Journal of Personality and Social Psychology
2004, Vol. 87, No. 1, 133–145

0022-3514/04/$12.00 DOI: 10.1037/0022-3514.87.1.133

The Relationship Between Self-Esteem Level, Self-Esteem Stability, and Cardiovascular Reactions to Performance Feedback

Mark D. Seery, Jim Blascovich, Max Weisbuch, and S. Brooke Vick
University of California, Santa Barbara

The authors examined the notion that individuals with unstable high self-esteem possess implicit self-doubt. They adopted the framework of the biopsychosocial model of challenge and threat and assessed spontaneous cardiovascular reactions in the face of success versus failure performance feedback. Study 1 revealed predicted interactions between feedback condition, self-esteem level, and self-esteem stability, such that participants with unstable high self-esteem exhibited relative threat (a negative reaction) in the failure condition, whereas those with stable high self-esteem exhibited relative challenge (a positive reaction). Study 2 replicated these results and provided additional evidence against plausible alternative explanations.

Over the past several years, interest in the apparently contradictory nature of high self-esteem has grown. High self-esteem has been associated with resilience and optimal functioning, on the one hand, and with self-aggrandizement and defensiveness, on the other (e.g., Baumeister, Smart, & Boden, 1996; Kernis, 2003). Kernis and colleagues (for reviews, see Kernis, 1993; Kernis & Waschull, 1995) have addressed this controversy by examining self-esteem stability along with the typically assessed self-esteem level (high vs. low). The existing body of research suggests that stable high self-esteem is associated with resilience in the face of negative self-relevant events (i.e., failure feedback), whereas unstable high self-esteem is associated with the use of outwardly defensive strategies. In the present investigation, we directly tested the possibility that individuals with unstable high self-esteem possess underlying self-doubt, which presumably motivates such strategies. We accomplished this by assessing individuals' implicit, covert responses with online psychophysiological measures.

Self-Esteem Level and Stability

Kernis and colleagues (e.g., Kernis, 1993; Kernis & Waschull, 1995) have repeatedly demonstrated the utility of assessing self-esteem stability along with self-esteem level. They describe self-esteem stability as the magnitude of short-term fluctuations around a baseline level of self-esteem. Kernis and colleagues have assessed stability of self-esteem over varying time periods, but the basic procedure has remained the same. Participants complete state self-esteem scales at regular intervals (e.g., 12-hr periods). The items on the scale are worded so that they ask participants how they feel at that moment, rather than in general, as is done in trait self-esteem scales. Calculating the standard deviation of the multiple state self-esteem scores yields an index of stability, such that a person with unstable self-esteem exhibits greater shifts and, consequently, a higher standard deviation.

Relative to stable high self-esteem, unstable high self-esteem has been linked to greater defensiveness in a number of contexts. Kernis, Grannemann, and Barclay (1989) found that persons with unstable high self-esteem had the highest propensity for anger, as assessed by self-report, whereas those with stable high self-esteem had the lowest (those with either stable or unstable low self-esteem were in between). Investigating the joint effects of self-esteem and mood, Kernis, Greenier, Herlocker, Whisenhunt, and Abend (1997) established that after negative mood induction, participants with unstable high self-esteem, relative to others, reported they would be more likely to react defensively in response to a negative event. Regardless of mood induction, participants with stable high self-esteem reported they would be less likely to doubt themselves after a negative event than did others. Kernis, Cornell, Sun, Berry, and Harlow (1993, Study 1) provided participants with experimentally manipulated feedback about their social skills after giving a speech. Compared with participants with stable high self-esteem, those with unstable high self-esteem reported positive feedback more accurate, liked the evaluator more after positive feedback, and judged the evaluator as more competent. After negative feedback, however, participants with unstable high self-esteem liked the evaluator less and judged him or her to be less competent than did participants with stable high self-esteem.

We argue that such reactions on the part of people with unstable high self-esteem reflect attempts to defend underlying self-doubt. Existing research implicates a link between such self-doubt and unstable self-esteem in general, regardless of self-esteem level. In a naturalistic diary study, Greenier et al. (1999) demonstrated that individuals with unstable self-esteem were more affected by the vicissitudes of daily events than were those with stable self-esteem. Negative events had a larger negative impact on participants with unstable self-esteem relative to those with stable self-esteem; positive events had a marginally larger positive impact.

Mark D. Seery, Jim Blascovich, Max Weisbuch, and S. Brooke Vick, Department of Psychology, University of California, Santa Barbara.

This research was partially supported by National Science Foundation Graduate Research Fellowships to Mark D. Seery and Max Weisbuch as well as by a Ford Foundation Predoctoral Fellowship to S. Brooke Vick. Portions of Study 2 were presented at the annual meeting of the Society for Personality and Social Psychology in Los Angeles, CA (February 2003).

Correspondence concerning this article should be addressed to Mark D. Seery, Department of Psychology, University of California, Santa Barbara, CA 93106-9660. E-mail: seery@psych.ucsb.edu

Independent raters judged the negative events reported by participants with unstable versus stable self-esteem to be more esteem relevant. The authors suggested that relative to those with stable self-esteem, persons with unstable self-esteem are more likely to perceive events as self-relevant and link their self-worth to them. Consistent with this idea, Kernis et al. (1998) found that, over time, persons with unstable self-esteem who also reported a high level of daily hassles became more depressed than did others.

If defense of underlying self-doubt in those with unstable high self-esteem motivates their responses to negative feedback, such individuals should exhibit an unprompted "self-doubting" response in the face of failure. In the current investigation, we assessed this reaction by measuring the implicit effects of self-esteem level and stability in the domain of reactions to performance feedback. This paradigm has several advantages. First, the valence of performance feedback (success vs. failure) can be manipulated in a laboratory experiment. Second, participants directly experience such events, in contrast to imagining hypothetical situations and hypothetical reactions. Third, active task performance creates the context required by the biopsychosocial model of challenge and threat (to be discussed shortly) for indexing implicit motivational states online using physiological measures.

Our theory-relevant physiological measures, in turn, offer several new and important advantages for the study of self-esteem level and stability. First, level–stability research conducted to this point has relied exclusively on self-reported dependent variables. Merely using a different methodology can provide convergent validity and increase confidence in the existing collection of findings. Second, because of reliance on self-reported responses, it is unclear whether previously observed effects emerged spontaneously or only after participants were prompted to make retrospective self-reports. Assessing reactions covertly with physiological measures eliminates the possibility that effects occur only after such prompting. Third, the biopsychosocial model of challenge and threat (Blascovich & Mendes, 2000; Blascovich & Tomaka, 1996) in particular is well suited for measuring implicit self-doubt, making it possible to test the hypothesized reactions of people with unstable high self-esteem more directly than was possible in previous research.

Biopsychosocial Model of Challenge and Threat

The biopsychosocial model (Blascovich & Mendes, 2000; Blascovich, Mendes, Tomaka, Salomon, & Seery, 2003; Blascovich & Tomaka, 1996) holds that *challenge* and *threat* represent motivational states that include affective and cognitive, conscious and nonconscious components. Challenge and threat have been investigated primarily in *motivated performance situations*—goal-relevant situations requiring active coping, in which one must act instrumentally to achieve a self-relevant objective. Examples of such situations include taking a test, making a good impression, playing a game, giving a speech, and engaging in athletic competition.[1]

According to Blascovich and colleagues (Blascovich & Mendes, 2000; Blascovich, Mendes, Tomaka, Salomon, & Seery, 2003; Blascovich & Tomaka, 1996), challenge or threat can occur only within the context of *goal relevance* and *task engagement*—essentially, psychological involvement in the task. In fact, a motivated performance situation is defined in part by task engagement. Even if a task would normally require active responses, if an individual lacks a goal or simply does not care enough about the goal, he or she can be neither challenged nor threatened. A goal can be self-relevant for a variety of reasons, including both tangible and intangible consequences for reaching the goal or failing to reach it. For example, receiving a monetary incentive or making an impression on an audience could create or increase self-relevance. Relative differences in task engagement are also meaningful (see Blascovich, Mendes, Hunter, & Salomon, 1999), such that greater task engagement reflects striving toward a goal that possesses greater self-relevance.

Given task engagement, the relative balance of an individual's evaluated coping *resources* and the evaluated *demands* of the situation determine to what extent he or she will experience challenge versus threat. According to the biopsychosocial model, resources include skills, knowledge, and abilities; dispositions; and external support. Demands include danger, uncertainty, and required effort. Evaluations of resources and demands need not be conscious and can be affected by factors outside of conscious awareness. Challenge occurs when evaluated resources meet or exceed evaluated demands, whereas threat occurs when demands outweigh resources. Although sometimes labeled as discrete states, challenge and threat actually reflect opposite ends of a single continuum, such that relative differences in challenge and threat (e.g., greater vs. lesser challenge) are meaningful.

