Emotion Rating 4.
> Using the scale from 0-100, how distressed are you feeling right now? _____

Emotion Rating 5.
> Using the scale from 0-100, how distressed are you feeling right now? _____

Emotion Rating 6.
> Using the scale from 0-100, how distressed are you feeling right now? _____

Emotion Rating 7.
> Using the scale from 0-100, how distressed are you feeling right now? _____

Emotion Rating 8.
> Using the scale from 0-100, how distressed are you feeling right now? _____

Emotion Rating 9.
> Using the scale from 0-100, how distressed are you feeling right now? _____

Emotion Rating 10.
> Using the scale from 0-100, how distressed are you feeling right now? _____

Funnel Debriefing Interview Form

1. Do you have any general comments about the study and what you did here?

2. What did you think the point of the study was?

3. Did any part of the study seem strange to you?

    Y   N

4. What about the instructions you were given for the thinking exercises or the academic aptitude tests?

    Y   N

5. What about the feedback you were given about your performance on the first set of academic aptitude tests?

    Y   N

6. *QUERY FOR ALL YES ANSWERS*: You mentioned that you thought…
    • When would you say you started to notice something strange? Was it during the experiment or after it was over and you were thinking back on it?
    • What exactly did you think?
    • What caused you to think that?
    • How do you think it affected your answers, if at all?

7. Did you hear anything about this study before coming in today? *IF YES*: What did you hear?

Other experimenter comments

Appendix E
Factor Analyses to Derive SSSQ Rumination Subscale

Exploratory factor analysis (EFA).

An EFA was first run on all 24 items of the pre-training (Test Period 1) Short Stress State

Questionnaire (SSSQ; Helton, 2004) to examine and compare individual factor loadings. Based

on the results of the scree plot, six factors were initially included in the model (which was run

via the "psych" package in R, using maximum likelihood with a "promax" rotation). After

removing items that did not load highly (<.5) on any factor or loaded highly (>.5) on multiple

factors, a 2-factor solution emerged as the best-fitting model, with a 4-item "Rumination" (i.e.,

negative self-focus) subscale that explained 34% of the total variance, followed by a 3-item

"Motivation" factor that explained 29% of the total variance (for a total of 63% of variance

explained). See Table 1 (below) for factor loadings at Test Periods 1 and 2.

Table 1. Factor loadings on "Rumination" (and "Motivation") subscales for each item retained in
abbreviated, 7-item SSSQ measure.

	Test Period 1		Test Period 2	
Item	**Factor1: Rumination**	Factor2: Motivation	**Factor1: Rumination**	Factor2: Motivation
1. I was committed to attaining my performance goals.	-.01	**.80**	-.03	**.89**
2. I wanted to succeed on the task.	.04	**.91**	.02	**.88**
3. I was motivated to do the task.	-.07	**.77**	-.04	**.75**
4. I felt self-conscious.	**.79**	.01	**.74**	.04
5. I worried about what other people think of me.	**.91**	-.04	**.98**	-.05
6. I felt concerned about the impression I was making.	**.73**	.05	**.77**	.02
7. I thought about how others have done on this task.	**.59**	.02	**.61**	.02

Confirmatory factor analysis (CFA).

A CFA was then conducted, using the "semTools" package in R (Pornprasertmanit,

Miller, Schoemann, & Rosseel, 2014), to test both overall model fit and longitudinal factor

invariance across the two test periods. Specifically, we tested the assumption of "strong" factor invariance, defined as a minimal loss of model fit when the Test Period 1 and Test Period 2 factor models are constrained to have the same item loadings and intercepts. Following Little (2013), we set a ΔCFI < .01 as the criterion for factor invariance. See Table 2 (below) for model fit and longitudinal invariance statistics for each factor.

Table 2. Model fit statistics and factor invariance of the SSSQ "Rumination" (2a) and "Motivation" (2b) subscales.