The work of Dienstbier (1989) provides a basis for physiological indexes of challenge and threat motivational states. Dienstbier argued that the body prepares itself for motivated performance situations through activation of the sympathetic–adrenomedullary (SAM) and pituitary–adrenocortical (PAC) axes, both of which serve to mobilize energy reserves. However, the product of SAM activation is the release of catecholamines, including epinephrine and norepinephrine, which have a half-life in the body of only a few minutes, whereas the product of PAC activation is the release of cortisol, which has a half-life in the body of approximately 90 min. Thus, SAM activation allows for a fast spike of energy mobilization, whereas PAC activation does not. A fast onset and offset of SAM activation is characteristic of what Dienstbier referred to as "toughened" individuals. Toughness and SAM activation—relative to lack of toughness and PAC activation—are in turn associated with favorable outcomes, including better task performance, lower anxiety, and improved immune function.

The biopsychosocial model provides for the assessment of cardiovascular responses that are sensitive to SAM versus PAC activation. These physiological changes are then used to index the motivational states (challenge vs. threat) that engender them (for additional discussion, see Blascovich & Tomaka, 1996). Three physiological recording techniques are used in challenge and threat research: electrocardiography, which assesses electrical depolarization of the heart muscle; impedance cardiography, which assesses blood movement in the chest cavity; and blood pressure

[1] Challenge and threat motivational states may also occur during passive-coping tasks that do not require instrumental responses—such as watching a disturbing movie—but the physiological markers of challenge and threat that apply to motivated performance situations have not been validated in passive ones.

measurement. Using these techniques, one can derive a constellation of four cardiovascular measures that identify task engagement and differentiate challenge from threat: heart rate (HR); ventricular contractility (VC), an index of the left ventricle's contractile force; cardiac output (CO), the amount of blood in liters pumped by the heart per minute; and total peripheral resistance (TPR), an index of net constriction versus dilation in the vascular system. For presentational purposes, VC is calculated by multiplying changes in preejection period by –1, where preejection period represents the time in milliseconds in the cardiac cycle from initiation of ventricular depolarization to opening of the aortic valve and ejection of blood; a larger VC value thus corresponds to greater contractility. TPR is calculated by dividing mean arterial pressure by cardiac output and multiplying the total by 80 (Sherwood et al., 1990). For all four measures, baseline resting levels are subtracted from levels exhibited during a motivated performance situation, yielding reactivity scores.

Task engagement is indexed primarily by increases in HR but also by increases in VC from baseline to task; larger increases reflect relatively greater task engagement. Challenge results in higher CO and lower TPR than threat, such that relatively higher CO and lower TPR reflect relatively greater challenge or lesser threat.[2] Differential activation of the SAM and PAC axes underlie these cardiovascular changes. Both challenge and threat result in heightened SAM activation and thus an increase in HR and VC, but threat also results in heightened PAC activation, which inhibits the CO increase and TPR decrease—mediated by the release of epinephrine (see Brownley, Hurwitz, & Schneiderman, 2000)—that would otherwise occur.[3]

Following Dienstbier (1989), the SAM activation that accompanies challenge is well-suited for a situation in which resources meet or exceed demands: The relatively short half-life of epinephrine leads to short-lived mobilization of energy reserves, appropriate for expectations of successful and presumably short-lived coping. The challenge response functions to increase blood flow (greater CO) to skeletal muscles and dilate arteries to accommodate it (lower TPR), which primes the body for potential physical activity. The additional PAC activation that accompanies threat is well-suited for a situation in which demands exceed resources: The relatively long half-life of cortisol leads to long-lived mobilization of energy reserves, appropriate for the possibility of an extended struggle. The threat response results in constriction of arteries (higher TPR), which is characteristic of a vigilance response (Hunter, 2001; Williams, Barefoot, & Shekelle, 1985) also appropriate for an extended struggle; for example, it could be beneficial to watch for subtle changes in conditions or ways to escape the situation.

The utility of challenge and threat and their physiological indexes have been demonstrated in numerous experimental contexts (for reviews, see Blascovich & Mendes, 2000; Blascovich & Tomaka, 1996). Recently, for example, challenge has been shown to result from social comparison with a less capable other (downward comparison) and threat from comparison with a more capable other (upward comparison; Mendes, Blascovich, Major, & Seery, 2001). Greater threat has also been demonstrated during interactions with stigmatized versus nonstigmatized others (Blascovich, Mendes, Hunter, Lickel, & Kowai-Bell, 2001; Mendes, Blascovich, Lickel, & Hunter, 2002).

Self-Esteem, Task Engagement, and Challenge and Threat

We argue that people with unstable high self-esteem possess underlying self-doubt that should affect evaluations of demands and resources in the face of failure feedback, such that individuals with unstable high self-esteem should evaluate lower resources and higher demands than those with stable high self-esteem. Given that lower resources and/or higher demands lead to relative threat, individuals with unstable high self-esteem should exhibit greater threat relative to those with stable high self-esteem. In fact, this relationship should parallel the one between stable low and stable high self-esteem. By virtue of the self-doubt and consistently low self-regard that it entails (e.g., greater depression than others; Kernis, Grannemann, & Mathis, 1991), stable low self-esteem should also be associated with low resources–high demand evaluations. Thus, individuals with stable low self-esteem should respond similarly to those with unstable high self-esteem when confronted with failure, exhibiting threat relative to people with stable high self-esteem.

Study 1

Study Overview

Participants first completed measures of self-esteem level and stability. They then arrived for a separate individual laboratory session, whereupon they received veridical success or failure feedback after taking a relatively easy (success feedback) or difficult (failure feedback) version of the Remote Associates Test (RAT). Participants then completed a moderate-difficulty version of the

[2] In previous discussions of challenge and threat, VC has been used along with CO and TPR to differentiate the two states. However, VC does not always differ between challenge and threat and so is treated here as a measure of task engagement. The previously observed association between greater VC and greater challenge may at least in part be due to the strategy used to identify the opening of the aortic valve when the point is ambiguous. Specifically, if the inflection in the impedance cardiography waveform that is used to identify the point is very subtle—perhaps because of poor signal quality—it can be unclear whether the valve opening is not visible or whether it actually occurs earlier in the cardiac cycle. The earlier the point is marked, the higher VC and CO will be, thus possibly inflating the degree of their association. In the current research, ambiguous cases were excluded rather than marked earlier on the waveform. Although VC should increase from baseline during both challenge and threat, it is unclear how effectively it indexes relative levels of engagement because it may also partially reflect differences in challenge versus threat. Therefore, only CO and TPR were used here to differentiate challenge from threat, whereas HR was used as the primary measure of task engagement and VC was used as the secondary measure.

[3] During both challenge and threat, SAM activation has two principal effects via direct innervation of tissue: (a) it stimulates the heart muscle, which increases HR and VC, and (b) it constricts veins and venules, which increases venous return, thereby further increasing VC and potentially increasing CO. During challenge, SAM activation also stimulates the preferential release of epinephrine from the adrenal medulla, the primary effect of which is to act on beta-2 receptors and cause vasodilation in skeletal muscle beds, resulting in a net decrease in TPR and facilitating an increase in CO (Brownley et al., 2000; Papillo & Shapiro, 1990). During threat, however, SAM activation is tempered by increased PAC activation, which inhibits this release of epinephrine.

RAT. Our predictions concerned cardiovascular responses exhibited during the second task.

We hypothesized that the combination of feedback condition, self-esteem level, and self-esteem stability would predict TPR and CO reactivity, which differentiate challenge from threat. We expected a three-way interaction, such that a two-way interaction between level and stability would only be observed in the failure feedback condition. Because these interactions best capture the comprehensive pattern of differences we predicted, they provide the best statistical tests of our hypotheses. However, we anticipated that the interactions would be driven by three simple effects in particular: (a) Within the two-way interaction, we predicted a simple effect of stability within high self-esteem, such that unstable high self-esteem would be associated with relative threat (higher TPR and lower CO) compared with stable high self-esteem; (b) within the two-way interaction, we predicted a simple effect of level within stable self-esteem, such that stable low self-esteem would also be associated with relative threat compared with stable high self-esteem; and (c) we predicted a simple effect of condition within unstable high self-esteem, such that those participants would exhibit relative threat after failure feedback compared with success feedback.