2a. "Rumination" factor:

Model 1: Configural invariance (overall model fit):					
χ^2	df	*p*	CFI	RMSEA	BIC
23.58	15	.073	.99	.05	5241.56
Model 2: Strong invariance (equal loadings + intercepts):					
χ^2	df	*p*	CFI	RMSEA	BIC
23.58	14	.051	.99	.05	5247.06
Model 1 versus Model 2:					
$\Delta\chi^2$	Δdf	**ΔCFI**			
.00	1	**-.001**			

2b. "Motivation" factor:

Model 1: Configural invariance (overall model fit):					
χ^2	df	*p*	CFI	RMSEA	**BIC**
4.00	5	.550	1	.00	3272.03
Model 2: Strong invariance (equal loadings + intercepts):					
χ^2	df	*p*	CFI	RMSEA	BIC
4.00	4	.406	1	.00	3277.54
Model 1 versus model 2:					
$\Delta\chi^2$	Δdf	**ΔCFI**			
.00	1	**.000**			

Appendix F
Supplementary regression statistics for training effects

Table F1. Omnibus regression statistics for the effects of Condition (How-Only, Why-Only, Combined, Control) x Time (Test Period 1, Test Period 2) on Cognitive Performance across 4 Test Types (Reading Comprehension, Fluency, Creativity, Problem-Solving).

	Numerator DF	Demoninator DF	F	p
Training Condition	3	209.45	.27	.845
Test Period *(positive main effect; increase from pre- to post-training)*[***]	1	1465.03	17.28	<.001
Test Type	3	1457.88	.01	.999
Average SAT (covariate)[***]	1	21.77	13.29	<.001
Running Span Score (covariate)[*]	1	21.93	5.92	.016
Training Condition x Test Period	3	1464.96	1.58	.193
Training Condition x Test Type[*]	9	1457.89	2.01	.035
Test Period x Test Type	3	1454.66	2.06	.103
Training Condition x Test Period x Test Type[*]	**9**	**1454.65**	**1.92**	**.045**

Table F2. Omnibus multivariate regression statistics for the effects of Condition (How-Only, Why-Only, Combined, Control) X Time (Test Period 1, Test Period 2) on Off-task Rumination Frequency across 2 Episode Types (Probe-Caught, Self-Caught).

	Numerator DF	Demoninator DF	F	p
Training Condition	3	197.20	.71	.545
Test Period[***] *(negative main effect; decrease from pre- to post-training)*	1	572.28	6.86	.009
Episode Type	1	566.12	1.34	.247
Average SAT (covariate)	1	198.64	.08	.781
Running Span Score (covariate)	1	197.05	.00	.994
Training Condition x Episode Type	3	566.11	2.23	.084
Training Condition x Test Period	3	572.32	1.70	.167
Test Period x Episode Type[***]	1	564.43	11.11	.001
Training Condition x Test Period x Episode Type	3	564.43	1.07	.363

Table F3. Omnibus multivariate regression statistics for the effects of Condition (How-Only, Why-Only, Combined, Control) X Time (Test Period 1, Test Period 2) on Negative Affect during Testing (SUDS-Distress) across 2 mid-test-battery Timepoints (post-Test1, post-Test2).

	Numerator DF	Demoninator DF	F	p
Training Condition[*]	1	207.47	4.87	.028
Test Period	1	207.27	.12	.725
Timepoint	3	207.18	.75	.524
Average SAT (covariate)[*]	1	621.27	4.90	.027
Running Span Score (covariate)	1	62.33	.05	.825
Training Condition x Test Period	3	621.27	1.91	.127
Training Condition x Timepoint[†]	3	62.33	2.20	.087
Test Period x Timepoint[*]	1	62.32	4.04	.045
Training Condition x Test Period x Timepoint	3	62.32	.64	.590

Table F4. Planned contrast statistics for non-significant moderation effects predicting Cognitive Performance.

Reading Comprehension	Unstandardized b	Standardized b	SE	z	p	Uncorrected p
Depression Symptoms (DASS-D) as moderator						
Condition (Why-Only vs. Combined) x Time x DASS-D	-.55	-.34	.26	-1.32	.264	.109
Condition (How-Only vs. Combined) x Time x DASS-D	-.45	-.27	.23	-1.22	.258	.106
Condition (Control vs. Combined) x Time x DASS-D	-.52	-.32	.23	-1.38	.392	.174
Condition (How-Only vs. Why-Only) x Time x DASS-D	.11	.07	.25	.27	.985	.791
Condition (Control vs. Why-Only) x Time x DASS-D	.03	.02	.25	.07	1.000	.943
Condition (How-Only vs. Control) x Time x DASS-D	.08	.05	.22	.21	.994	.832
Anxiety Symptoms (DASS-A) as moderator						
Condition (Why-Only vs. Combined) x Time x DASS-A	-.81	-.50	.26	-1.94	.133	.054
Condition (How-Only vs. Combined) x Time x DASS-A	-.63	-.39	.23	-1.70	.223	.090
Condition (Control vs. Combined) x Time x DASS-A	-.61	-.38	.24	-1.58	.279	.116
Condition (How-Only vs. Why-Only) x Time x DASS-A	.17	.11	.24	.44	.943	.662
Condition (Control vs. Why-Only) x Time x DASS-A	.19	.12	.25	.47	.930	.638
Condition (How-Only vs. Control) x Time x DASS-A	-.02	-.01	.23	-.06	.999	.955