Method

Participants

One hundred thirteen (82 women, 31 men) undergraduates at the University of California, Santa Barbara, participated in the study for introductory psychology course credit.[4]

Laboratory Setting

The laboratory portion of the study took place in the social psychophysiology laboratory at the University of California, Santa Barbara. The laboratory consisted of a control room and separate experimental rooms. The control room contained physiological equipment, video monitors and recorders connected to two cameras in each experimental room, and an audio tape player and intercom system connected to speakers in the experimental rooms. Each experimental room was divided into a preparation room, where participants completed forms and where an experimenter applied sensors to participants, and a recording room, where data collection took place. The acoustically and environmentally controlled recording room, measuring approximately 3.0 × 3.5 m, contained physiological, audiovisual, and computer equipment. Participants received instructions through a speaker and interacted with the experimenter through an intercom. Participants sat upright in a comfortable easy chair throughout the experiment with a tray on their laps on which they filled out questionnaires and used a computer keyboard.

Cardiovascular Measures

Cardiovascular measures were recorded noninvasively, following commonly accepted guidelines (Sherwood et al., 1990) and utilizing a Minnesota Impedance Cardiograph (Model 304B) and a Cortronics (Model 7000) continuously inflated blood pressure monitor. Signals were conditioned using Coulbourn amplifiers (Models S75-11 and S79-02, Coulbourn Instruments, Allentown, PA) and were stored on a desktop computer.

Impedance cardiograph (ZKG) and electrocardiograph (EKG) recordings provided continuous measures of cardiac performance. The impedance cardiograph utilized a tetrapolar aluminum/mylar tape electrode system to record basal transthoracic impedance (Z0) and the first derivative of basal impedance (dZ/dt). Two pairs of band electrodes completely encircled participants' bodies. The two inner electrodes were placed at the base of the neck and at the xiphisternal junction (approximately midchest); the two outer electrodes were placed on the neck and abdomen, separated from the respective inner electrodes by a distance of at least 3 cm. EKG signals were detected using either a Standard Lead II electrode configuration (additional spot electrodes on the right arm and both legs) or through the band electrodes. The Cortronics blood pressure monitor collected continuous noninvasive recordings of blood pressure from the brachial artery of participants' nondominant arm. In combination, ZKG and EKG recordings allow computation of HR, VC, and CO; the addition of blood pressure monitoring allows computation of TPR. The recorded data were scored using an interactive MS-DOS software program (Kelsey & Guethlein, 1990); scoring was performed blind to condition and self-esteem.

RAT

Adopted from McFarlin and Blascovich (1984), the items of the RAT each consisted of three related words; participants' task was to generate the single word that linked the other three together (e.g., in a difficult item, *deep* is the correct response to *bass*, *complex*, *sleep*). Each version of the RAT—easy, difficult, and moderate—consisted of 12 items presented serially on a computer. Participants had 15 s to answer each item aloud, after which time the program automatically advanced to the next item. Participants could also manually advance to the next item before time expired by pressing the space bar, although no extra time carried over to the next item after doing so.

The RAT allows manipulation of performance feedback while minimizing deception and the concomitant risk of suspicion. This is accomplished by administering easy and difficult versions of the RAT. McFarlin and Blascovich (1984) found that participants completing an easy version were reliably more successful than those completing a difficult version. In addition, participants' perceptions of their performance matched their actual performance, such that participants who completed an easy version thought they had done well, whereas those who completed a difficult version thought they had done poorly. Because of the veridical performance information that the RAT inherently offers, the experimenter need not provide deceptive feedback.

Directly relevant to the nature of this study (i.e., requirement of a motivated performance situation), McFarlin and Blascovich (1984) found that participants viewed the RAT as a fun and interesting exercise. Simply caring about the task should have created goal relevance for participants, resulting in task engagement. Presenting the task as a test of reasoning ability, as was done in the present studies, should have amplified goal relevance because performance quality had implications for a domain that is likely to be personally relevant to college students. Goal relevance should have resulted both from the potential for self-evaluation and evaluation by the experimenter. Given the nature of the RAT and its presentation in the present studies, no tangible incentives (e.g., a monetary reward for reaching a performance criterion) should have been required to engage participants in the task.

[4] Nine additional participants completed all elements of the study but were excluded from analyses: Two participants yielded cardiovascular data which were impossible to score reliably because of poor impedance cardiograph signal quality (e.g., ambiguous aortic valve opening); 4 participants were excluded because of blood pressure monitor malfunction; 2 participants were excluded because of experimenter error in conducting the study; and 1 participant was unable to perform the tasks because of difficulty with the English language.

Procedure

Assessment of self-esteem level and stability. Participants met in groups of up to 20 for an introduction to the study—which was described simply as a multiple-part study that involved asking about their thoughts and feelings and later assessing their physiological responses—and to receive materials. Participants then completed a modified version of the Rosenberg Self-Esteem Scale (1965; in Blascovich & Tomaka, 1991) eight times over the course of 1 week. The measure consisted of 10 items that were recast to ask how respondents felt at that moment, rather than in general, thus capturing state rather than trait self-esteem (e.g., "Right now I feel that I am a person of worth, at least on an equal basis with others"). Participants responded using a 9-point, Likert-type scale anchored at *strongly disagree* and *strongly agree*. Possible scores range from 10 to 90; higher scores represent higher self-esteem. In this sample, Cronbach's alpha for each of the questionnaires ranged from .90 to .92. Following the procedure outlined by Kernis et al. (1993), participants were instructed to complete the forms approximately 12 hr apart, starting that night and ending 4 days later in the morning. For each participant, the standard deviation of scores over these multiple assessments was calculated, reflecting the degree of fluctuation in state self-esteem, such that a higher standard deviation indicated greater day-to-day fluctuation and greater self-esteem instability. We assessed self-esteem level by calculating the mean state self-esteem score.[5] Consistent with previous work (see Kernis, 1993), self-esteem level and stability were negatively correlated, $r = -.53$. Participants were retained for the laboratory phase of the experiment if they completed at least six of the eight questionnaires; 3 participants were excluded on this basis. Participant instructions emphasized the importance of completing forms at the proper times, but they also stressed that the worst possible outcome would be if participants completed multiple forms at the same time. To avoid this possibility, instructions urged participants to leave forms blank if they forgot to complete them at the scheduled times.

Laboratory procedure. After participants completed and returned their state self-esteem forms, a research assistant contacted them by phone to schedule the laboratory component of the experiment, which participants completed individually. An experimenter greeted participants at the lab and gave them an information sheet to read describing the cardiovascular measurement procedures they would experience during the study. No other information about the study was provided. The experimenter applied cardiovascular sensors to participants, who then heard taped audio instructions that welcomed them to the lab and asked them to sit quietly for the next few minutes. During this time, 5 min of baseline cardiovascular data were recorded.

Next, taped instructions informed participants that the study entailed measuring physiological responses during tests of reasoning ability. The instructions explained the nature of the upcoming task to participants as they completed a short tutorial on the computer, including two sample RAT items. The experimenter then prompted participants to begin the first task. In the success feedback condition, participants completed the easy version of the RAT; in the failure feedback condition, participants completed the difficult version of the RAT. The experimenter tallied participants' answers as they were said aloud.

At the end of the first task, the experimenter used the intercom to inform participants how many items they answered correctly. The substantive content of the feedback was veridical, but the delivery differed between conditions. In the success feedback condition, the experimenter said, "Okay, (participant's name), you did a great job! You got (number) items right," with a slightly higher pitch at the end of the sentences, designed to sound encouraging. In the failure feedback condition, the experimenter said, "Okay, (participant's name), you didn't do that well," followed by either, "you only got (number) item(s) right," or, "you didn't get any items right," as appropriate, using a hesitant tone.

Next, the experimenter played taped instructions that asked participants to complete the manipulation check questionnaire, which was facedown underneath the computer keyboard. The first two items ("The task I just completed was a difficult one," and "I feel that I performed well on this task") used a 9-point, Likert-type scale anchored at *strongly disagree* and *strongly agree*. For the third item ("How many items out of twelve did you answer correctly?"), participants wrote a number in a blank. The experimenter then played additional taped instructions that asked participants to sit quietly for several minutes. During this rest period, 5 min of cardiovascular data were recorded.