Working Memory (Running Span) as moderator						
Condition (Why-Only vs. Combined) x Time x RunningSpan	-.03	-.02	.27	-.06	1.000	.951
Condition (How-Only vs. Combined) x Time x RunningSpan	.13	.08	.23	.36	.973	.723
Condition (Control vs. Combined) x Time x RunningSpan	.15	.09	.22	.41	.959	.682
Condition (How-Only vs. Why-Only) x Time x RunningSpan	.10	.10	.27	.37	.960	.711
Condition (Control vs. Why-Only) x Time x RunningSpan	.11	.11	.26	.42	.944	.677
Condition (How-Only vs. Control) x Time x RunningSpan	-.01	-.01	.23	-.04	1.000	.968

Creative Quality	*Unstandardized b*	*Standardized b*	*SE*	*z*	*p*	*Uncorrected p*
Trait Rumination (RRS) as moderator						
Condition (Why-Only vs. Combined) x Time x RRS	-.02	-.07	.22	-.33	.974	.739
Condition (How-Only vs. Combined) x Time x RRS	.08	.32	.21	1.55	.284	.123
Condition (Control vs. Combined) x Time x RRS	.06	.24	.22	1.06	.579	.290
Condition (How-Only vs. Why-Only) x Time x RRS	.10	.39	.20	1.97	.132	.050
Condition (Control vs. Why-Only) x Time x RRS	.08	.31	.22	1.43	.355	.153
Condition (How-Only vs. Control) x Time x RRS	.02	.09	.20	.42	.954	.677
Depression Symptoms (DASS-D) as moderator						
Condition (Why-Only vs. Combined) x Time x DASS-D	-.01	-.02	.23	-.11	.999	.912
Condition (How-Only vs. Combined) x Time x DASS-D	.09	.37	.20	1.83	.179	.069
Condition (Control vs. Combined) x Time x DASS-D	.09	.36	.22	1.63	.257	.104
Condition (How-Only vs. Why-Only) x Time x DASS-D	.10	.40	.21	1.84	.166	.067
Condition (Control vs. Why-Only) x Time x DASS-D	.09	.39	.23	1.66	.232	.099
Condition (How-Only vs. Control) x Time x DASS-D	.00	.01	.21	.05	1.000	.964
Anxiety Symptoms (DASS-A) as moderator (SEE FIGURE F1 BELOW)						
Condition (Why-Only vs. Combined) x Time x DASS-A	.02	.07	.23	.32	.979	.749
Condition (How-Only vs. Combined) x Time x DASS-A	**.12**	**.46**	**.21**	**2.22**	**.077**	**.027**
Condition (Control vs. Combined) x Time x DASS-A	**.14**	**.55**	**.24**	**2.32**	**.056**	**.021**
Condition (How-Only vs. Why-Only) x Time x DASS-A	**.10**	**.39**	**.21**	**1.81**	**.182**	**.073**
Condition (Control vs. Why-Only) x Time x DASS-A	**.12**	**.47**	**.24**	**1.96**	**.133**	**.051**
Condition (How-Only vs. Control) x Time x DASS-A	-.02	-.09	.22	-.40	.957	.694
Working Memory (Running Span) as moderator						
Condition (Why-Only vs. Combined) x Time x RunningSpan	-.02	.07	.23	.32	.981	.753
Condition (How-Only vs. Combined) x Time x RunningSpan	-.02	-.08	.21	-.40	.962	.690
Condition (Control vs. Combined) x Time x RunningSpan	-.01	-.03	.20	-.15	.998	.879