On conclusion of the second rest period, the experimenter explained over the intercom that participants would hear additional taped instructions. These instructions stated that participants were about to begin a different version of the reasoning ability test they completed earlier; the items would be different, but the format would be the same. The instructions included additional material designed to heighten goal relevance (i.e., the importance of doing well or poorly)—and thus task engagement—during the upcoming task (all participants heard the entire set of instructions): The first test was a practice test, and only the results from the second test would be recorded; the research team would use results from the second test to determine whether participants were among the top performers; the research team would videotape the performance; and several judges would be watching participants carefully from the control room. After the instructions, all participants completed the moderate-difficulty version of the RAT while 3 min of cardiovascular data were recorded. Finally, the experimenter removed all sensors from participants and fully debriefed them. During the debriefing, the experimenter probed participants as to whether they had filled out state self-esteem forms at the proper times, emphasizing that they would not be penalized in any way for failure to do so. In Study 1, 1 participant admitted to completing a questionnaire late, immediately

[5] Before participants heard instructions for the state self-esteem forms, we also administered the Rosenberg Self-Esteem Scale worded for trait self-esteem, assessing general instead of current feelings. Typically (e.g., Greenier et al., 1999; Kernis et al., 1989), assessing self-esteem level by calculating the mean state self-esteem score yields results that are similar to those obtained from assessing level with the single trait scale. However, in Study 1, analyses that used the trait scale generally yielded results that did not approach significance. It is possible that elements of the trait scale administration in Study 1 contributed to its lack of predictive ability, so we have reported results from analyses that used the mean state self-esteem score. We made several changes in Study 2 to address this issue. To help ensure that participants did not feel rushed as they completed the trait self-esteem scale, the experimenter read instructions for the state self-esteem forms before participants completed any questionnaires. Participants were then able to complete the trait scale at their own pace, after which they were free to go. In addition, in an attempt to reduce misreading of questions and measurement error in general, we changed the response scale for the trait and state forms. Instead of a nine-point scale in which points were represented by dots and only the endpoints were labeled—as was used in Study 1—the forms used a 5-point scale, with all points represented by both numbers and labels (*strongly disagree* to *strongly agree*). In combination with these factors, it is conceivable that participants were more likely to inaccurately answer items on the trait scale than on the state scales in Study 1 because they were more accustomed to the general scale format when completing the state scales. This could explain why analyses using the mean state self-esteem score yielded significant effects whereas analyses using the trait scale did not. In Study 2, consistent with previous work, using the trait scale and mean state score to assess self-esteem level yielded similar patterns of results, although effect sizes tended to be slightly larger when using the trait scale (see Kernis et al., 1989). We planned a priori to use the trait scale because it has been preferred in previous research, so only those analyses are reported for Study 2. The convergence of findings between Study 1 and Study 2 suggests that these differences in assessing self-esteem level have little bearing on our conclusions.

before the next scheduled questionnaire (7 admitted to doing so in Study 2); any questionnaires completed retrospectively in this fashion were excluded.

Results

Analytical Strategy

Following Kernis and colleagues (e.g., Kernis et al., 1993), we treated self-esteem level and stability as continuous variables on the grounds that median splits result in information loss. Continuous predictor variables were centered by subtracting the variable mean from all scores, thereby setting means equal to zero (noncentered means are reported in the text).

In preliminary analyses, we tested manipulation checks, quality of performance in the second task, and task engagement (HR and VC reactivity) during the second task; in primary analyses—corresponding to our predictions—we tested challenge–threat (TPR and CO reactivity) during the second task. We expected that explicit feedback delivered by the experimenter in combination with the experience of performing relatively well or poorly during the first task would yield larger effect sizes than the performance experience alone. Thus, our predictions concerned cardiovascular reactivity during the second task, not the first. In addition, our studies were designed to assess reactions to feedback between subjects, not within, so we did not attempt analyses that compared pre–post explicit feedback reactivity (i.e., comparing the first task to the second task); indeed, such analyses would have little meaning because the first task itself provides an important part of the feedback manipulation.

We used multiple regression in four steps to test the effects of self-esteem level, stability, condition, and their interactions on dependent variables; at each step, all terms were entered simultaneously. Step 1 included any relevant covariates. Although no covariates were included for nonphysiological dependent variables, the same step labels are used to avoid confusion. Step 2 included terms for level, stability, and condition, allowing us to assess the regression analog of main effects. Step 3 included the three possible two-way interactions: Level × Stability, Level × Condition, and Stability × Condition. Finally, Step 4 included the three-way interaction: Level × Stability × Condition. Our primary hypotheses depended on the significance of terms after the addition of Step 4 to the regression model. To interpret interactions, we conducted simple effects tests by shifting the zero point of variables to one standard deviation above or below the mean, recalculating interaction terms, and repeating the regression analysis, thereby allowing a test of, for example, the effect of stability within high self-esteem. For effects that we did not explicitly predict, we report results as nonsignificant without additional elaboration when $p > .10$.

Cardiovascular reactivity values were calculated by subtracting the baseline value—the last minute of the initial rest period—from the value obtained during the relevant task minute. Values exceeding 3.3 *SD*s from the mean ($p = .001$ in a normal distribution) were identified as extreme and were winsorized by assigning them a value 1% higher than the next-highest nonextreme value, thereby decreasing the influence of the extreme value while maintaining the rank order of the distribution. In both studies, between zero and three values were winsorized for each dependent variable. Reactivity values were averaged across the 1st and 2nd minutes of each task. Although the use of change scores (of which reactivity is one example) is sometimes discouraged on psychometric grounds (e.g., Cronbach & Furby, 1970), their use is common in psychophysiological work. In the context of assessing task reactivity from baseline, Llabre, Spitzer, Saab, Ironson, and Schneiderman (1991) concluded that the reliability of change scores typically is comparable to or exceeds that of residualized change scores, calculated by regressing task levels on baseline levels and then subtracting the generated predicted values from observed values. Unlike some other potential applications, changes from the baseline zero point do have meaning for our purposes in that we typically examine both relative differences in challenge versus threat (as is done in this article) and—when using a factorial design—absolute levels of challenge versus threat by testing reactivity against zero. However, because of the possibility that change scores can produce artifactual results because of correlations between baseline levels and magnitude of change, we controlled for baseline levels when predicting reactivity. This should have accounted for any confounding effect that magnitude of baseline level had on magnitude of reactivity.[6]

In all analyses of challenge and threat using TPR and CO reactivity, we also controlled for task engagement (HR and VC reactivity). This served two purposes: First, task engagement reflects SAM activation, which is common to both challenge and threat, so controlling for SAM activation should increase power to detect differences in PAC activation, which differentiates challenge versus threat; second, we sought to demonstrate that any differences in task engagement did not account for challenge and threat effects.

Finally, because task engagement is a prerequisite for challenge and threat, we established that participants in the sample were, on average, engaged in the second task before we tested for TPR and CO differences. We used *t* tests to assess whether HR and VC increased significantly from baseline. In addition, when predicting TPR and CO, we excluded participants who were not engaged in the task because any changes in their TPR and CO would not be interpretable. We adopted a conservative criterion, dropping participants who exhibited decreases in both HR and VC during the second task; this resulted in the exclusion of 3 participants in Study 1 and 9 in Study 2. These participants were included in all analyses of task engagement.[7]

Preliminary Analyses

Manipulation checks. We conducted a series of *t* tests to assess whether the difficult RAT (failure feedback condition; $n = 57$) was indeed more difficult than the easy RAT (success feedback condition; $n = 56$). Confirming the intent of the manipulation, participants answered fewer items correctly on the difficult RAT ($M = 0.93$, $SD = 0.90$) than on the easy RAT ($M = 8.20$, $SD = 2.03$), $t(111) = 24.65$, $p < .01$. In addition, participants who completed the difficult RAT reported that their task was more

[6] In all analyses of the second task in both studies, also controlling for the last minute of the second rest period had a negligible effect on the reported results, as did including second-task performance as a covariate.

[7] Including disengaged participants in analyses of TPR and CO had no effect on overall patterns of results.

difficult ($M = 8.30$, $SD = 1.30$) than participants who completed the easy RAT ($M = 4.40$, $SD = 1.87$), $t(110) = 12.86$, $p < .01$. Participants who completed the difficult RAT also reported performing worse ($M = 1.68$, $SD = 1.28$) than those who took the easy RAT reported ($M = 6.11$, $SD = 1.64$), $t(110) = 15.93$, $p < .01$.

In regression analyses, the magnitude of these condition effects varied little across combinations of self-esteem level ($M = 74.92$, $SD = 11.06$) and stability ($M = 6.06$, $SD = 3.97$). The only other effect that approached significance was a marginal two-way interaction between level and stability in Step 3 for difficulty ratings ($sr^2 = .013$, $p = .064$). Testing simple effects revealed that within high self-esteem, unstable self-esteem was associated with higher difficulty ratings than those associated with stable self-esteem ($B = 0.18$, $\beta = .28$, $sr^2 = .014$, $p = .055$), regardless of feedback condition.

Finally, the number of items that participants answered correctly—which the experimenter said aloud when delivering feedback—correlated .99 with the number of items that participants reported answering correctly in the third item on the manipulation check questionnaire. Therefore, as intended, participants in the failure feedback condition not only performed objectively worse than those in the success feedback condition but also judged the first task to be more difficult and reported performing worse as compared with participants in the success feedback condition.

Performance in the second task. There were no significant effects of condition, self-esteem, or their interaction on number of items answered correctly in the second task ($M = 5.92$, $SD = 1.98$).

Task engagement in the second task. Regressing HR in the second task on feedback condition, self-esteem level, self-esteem stability, and their interactions yielded a marginal effect of stability ($B = -0.35$, $\beta = -.21$, $sr^2 = .031$, $p = .058$) in Step 2, such that greater instability was associated with smaller increases in HR, consistent with lower task engagement. The regression for VC reactivity yielded a significant effect of stability ($B = -0.29$, $\beta = -.24$, $sr^2 = .039$, $p < .05$) in Step 2, such that greater instability was associated with smaller increases in VC, consistent with lower task engagement. No interactions approached significance for HR or VC.

Testing sample means with t tests revealed that HR reactivity ($M = 6.78$, $SD = 6.73$) was significantly greater than zero, $t(112) = 10.70$, $p < .01$, as was VC reactivity ($M = 4.69$, $SD = 4.99$), $t(112) = 9.99$, $p < .01$. This demonstration of task engagement met the assumptions of the biopsychosocial model and allowed us to assess challenge and threat.

Table 1
Summary of Regression Analysis Predicting TPR Reactivity in the Second Task, Study 1

Variable	B	$SE\ B$	β	sr^2
Step 1				
Baseline TPR	0.053	0.037	.139	.019
HR reactivity	0.053	1.775	.003	.000
VC reactivity	−2.226	2.389	−.097	.008
Step 2				
Feedback condition	−26.459	21.120	−.120	.014
Self-esteem level	−1.909	1.115	−.192	.027†
Self-esteem stability	−4.142	3.231	−.146	.015
Step 3				
Condition × Level	0.286	2.252	.021	.000
Condition × Stability	−7.879	6.371	−.203	.014
Level × Stability	0.315	0.266	.126	.013
Step 4				
Condition × Level × Stability	−1.031	0.534	−.292	.033†

Note. For the feedback condition variable, failure was set to 0 and success was set to 1. TPR = total peripheral resistance; HR = heart rate; VC = ventricular contractility.
† $p < .10$.

Primary Analyses: Challenge and Threat in the Second Task

For TPR reactivity in the second task (see Table 1), the regression in Step 2 revealed a marginal effect for self-esteem level ($p = .090$), such that higher self-esteem was associated with lower TPR, consistent with relative challenge. However, this effect was qualified by the predicted three-way interaction in Step 4 ($p = .057$; see Figure 1). Also as predicted, the two-way interaction between self-esteem level and stability (within the three-way interaction) only reached significance in the failure condition ($sr^2 = .042$, $p < .05$). Consistent with predictions, testing the three hypothesized simple effects yielded: (a) a marginal effect of stability within high self-esteem after failure ($B = 12.12$, $\beta = .43$, $sr^2 = .024$, $p = .102$), such that greater instability was associated with higher TPR, consistent with relative threat; (b) a significant effect of level within stable self-esteem after failure ($B = -5.09$, $\beta = -.51$, $sr^2 = .050$, $p < .05$), such that lower self-esteem was associated with higher TPR; and (c) a significant effect of condition within unstable high self-esteem ($B = -142.38$, $\beta = -.65$, $sr^2 = .041$, $p < .05$), such that participants exhibited higher TPR after failure feedback than after success feedback. In addition, although not predicted, a marginal simple effect emerged for condition within unstable low self-esteem ($B = -55.96$, $\beta = -.25$, $sr^2 = .026$, $p = .089$), such that participants exhibited higher TPR after failure feedback than after success feedback. No other significant effects emerged for TPR.

For CO reactivity in the second task (see Table 2), no significant effects emerged in Steps 2 or 3. The predicted three-way interaction in Step 4 was marginally significant ($p = .106$);[8] within the failure condition, the predicted two-way interaction between self-esteem level and stability was significant ($sr^2 = .037$, $p < .05$). Testing the three hypothesized simple effects yielded: (a) a nonsignificant effect of stability within high self-esteem after failure ($B = -0.07$, $\beta = -.37$, $sr^2 = .018$, $p = .128$), although the direction was such that greater instability was associated with lower CO, consistent with relative threat; (b) a significant effect of level within stable self-esteem after failure ($B = 0.029$, $\beta = .43$, $sr^2 = .036$, $p < .05$), such that lower self-esteem was associated with lower CO; and (c) a marginal effect of condition within unstable high self-esteem ($B = 0.71$, $\beta = .48$, $sr^2 = .022$, $p = .089$), such that participants exhibited lower CO after failure feedback than after success feedback. The simple effect of condi-

[8] Because tests of higher order interactions in regression are typically marked by relatively low statistical power, we interpreted predicted three-way interactions that approached significance.

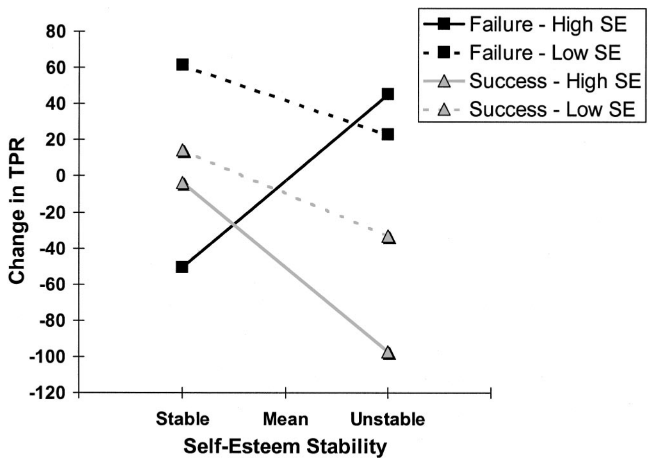

Figure 1. The regression of total peripheral resistance (TPR) reactivity in the second task on self-esteem (SE) level, self-esteem stability, and condition in Study 1; higher TPR reflects greater relative threat. "High SE" and "unstable" represent values 1 standard deviation above the mean of level and stability, respectively, whereas "low SE" and "stable" represent values 1 standard deviation below the mean.

tion within unstable low self-esteem that was observed for TPR no longer approached significance ($B = 0.28$, $\beta = .19$, $sr^2 = .015$, $p = .167$), although the direction of the effect was consistent with TPR results. No other significant effects emerged for CO.

Discussion

Primary analyses of TPR and CO during the second task revealed the predicted interactions. After failure feedback—consistent with our hypotheses—participants with unstable high self-esteem exhibited threat relative to those with stable high self-esteem, as did participants with stable low self-esteem. In addition, the reaction of participants with unstable high self-esteem was affected by feedback, such that they exhibited relative challenge after success but relative threat after failure. These results are consistent with our argument that individuals with unstable high self-esteem possess implicit self-doubt. If their underlying self-doubt is triggered when they are faced with failure feedback, they should evaluate lower resources–higher demands, leading to the experience of threat relative to individuals with stable high self-esteem. This should parallel the reaction of individuals with stable low self-esteem, who should also possess self-doubt. The observed data from Study 1 support these contentions.

However, alternative explanations are also possible. To draw conclusions about people with unstable high self-esteem, we needed to contrast them with people with stable high self-esteem. The latter exhibited relative challenge in the face of negative feedback, presumably because they possess implicit confidence rather than self-doubt, leading to relatively high resource–low demand evaluations. It is possible that this challenge response does not reflect confidence, however, but instead one of the following: (a) they disregarded the feedback information; (b) they disengaged from the task, thus possibly decreasing threat as well; or (c) they were differentially affected by the goal-relevance-heightening instructions immediately before the second task. Our preliminary analyses can shed additional light on these alternatives.

The manipulation check data do not support the first possibility: Participants with stable high self-esteem—like other participants—rated the difficult task (failure condition) more difficult than the easy task (success condition) and reported performing less well, indicating that they were sensitive to the feedback. Similarly, task engagement data from the second task do not support the second possibility: Participants were on average engaged in the task, and the only effect to emerge was for stability, such that greater stability was associated with *higher* task engagement. Regarding the third possibility, participants with stable high self-esteem may have reacted differently than others to the information that the first test was a practice trial; specifically, they may have been better able to discount the failure feedback by making a favorable attribution to inexperience. We addressed this possibility in Study 2.

Study 2

In Study 2, we sought to replicate and extend the findings from Study 1 after making two changes. First, we made the failure feedback subtler and eliminated the goal-relevance-heightening instructions that occurred immediately before the second task in Study 1. Although we anticipated that this could decrease effect sizes for cardiovascular reactivity in the second task (i.e., task engagement and challenge–threat), the change provided a more stringent test of our hypotheses; specifically, by reducing or eliminating the possibility that observed effects were driven by participants with stable high self-esteem responding differently than others to the instructions (e.g., that the first test was a practice trial). Second, we collected data from a larger sample than was used in Study 1 to guard against this potential loss of statistical power to detect effects.