	Unstandardized b	Standardized b	SE	z	p	Uncorrected p
Condition (How-Only vs. Why-Only) x Time x RunningSpan	-.04	-.16	.23	-.67	.822	.503
Condition (Control vs. Why-Only) x Time x RunningSpan	-.03	-.10	.23	-.46	.932	.648
Condition (How-Only vs. Control) x Time x RunningSpan	-.01	-.05	.20	-.26	.989	.795
Creative Fluency	*Unstandardized b*	*Standardized b*	*SE*	*z*	*p*	*Uncorrected p*
Trait Rumination (RRS) as moderator						
Condition (Why-Only vs. Combined) x Time x RRS	.19	.02	.15	.16	.997	.876
Condition (How-Only vs. Combined) x Time x RRS	1.03	.13	.14	.90	.687	.369
Condition (Control vs. Combined) x Time x RRS	.49	.06	.16	.40	.957	.690
Condition (How-Only vs. Why-Only) x Time x RRS	.84	.11	.14	.77	.787	.442
Condition (Control vs. Why-Only) x Time x RRS	.30	.04	.15	.26	.989	.798
Condition (How-Only vs. Control) x Time x RRS	.54	.07	.14	.48	.932	.631
Depression Symptoms (DASS-D) as moderator						
Condition (Why-Only vs. Combined) x Time x DASS-D	-.56	-.07	.16	-.45	.946	.654
Condition (How-Only vs. Combined) x Time x DASS-D	.73	.09	.14	.65	.857	.515
Condition (Control vs. Combined) x Time x DASS-D	-1.00	-.13	.15	-.82	.758	.414
Condition (How-Only vs. Why-Only) x Time x DASS-D	1.29	.16	.15	1.09	.553	.276
Condition (Control vs. Why-Only) x Time x DASS-D	-.44	-.06	.16	-.35	.971	.730
Condition (How-Only vs. Control) x Time x DASS-D	1.73	.22	.15	1.50	.312	.136
Anxiety Symptoms (DASS-A) as moderator						
Condition (Why-Only vs. Combined) x Time x DASS-A	-1.83	-.23	.16	-1.44	.352	.151
Condition (How-Only vs. Combined) x Time x DASS-A	-.85	-.11	.15	-.74	.810	.463
Condition (Control vs. Combined) x Time x DASS-A	-1.33	-.17	.17	-1.00	.632	.317
Condition (How-Only vs. Why-Only) x Time x DASS-A	.98	.12	.15	.82	.748	.411
Condition (Control vs. Why-Only) x Time x DASS-A	.50	.06	.17	.37	.967	.712
Condition (How-Only vs. Control) x Time x DASS-A	.48	.06	.16	.38	.960	.702
Working Memory (Running Span) as moderator						
Condition (Why-Only vs. Combined) x Time x RunningSpan	.66	.08	.16	.52	.924	.607
Condition (How-Only vs. Combined) x Time x RunningSpan	-.44	-.06	.14	-.39	.965	.700
Condition (Control vs. Combined) x Time x RunningSpan	-.72	-.09	.14	-.65	.860	.514
Condition (How-Only vs. Why-Only) x Time x RunningSpan	-1.09	-.14	.16	-.86	.697	.390
Condition (Control vs. Why-Only) x Time x RunningSpan	-1.38	-.17	.16	-1.11	.524	.268
Condition (How-Only vs. Control) x Time x RunningSpan	.28	-.04	.14	-.26	.989	.798
MEPS Problem-Solving Task	*Unstandardized b*	*Standardized b*	*SE*	*z*	*p*	*Uncorrected p*
Trait Rumination (RRS) as moderator						
Condition (Why-Only vs. Combined) x Time x RRS	-.31	-.13	.22	-.58	.886	.561

	Unstandardized b	Standardized b	SE	z	p	Uncorrected p
Condition (How-Only vs. Combined) x Time x RRS	.06	.03	.21	.13	.999	.901
Condition (Control vs. Combined) x Time x RRS	.05	.02	.22	.10	.999	.925
Condition (How-Only vs. Why-Only) x Time x RRS	.38	.16	.21	.75	.797	.453
Condition (Control vs. Why-Only) x Time x RRS	.37	.15	.22	.70	.828	.486
Condition (How-Only vs. Control) x Time x RRS	.01	.01	.21	.03	1.000	.978
Depression Symptoms (DASS-D) as moderator						
Condition (Why-Only vs. Combined) x Time x DASS-D	-.30	-.13	.23	-.54	.910	.590
Condition (How-Only vs. Combined) x Time x DASS-D	-.17	-.07	.21	-.33	.977	.741
Condition (Control vs. Combined) x Time x DASS-D	.00	.00	.22	.00	1.000	.999
Condition (How-Only vs. Why-Only) x Time x DASS-D	.13	.06	.22	.25	.989	.807
Condition (Control vs. Why-Only) x Time x DASS-D	.30	.13	.23	.55	.898	.582
Condition (How-Only vs. Control) x Time x DASS-D	-.17	-.07	.21	-.34	.976	.736
Anxiety Symptoms (DASS-A) as moderator						
Condition (Why-Only vs. Combined) x Time x DASS-A	-.74	-.31	.24	-1.28	.445	.201
Condition (How-Only vs. Combined) x Time x DASS-A	-.09	-.04	.22	-.16	.997	.870
Condition (Control vs. Combined) x Time x DASS-A	.14	.06	.23	.26	.988	.793
Condition (How-Only vs. Why-Only) x Time x DASS-A	.65	.27	.24	1.15	.513	.250
Condition (Control vs. Why-Only) x Time x DASS-A	.89	.37	.24	1.55	.283	.123
Condition (How-Only vs. Control) x Time x DASS-A	-.23	-.10	.22	-.44	.951	.663
Working Memory (Running Span) as moderator						
Condition (Why-Only vs. Combined) x Time x RunningSpan	-.56	-.23	.24	-.96	.671	.338
Condition (How-Only vs. Combined) x Time x RunningSpan	-.79	-.33	.22	-1.50	.326	.135
Condition (Control vs. Combined) x Time x RunningSpan	-.43	-.18	.21	-.85	.746	.398
Condition (How-Only vs. Why-Only) x Time x RunningSpan	-.23	-.10	.25	-.39	.957	.701
Condition (Control vs. Why-Only) x Time x RunningSpan	.14	.06	.24	.24	.988	.809
Condition (How-Only vs. Control) x Time x RunningSpan	-.37	-.15	.21	-.72	.834	.474

Table F5. Planned contrast statistics for non-significant moderation effects predicting Off-task Rumination Frequency.

Off-task Rumination Frequency (via thought tracking)	*Unstandardized b*	*Standardized b*	*SE*	*z*	*p*	*Uncorrected p*
Trait Rumination (RRS) as moderator						
Condition (Why-Only vs. Combined) x Time x RRS	.04	.06	.16	.38	.966	.672
Condition (How-Only vs. Combined) x Time x RRS	.00	.02	.15	.13	.998	.998
Condition (Control vs. Combined) x Time x RRS	.07	.12	.15	.80	.768	.431
Condition (How-Only vs. Why-Only) x Time x RRS	-.04	-.04	.15	-.26	.988	.663

	Unstandardized b	Standardized b	SE	z	p	Uncorrected p
Condition (Control vs. Why-Only) x Time x RRS	.03	.06	.15	.41	.955	.719
Condition (How-Only vs. Control) x Time x RRS	-.07	-.10	.15	-.69	.836	.415
Depression Symptoms (DASS-D) as moderator						
Condition (Why-Only vs. Combined) x Time x DASS-D	-.10	-.19	.17	-1.13	.538	.306
Condition (How-Only vs. Combined) x Time x DASS-D	-.13	-.23	.16	-1.48	.323	.124
Condition (Control vs. Combined) x Time x DASS-D	.01	.02	.16	.13	.999	.915
Condition (How-Only vs. Why-Only) x Time x DASS-D	-.04	-.04	.16	-.24	.990	.688
Condition (Control vs. Why-Only) x Time x DASS-D	.11	.21	.16	1.29	.423	.246
Condition (How-Only vs. Control) x Time x DASS-D	-.14	-.25	.15	-1.68	.240	.088
Anxiety Symptoms (DASS-A) as moderator						
Condition (Why-Only vs. Combined) x Time x DASS-A	.03	.05	.17	.30	.982	.730
Condition (How-Only vs. Combined) x Time x DASS-A	-.03	-.05	.16	-.33	.977	.723
Condition (Control vs. Combined) x Time x DASS-A	-.12	-.20	.16	-1.22	.481	.195
Condition (How-Only vs. Why-Only) x Time x DASS-A	-.06	-.10	.16	-.64	.857	.479
Condition (Control vs. Why-Only) x Time x DASS-A	-.15	-.25	.17	-1.49	.315	.108
Condition (How-Only vs. Control) x Time x DASS-A	.09	.15	.16	.97	.656	.307
Working Memory (Running Span) as moderator						
Condition (Why-Only vs. Combined) x Time x RunningSpan	-.12	-.22	.17	-1.28	.460	.215
Condition (How-Only vs. Combined) x Time x RunningSpan	-.09	-.17	.16	-1.12	.562	.297
Condition (Control vs. Combined) x Time x RunningSpan	-.16	-.31	.15	-2.10	.105	.054
Condition (How-Only vs. Why-Only) x Time x RunningSpan	.03	.04	.17	.25	.988	.779
Condition (Control vs. Why-Only) x Time x RunningSpan	-.04	-.09	.17	-.54	.898	.672
Condition (How-Only vs. Control) x Time x RunningSpan	.07	.13	.15	.87	.739	.430

Table F6. Planned contrast statistics for non-significant moderation effects predicting Retrospective State Rumination.