Table 2
Summary of Regression Analysis Predicting CO Reactivity in the Second Task, Study 1

Variable	B	$SE\ B$	β	sr^2
Step 1				
Baseline CO	−0.751	0.031	.236	.054**
HR reactivity	0.020	0.011	.171	.024†
VC reactivity	0.044	0.015	.289	.070**
Step 2				
Feedback condition	0.144	0.132	.097	.009
Self-esteem level	0.008	0.007	.126	.012
Self-esteem stability	0.024	0.020	.127	.011
Step 3				
Condition × Level	−0.003	0.014	−.031	.000
Condition × Stability	0.035	0.040	.135	.006
Level × Stability	−0.002	0.002	.146	.017
Step 4				
Condition × Level × Stability	0.005	0.003	.228	.020

Note. For the feedback condition variable, failure was set to 0 and success was set to 1. CO = cardiac output; HR = heart rate; VC = ventricular contractility.
† $p < .10$. ** $p < .01$.

Study Overview

Participants first completed measures of self-esteem level and stability. They then arrived for a separate individual laboratory session, whereupon they received veridical success or failure feedback after taking a relatively easy (success feedback) or difficult (failure feedback) version of the RAT. Participants then completed a second moderate-difficulty version of the RAT. Cardiovascular responses were assessed during the second task. Unless otherwise noted, elements of the method in Study 2 were identical to those in Study 1.

As in Study 1, we hypothesized that the combination of feedback condition, self-esteem level, and self-esteem stability would predict TPR and CO reactivity (challenge vs. threat). We expected a three-way interaction in the second task, such that a two-way interaction between level and stability would only be observed in the failure feedback condition. We also predicted the same three simple effects as in Study 1: (a) Within the two-way interaction, a simple effect of stability within high self-esteem, such that unstable high self-esteem would be associated with relative threat (higher TPR and lower CO) compared with stable high self-esteem; (b) within the two-way interaction, a simple effect of level within stable self-esteem, such that stable low self-esteem would also be associated with threat relative to stable high self-esteem; and (c) a simple effect of condition within unstable high self-esteem, such that those participants would exhibit relative threat after failure feedback compared with success feedback.

Method

Participants

One hundred seventy-five (136 women, 39 men) undergraduates at the University of California, Santa Barbara, participated in the study for introductory psychology course credit.[9]

Procedure

Assessment of self-esteem level and stability. After hearing an introduction to the study and instructions for the stability forms, participants completed the Rosenberg Self-Esteem Scale worded for trait self-esteem, which we used to assess self-esteem level. For both level and stability forms, a 5-point response scale was used, such that possible scores range from 10 to 50; higher scores represent higher self-esteem. In this sample, Cronbach's alpha was .88 for the level questionnaire and ranged from .90 to .94 for the stability questionnaires. As in Study 1, self-esteem level and stability were negatively correlated, $r = -.32$. Participants were retained for the laboratory phase of the experiment if they completed at least six of the eight stability questionnaires; 7 participants were excluded on this basis.

Laboratory procedure. The procedure in Study 2 was identical to the procedure in Study 1 in all respects except the following. The content of feedback changed slightly in the success condition to, "Okay, (participant's name), you actually did a great job! You got (number) items right." The feedback in the failure condition changed to, "Okay, (participant's name)," followed by either, "you only got (number) item(s) right," or, "you didn't get any items right," as appropriate. All questionnaires used a 5-point Likert-type response scale, matching the one used in the level and stability forms. Before the second task, participants heard nothing about a practice test, an audience, or other notable circumstances, only that they were about to begin a different version of the reasoning ability test that they completed earlier.

Results

Preliminary Analyses

Manipulation checks. As in Study 1, we conducted a series of *t* tests to assess whether the difficult RAT (failure feedback condition; $n = 89$) was indeed more difficult than the easy RAT (success feedback condition; $n = 86$). Confirming the intent of the manipulation, participants answered fewer items correctly on the difficult RAT ($M = 1.04$, $SD = 1.22$) than on the easy RAT ($M = 7.99$, $SD = 1.64$), $t(173) = 31.80$, $p < .01$. In addition, participants who completed the difficult RAT reported that their task was more difficult ($M = 4.51$, $SD = 0.80$) than participants who completed the easy RAT reported ($M = 3.07$, $SD = 0.97$), $t(173) = 10.72$, $p < .01$. Participants who completed the difficult RAT also reported performing worse ($M = 1.29$, $SD = 0.64$) than those who took the easy RAT ($M = 3.34$, $SD = 0.86$), $t(173) = 17.82$, $p < .01$.

In regression analyses, the magnitude of these condition effects varied little across combinations of self-esteem level ($M = 39.63$, $SD = 6.40$) and stability ($M = 3.53$, $SD = 2.32$), although several other significant or marginal effects emerged. For difficulty ratings, the regression in Step 3 revealed a significant interaction between self-esteem level and stability ($sr^2 = .016$, $p < .05$). Within this interaction, the simple effect of level within stable self-esteem was marginally significant ($B = -0.041$, $\beta = -.23$, $sr^2 = .012$, $p = .068$), such that higher self-esteem was associated with lower difficulty ratings. The regression for ratings of performing well yielded a significant effect of level in Step 2 ($B = 0.023$, $\beta = .11$, $sr^2 = .011$, $p < .05$), such that higher self-esteem was associated with better reported performance, although this was qualified by a marginal interaction between condition and level in Step 3 ($sr^2 = .006$, $p = .076$), such that the effect was only significant in the success condition ($B = 0.038$, $\beta = .19$, $sr^2 = .017$, $p < .01$).

Finally, the number of items that participants answered correctly—which the experimenter said aloud when delivering feedback—correlated .98 with the number of items that participants reported answering correctly in the third item on the manipulation check questionnaire. Therefore, as intended, participants in the failure feedback condition not only performed objectively worse than those in the success feedback condition but also judged the first task to be more difficult and reported performing worse compared with participants in the success feedback condition.

Performance in the second task. For the regression of number of items answered correctly in the second task ($M = 5.71$, $SD = 2.30$) on level and stability, a significant effect of condition emerged in Step 2 ($B = 0.76$, $\beta = .17$, $sr^2 = .027$, $p < .05$), such

[9] Nineteen additional participants completed all elements of the study but were excluded from analyses: 6 participants yielded cardiovascular data that were impossible to score reliably because of poor impedance cardiograph signal quality (e.g., ambiguous aortic valve opening); 3 participants were excluded because of blood pressure monitor malfunction; 8 participants were excluded because of experimenter error in conducting the study (e.g., providing incorrect feedback, failing to collect cardiovascular data); and 2 participants were excluded because of suspicion about the study's hypotheses.

that participants answered more items correctly after success feedback than after failure feedback.[10]

Task engagement in the second task. The regression for HR reactivity in the second task yielded effects in Step 2 for level ($B = -0.15$, $\beta = -.16$, $sr^2 = .023$, $p < .05$) and—as in Study 1—stability ($B = -0.40$, $\beta = -.15$, $sr^2 = .020$, $p < .05$), such that higher level and greater instability were both associated with lower HR, consistent with lower task engagement. The regression for VC failed to yield significant effects.

Testing sample means with t tests revealed that HR reactivity ($M = 4.24$, $SD = 6.07$) was significantly greater than zero, $t(174) = 9.25$, $p < .01$, as was VC reactivity ($M = 3.39$, $SD = 4.99$), $t(174) = 8.98$, $p < .01$, demonstrating task engagement in the second task.

Primary Analyses: Challenge and Threat in the Second Task

For the regression of TPR reactivity in the second task (see Table 3), only the predicted three-way interaction in Step 4 approached significance ($p = .070$; see Figure 2). Also as predicted, the two-way interaction between self-esteem level and stability only reached significance in the failure condition ($sr^2 = .022$, $p < .05$). Consistent with predictions, testing the three hypothesized simple effects yielded: (a) a marginal effect of stability within high self-esteem after failure ($B = 10.93$, $\beta = .30$, $sr^2 = .019$, $p = .067$), such that greater instability was associated with higher TPR, consistent with relative threat; (b) a marginal effect of level within stable self-esteem after failure ($B = -4.21$, $\beta = -.33$, $sr^2 = .015$, $p = .098$), such that lower self-esteem was associated with higher TPR; and (c) a nonsignificant effect of condition within unstable high self-esteem ($B = -38.51$, $\beta = -.23$, $sr^2 = .009$, $p = .196$), although the direction was such that participants exhibited higher TPR after failure feedback than after success feedback. No other significant effects emerged for TPR.