Retrospective State Rumination (via SSSQ Rumination)	*Unstandardized b*	*Standardized b*	*SE*	*z*	*p*	*Uncorrected p*
Trait Rumination (RRS) as moderator						
Condition (Why-Only vs. Combined) x Time x RRS	.10	.10	.16	.62	.865	.507
Condition (How-Only vs. Combined) x Time x RRS	.02	.02	.15	.11	.999	.914
Condition (Control vs. Combined) x Time x RRS	.20	.20	.16	1.27	.443	.207
Condition (How-Only vs. Why-Only) x Time x RRS	-.09	-.08	.14	-.56	.901	.546
Condition (Control vs. Why-Only) x Time x RRS	.10	.10	.15	.67	.849	.536
Condition (How-Only vs. Control) x Time x RRS	-.18	-.18	.14	-1.25	.460	.213
Anxiety Symptoms (DASS-A) as moderator						

Condition (Why-Only vs. Combined) x Time x DASS-A	.30	.28	.16	1.72	.215	.077
Condition (How-Only vs. Combined) x Time x DASS-A	.08	.07	.15	.48	.932	.602
Condition (Control vs. Combined) x Time x DASS-A	.33	.32	.16	2.02	.119	.042
Condition (How-Only vs. Why-Only) x Time x DASS-A	-.22	-.21	.16	-1.33	.399	.173
Condition (Control vs. Why-Only) x Time x DASS-A	.03	.03	.17	.20	.994	.849
Condition (How-Only vs. Control) x Time x DASS-A	.24	-.24	.15	-1.62	.263	.101
Working Memory (Running Span) as moderator						
Condition (Why-Only vs. Combined) x Time x RunningSpan	-.18	-.18	.17	-1.06	.603	.288
Condition (How-Only vs. Combined) x Time x RunningSpan	.01	.01	.15	.06	1.000	.972
Condition (Control vs. Combined) x Time x RunningSpan	-.10	-.10	.14	-.73	.820	.471
Condition (How-Only vs. Why-Only) x Time x RunningSpan	.19	.19	.17	1.10	.525	.279
Condition (Control vs. Why-Only) x Time x RunningSpan	.08	.07	.16	.45	.934	.646
Condition (How-Only vs. Control) x Time x RunningSpan	-.11	-.11	.14	-.78	.785	.456

Figure F1. Simple slopes for the Condition x Test Period x Anxiety Symptoms interaction predicting Creativity.
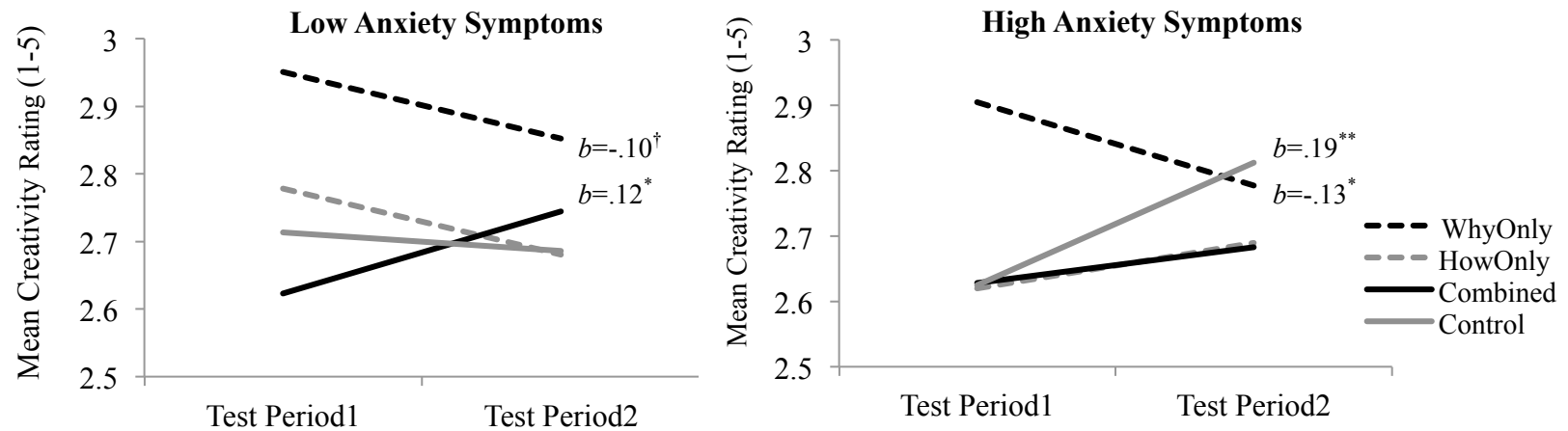
Note. Low and High Anxiety Symptoms reflect model-predicted values at -1 and +1SD from mean of Depression Anxiety Stress Scales – Anxiety (DASS-A) subscale. Unstandardized betas (b) are shown only for significant or marginally significant simple slopes of Time (i.e., significant increases or decreases in Creativity from Test Period1 to Test Period2).