Table 3
Summary of Regression Analysis Predicting TPR Reactivity in the Second Task, Study 2

Variable	B	$SE\ B$	β	sr^2
Step 1				
Baseline TPR	−0.055	0.027	−.149	.022*
HR reactivity	2.049	1.030	.149	.022*
VC reactivity	−4.975	1.275	−.293	.083**
Step 2				
Feedback condition	−3.830	12.558	−.023	.001
Self-esteem level	−0.574	1.036	−.045	.002
Self-esteem stability	0.664	2.948	.018	.000
Step 3				
Condition × Level	−0.617	2.176	−.039	.000
Condition × Stability	−2.783	6.013	−.052	.001
Level × Stability	0.417	0.439	.078	.005
Step 4				
Condition × Level × Stability	−1.616	0.885	−.247	.018†

Note. For the feedback condition variable, failure was set to 0 and success was set to 1. TPR = total peripheral resistance; HR = heart rate; VC = ventricular contractility.
† $p < .10$. * $p < .05$. ** $p < .01$.

For the regression of CO reactivity in the second task (see Table 4), no significant effects emerged in Steps 2 or 3. The predicted three-way interaction in Step 4 did not reach significance ($p = .119$); within the failure condition, the predicted two-way interaction between self-esteem level and stability also failed to reach significance ($sr^2 = .010$, $p = .172$). However, the direction of these effects was consistent with TPR results.[11]

Discussion

Primary analyses of TPR during the second task revealed the predicted interactions, replicating Study 1. After failure feedback—consistent with our hypotheses—participants with unstable high self-esteem exhibited threat relative to those with stable high self-esteem, as did participants with stable low self-esteem. The tendency for participants with unstable high self-esteem to react with relative challenge after success feedback but relative threat after failure feedback was in the predicted direction. Although CO results did not reach significance, their direction was consistent with TPR effects.

In combination with preliminary analyses, the observed challenge–threat effects offer additional support for our argument regarding unstable high self-esteem—that they exhibited relative threat because of underlying self-doubt—but not the previously identified alternative explanations for why individuals with stable high self-esteem might have exhibited relative challenge: (a) They disregarded the feedback information; (b) they disengaged from the task, thus possibly decreasing threat as well; or (c) they were differentially affected by the goal-relevance-heightening instruc-

[10] Given that challenge is typically associated with superior performance relative to threat (e.g., Blascovich et al., 1999), we attempted to test whether TPR and CO reactivity mediated this effect of condition on task performance. However, feedback condition did not significantly predict TPR and CO, excluding them as possible mediators. It is nonetheless worthy of note that the effect of condition on performance remained significant when controlling for TPR and CO, which remained significant and marginal predictors, respectively.

[11] Because Study 1 and Study 2 shared nearly identical designs, it was possible to conduct a meta-analysis. By combining effect sizes from the two studies—allowing for a single significance test for each effect—we were able to increase statistical power to detect effects and further demonstrate the reliability of our findings. To be appropriate for meta-analysis, effects needed to be unambiguous and in the same direction; given the complexity of three-way interactions (e.g., identical interaction term regression coefficients could reflect different patterns of simple effects), we did not include them. Results revealed that the combined effects from Step 2 (level and condition) for both studies did not reach significance. Within Step 4, the combined effect for the predicted two-way interaction between level and stability within the failure condition was significant for both TPR ($z = 3.02$, $p < .01$) and CO ($z = 2.52$, $p < .05$). Testing the three hypothesized simple effects revealed: (a) an effect of stability within high self-esteem in the failure condition, such that greater instability was associated with relative threat, that was significant for TPR ($z = 2.54$, $p < .05$) and marginal for CO ($z = 1.65$, $p = .10$); (b) an effect of level within stable self-esteem in the failure condition, such that lower self-esteem was associated with relative threat, that was significant for both TPR ($z = 2.86$, $p < .01$) and CO ($z = 2.61$, $p < .01$); and (c) an effect of condition within unstable high self-esteem, such that participants exhibited greater threat in the failure condition than in the success condition, that was significant for TPR ($z = 2.43$, $p < .05$) and marginal for CO ($z = 1.80$, $p < .10$).

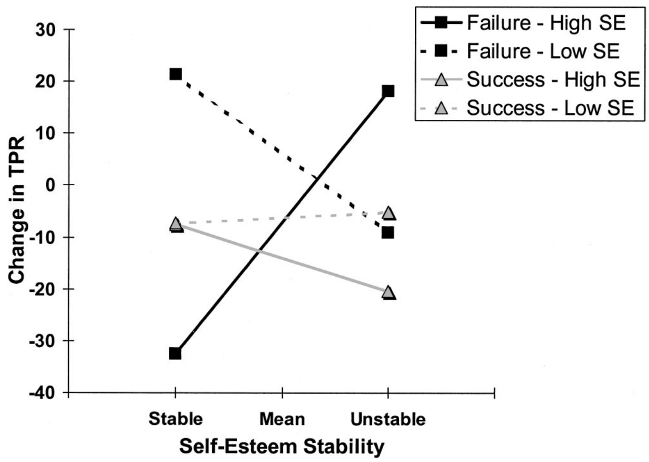

Figure 2. The regression of total peripheral resistance (TPR) reactivity in the second task on self-esteem (SE) level, self-esteem stability, and condition in Study 2; higher TPR reflects greater relative threat. "High SE" and "unstable" represent values 1 standard deviation above the mean of level and stability, respectively, whereas "low SE" and "stable" represent values 1 standard deviation below the mean.

tions immediately before the second task. As in Study 1, manipulation check data (i.e., task difficulty) suggest that all participants were sensitive to the feedback, discounting the first alternative. Also as in Study 1, task engagement data do not support the second alternative. We did find an effect for level, such that higher self-esteem was associated with lower task engagement, but we also found an effect for stability, such that greater stability was associated with higher task engagement. Together, these two additive effects indicate that participants with unstable high self-esteem exhibited the lowest task engagement in the second task. Finally, the absence of the goal-relevance-heightening instructions in Study 2 discounts the third alternative.

General Discussion

Our goal in this investigation was to test the notion that people with unstable high self-esteem possess underlying self-doubt, which is triggered when they are faced with failure. We believe that the present work represents the first direct demonstration of such self-doubt. In addition, we have demonstrated that the responses of individuals with unstable high self-esteem occur spontaneously, without prompting from a self-report questionnaire, and in an actual rather than a hypothetical experience. Because our predictor and criterion variables were assessed with different methodologies, we have also provided additional evidence that statistical stability is a meaningful construct (i.e., it does not merely represent measurement error).

In two studies, participants with unstable high self-esteem in the failure condition exhibited threat relative to participants with stable high self-esteem; no differences emerged in the success condition. As predicted, the reactions of participants with unstable high self-esteem paralleled those of participants with stable low self-esteem, who should also possess self-doubt. In addition, participants with unstable high self-esteem exhibited threat in the failure condition relative to the success condition. Alternative explanations, in contrast, were unable to convincingly account for these results.

In analyses of task engagement, effects of stability and level emerged. In combination, participants with unstable high self-esteem exhibited the lowest task engagement and those with stable low self-esteem exhibited the highest, regardless of feedback condition. These task engagement findings may reflect a preemptive strategy used by individuals with unstable high self-esteem to protect their underlying vulnerability. Unstable self-esteem is characterized by greater sensitivity to events in the environment than stable self-esteem (e.g., Greenier et al., 1999); if these individuals have some awareness that they tend to be sensitive in this way, they may adopt a self-protective approach to an upcoming esteem-relevant task by attempting to disengage from it. A form of self-handicapping, this could soften the blow of failure but accentuate the benefits of success. For example, by essentially convincing oneself that task performance does not matter or at least matters less—thereby decreasing goal relevance—it would be easier to discount failure as a function of not being invested in the task and perhaps not putting forth one's best effort. Conversely, success would seem all the more impressive if it occurred in spite of relative disengagement. This strategy does not seem to be completely effective, however, in that participants with unstable high self-esteem still experienced threat in the face of failure.