Secondary Analyses: Task Motivation and Self-Confidence

Given our theoretical interest in motivation and self-confidence as potential target mechanisms of our "why"-focused and "how"-focused training prompts, respectively, we conducted secondary planned contrast analyses to test whether there were training effects on 1) Task Motivation, as indexed by the SSSQ Motivation subscale derived from our factor analyses (see Appendix D); and 2) Self-Confidence, as indexed by the single SSSQ "confidence" item ("I felt confident about my abilities"). We followed the same analytic approach as for our primary outcome measures, with mixed-effects regression models estimating each Condition x Test Period interaction effect, both overall and as moderated by each baseline individual difference measure.

There were no overall main effects of Test Period for either outcome (both $p > .05$). As shown in Table F7 (below), the only training effect that reached significance was a 3-way Condition (HowOnly vs WhyOnly) x Test Period x DASS-A interaction ($p = .029$) predicting Task Motivation, such that, among more highly anxious participants, only the *WhyOnly* condition led to a marginal increase in motivation from pre- to post-training, in line with hypotheses. By contrast, among less anxious participants, only the *HowOnly* condition led to a marginal increase in motivation (see Figure F2 below). This pattern was replicated for the trait rumination moderator, though paired contrasts did not quite reach significance after correcting for multiple comparisons.

Table F7. Secondary outcome analyses: Planned contrasts to test hypothesized effects of Condition x Test Period for Motivation and Self-Confidence.

Task Motivation (via SSSQ Motivation subscale)	*Unstandardized b*	*Standardized b*	*SE*	*z*	*p*
Condition (Why-Only vs. Combined) x Time	.08	.09	.15	.58	.891
Condition (How-Only vs. Combined) x Time	.08	.09	.15	.59	.887
Condition (Control vs. Combined) x Time	.07	.08	.16	.50	.928
Condition (How-Only vs. Why-Only) x Time	.00	.00	.15	.01	1.000
Condition (Control vs. Why-Only) x Time	-.01	-.01	.15	-.08	1.000
Condition (How-Only vs. Control) x Time	.01	.01	.15	.09	1.000
Anxiety Symptoms (DASS-A) as moderator					
Condition (Why-Only vs. Combined) x Time x DASS-A	.07	.09	.17	.50	.927
Condition (How-Only vs. Combined) x Time x DASS-A	-.29	-.34	.16	-2.13	.093
Condition (Control vs. Combined) x Time x DASS-A	-.18	-.21	.16	-1.26	.454
Condition (How-Only vs. Why-Only) x Time x DASS-A	**-.36**	**-.42**	**.16**	**-2.58**	**.029**
Condition (Control vs. Why-Only) x Time x DASS-A	-.25	-.29	.17	-1.73	.206
Condition (How-Only vs. Control) x Time x DASS-A	-.11	-.13	.15	-.83	.750
Trait Rumination (RRS) as moderator					
Condition (Why-Only vs. Combined) x Time x RRS	.24	.28	.16	1.74	.203
Condition (How-Only vs. Combined) x Time x RRS	.04	.05	.16	.33	.975
Condition (Control vs. Combined) x Time x RRS	-.07	-.08	.16	-.49	.925
Condition (How-Only vs. Why-Only) x Time x RRS	-.20	-.23	.15	-1.53	.303
Condition (Control vs. Why-Only) x Time x RRS	-.32	-.37	.16	-2.28	.066
Condition (How-Only vs. Control) x Time x RRS	.11	.13	.15	.87	.721
Self-Confidence (via SSSQ "confidence" item)	*Unstandardized b*	*Standardized b*	*SE*	*z*	*p*
Condition (Why-Only vs. Combined) x Time	-.10	-.11	.18	-.59	.887
Condition (How-Only vs. Combined) x Time	.04	.04	.18	.23	.992
Condition (Control vs. Combined) x Time	.03	.03	.19	.17	.997
Condition (How-Only vs. Why-Only) x Time	.14	.15	.18	.83	.752
Condition (Control vs. Why-Only) x Time	.13	.14	.18	.76	.798
Condition (How-Only vs. Control) x Time	.01	.01	.18	.06	1.000

†p<.10; *p<.05; **p<.01; ***p<.001

Note. SSSQ=Short Stress State Questionnaire; DASS-A=Depression Anxiety Stress Scales - Anxiety subscale All continuous predictors and moderators were standardized for these analyses. Unstandardized betas represent the slope of change in the outcome variable for every 1 SD increase in the predictor. Standard errors (SEs) correspond to *standardized* beta estimates.

Figure F2. Simple slopes for the Condition x Test Period x Anxiety Symptoms interaction predicting State Motivation (SSSQ Motivation subscale).
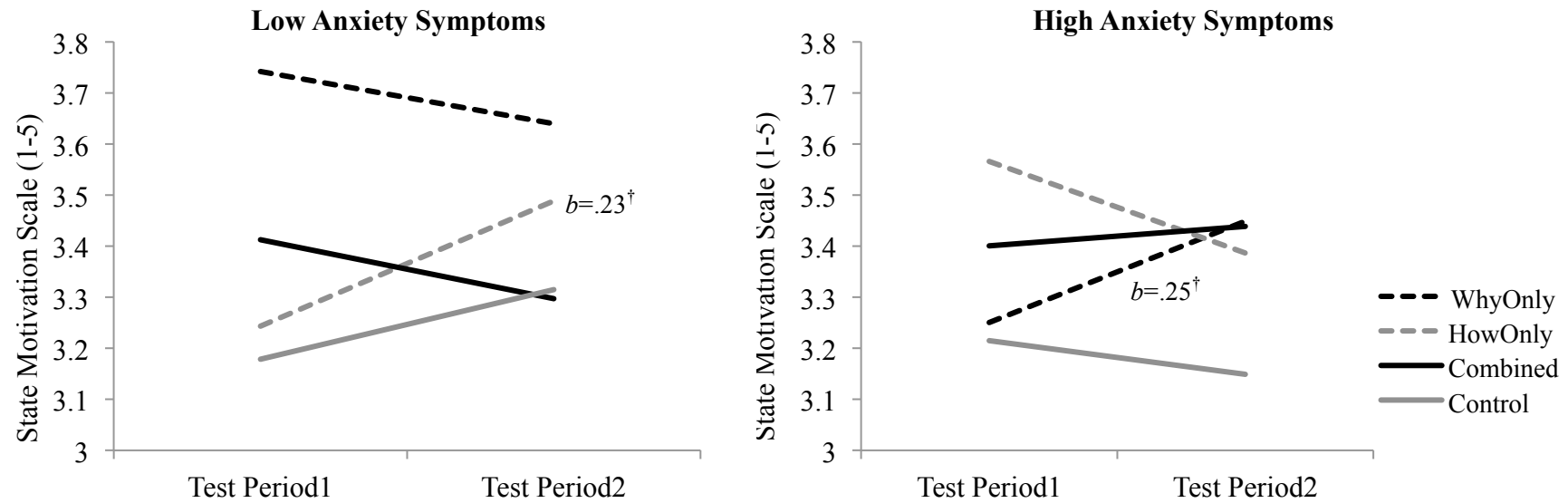
Note. Low and High Anxiety Symptoms reflect model-predicted values at -1 and +1SD from the mean of the Depression Anxiety Stress Scales – Anxiety (DASS-A) subscale. Unstandardized betas (*b*) are shown only for significant or marginally significant simple slopes of Test Period, and reflect the estimated change in State Motivation (on a 1-5 scale, where 1=Not at all and 5=Extremely) from pre- to post-training. The simple slopes for Low versus High Trait Rumination (Ruminative Responses Scale) follow a similar pattern to the one displayed above.

Appendix G:
Hypothesized Mediation Models