Limitations and Future Directions

One question left unanswered by the present research is what role post hoc self-esteem maintenance strategies play in the investigated processes. As already discussed, previous work has demonstrated that individuals with unstable high self-esteem are more likely to exhibit defensive reactions to failure, such as derogating an evaluator (Kernis et al., 1993) or responding with anger (Kernis et al., 1989). We did not explicitly give participants an opportunity to derogate or aggress against the experimenter or another target, so it remains unclear at what point such reactions might occur and

Table 4
Summary of Regression Analysis Predicting CO Reactivity in the Second Task, Study 2

Variable	*B*	*SE B*	β	sr^2
Step 1				
Baseline CO	−0.019	0.028	−.049	.002
HR reactivity	−0.010	0.010	−.075	.006
VC reactivity	0.074	0.012	.436	.183**
Step 2				
Feedback condition	0.124	0.119	.075	.005
Self-esteem level	0.009	0.010	.073	.005
Self-esteem stability	0.003	0.028	.008	.000
Step 3				
Condition × Level	−0.004	0.021	−.027	.000
Condition × Stability	−0.023	0.057	−.044	.001
Level × Stability	−0.001	0.004	−.021	.000
Step 4				
Condition × Level × Stability	0.013	0.008	.204	.012

Note. For the feedback condition variable, failure was set to 0 and success was set to 1. CO = cardiac output; HR = heart rate; VC = ventricular contractility.
** $p < .01$.

whether they are the means by which individuals with unstable high self-esteem attempt to repair damage to their senses of self-worth. It may be that people with stable high self-esteem—who appear to possess deep-seated self-confidence (e.g., Kernis et al., 1997), responding with challenge in the face of failure—are better able to prevent damage to self-esteem, whereas those with unstable high self-esteem (who responded with threat) are left to attempt remedies after the fact. Such post hoc strategies of responding to existing damage would necessarily result in self-esteem fluctuations, contributing to a cycle of perpetual instability. Presumably, it is the underlying self-doubt that we found in individuals with unstable high self-esteem that motivates their previously observed defensive responses, including derogating an evaluator and anger. A reasonable prediction for future research would thus be that the experience of threat mediates these responses.

Considering our task engagement findings, a related prediction for preemptive strategies would be that individuals with unstable high self-esteem would be more likely than those with stable high self-esteem to self-handicap before a self-relevant task. An opportunity to self-handicap should in turn result in lower goal relevance during the task and thus lower task engagement. Interestingly, this prediction implies that individuals with stable high self-esteem may have relatively little use for any such self-esteem maintenance strategies, whether preemptive or post hoc. Supportive results would provide further evidence that their responses are driven by fundamental self-confidence.

Because previous research (e.g., Kernis & Waschull, 1995) has found that people with unstable high self-esteem have unique post hoc strategies (e.g., they are more defensive than others) following failure experiences, it is worthy of note that in the current studies, participants with unstable high self-esteem exhibited a threat pattern similar to those with stable low self-esteem. This suggests that even though both types of participants exhibited threat in our studies, they differed in their awareness of this experience and its causes (e.g., underlying self-doubt) or their responses to it, which then resulted in differential use of post hoc strategies in the face of failure.

The current investigation does little, however, to clarify the difference between stable and unstable low self-esteem, which Kernis and colleagues (e.g., Kernis, 1993; Kernis & Waschull, 1995) have concluded remains less clear than the difference between stable and unstable high self-esteem. Consistent with this previous ambiguity, the simple effect of stability within low self-esteem failed to reach significance in our data. It may be that stability plays a greater role for low self-esteem in contexts other than performance feedback.

Similarly, it is possible that the effects we predicted and observed would not replicate in other domains. We are not aware of any reason they would not, but it remains an empirical question. According to the sociometer hypothesis (Leary, Tambor, Terdal, & Downs, 1995), self-esteem functions as an alarm system sensitive to social inclusion–exclusion, which suggests that social feedback is a particularly important domain for future research.

Finally, we included few self-reported dependent variables in the current studies. We did so because it remains unclear if or to what extent inducing self-reflection by completing a questionnaire affects the self-esteem processes we investigated. In subsequent work, however, other dependent variables (including self-reported ones) may offer additional insight into these processes. Specifically, they might facilitate a better understanding of how individuals with different combinations of level and stability interpret feedback information and what leads them to act (or fail to act) on this information.

Conclusion

Over the past decades, a great deal of time and effort has been invested in exploring the nature of self-esteem and its effects. Much has clearly been learned, but only relatively recently has emphasis shifted to other dimensions of self-esteem beyond high versus low. Self-esteem stability represents one such dimension that has demonstrated substantial utility with the potential for even more. We believe the methodological framework applied in these studies—utilizing theory-driven physiological measures to assess implicit psychological reactions—can continue to offer valuable insight in future self-esteem research.

References

Baumeister, R. F., Smart, L., & Boden, J. M. (1996). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological Review, 103,* 5–33.

Blascovich, J., & Mendes, W. B. (2000). Challenge and threat appraisals: The role of affective cues. In J. P. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition* (pp. 59–82). Paris: Cambridge University Press.

Blascovich, J., Mendes, W. B., Hunter, S. B., Lickel, B., & Kowai-Bell, N. (2001). Perceiver threat in social interactions with stigmatized others. *Journal of Personality and Social Psychology, 80,* 253–267.

Blascovich, J., Mendes, W. B., Hunter, S. B., & Salomon, K. (1999). Social "facilitation" as challenge and threat. *Journal of Personality and Social Psychology, 77,* 68–77.

Blascovich, J., Mendes, W. B., Tomaka, J., Salomon, K., & Seery, M. D. (2003). The robust nature of the biopsychosocial model of challenge and threat: A reply to Wright and Kirby. *Personality and Social Psychology Review, 7,* 234–243.

Blascovich, J., & Tomaka, J. (1991). Measures of self-esteem. In J. P. Robinson, P. R. Shaver, & L. R. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 115–160). New York: Academic Press.

Blascovich, J., & Tomaka, J. (1996). The biopsychosocial model of arousal regulation. *Advances in Experimental Social Psychology, 28,* 1–51.

Brownley, K. A., Hurwitz, B. E., & Schneiderman, N. (2000). Cardiovascular psychophysiology. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 224–264). New York: Cambridge University Press.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change"—Or should we? *Psychological Bulletin, 74,* 68–80.

Dienstbier, R. A. (1989). Arousal and physiological toughness: Implications for mental and physical health. *Psychological Review, 96,* 84–100.

Greenier, K. D., Kernis, M. H., McNamara, C. W., Waschull, S. B., Berry, A. J., Herlocker, C. E., & Abend, T. A. (1999). Individual differences in reactivity to daily events: Examining the roles of stability and level of self-esteem. *Journal of Personality, 67,* 185–208.

Hunter, S. B. (2001). *Performance under pressure: The impact of challenge and threat states on information processing*. Unpublished doctoral dissertation, University of California, Santa Barbara.

Kelsey, R. M., & Guethlein, W. (1990). An evaluation of the ensemble averaged impedance cardiogram. *Psychophysiology, 28,* 24–33.

Kernis, M. H. (1993). The roles of stability and level of self-esteem in psychological functioning. In R. F. Baumeister (Ed.), *Self-esteem: The puzzle of low self-regard* (pp. 167–182). New York: Plenum Press.

Kernis, M. H. (2003). Toward a conceptualization of optimal self-esteem. *Psychological Inquiry, 14,* 1–26.

Kernis, M. H., Cornell, D. P., Sun, C., Berry, A., & Harlow, T. (1993). There's more to self-esteem than whether it is high or low: The importance of stability of self-esteem. *Journal of Personality and Social Psychology, 65,* 1190–1204.

Kernis, M. H., Grannemann, B. D., & Barclay, L. C. (1989). Stability and level of self-esteem as predictors of anger arousal and hostility. *Journal of Personality and Social Psychology, 56,* 1013–1022.

Kernis, M. H., Grannemann, B. D., & Mathis, L. C. (1991). Stability of self-esteem as a moderator of the relation between level of self-esteem and depression. *Journal of Personality and Social Psychology, 61,* 80–84.

Kernis, M. H., Greenier, K. D., Herlocker, C. E., Whisenhunt, C. R., & Abend, T. A. (1997). Self-perceptions of reactions to doing well or poorly: The roles of stability and level of self-esteem. *Personality and Individual Differences, 22,* 845–854.

Kernis, M. H., & Waschull, S. B. (1995). The interactive roles of stability and level of self-esteem: Research and theory. *Advances in Experimental Social Psychology, 27,* 93–141.

Kernis, M. H., Whisenhunt, C. R., Waschull, S. B., Greenier, K. D., Berry, A. J., Herlocker, C. E., & Anderson, C. A. (1998). Multiple facets of self-esteem and their relations to depressive symptoms. *Personality and Social Psychology Bulletin, 24,* 657–668.

Leary, M. R., Tambor, E. S., Terdal, S. K., & Downs, D. L. (1995). Self-esteem as an interpersonal monitor: The sociometer hypothesis. *Journal of Personality and Social Psychology, 68,* 518–530.

Llabre, M. M., Spitzer, S. B., Saab, P. G., Ironson, G. H., & Schneiderman, N. (1991). The reliability and specificity of delta versus residualized change as measures of cardiovascular reactivity to behavioral challenges. *Psychophysiology, 28,* 701–711.

McFarlin, D. B., & Blascovich, J. (1984). On the Remote Associates Test (RAT) as an alternative to illusory performance feedback: A methodological note. *Basic and Applied Social Psychology, 5,* 223–229.

Mendes, W. B., Blascovich, J., Lickel, B., & Hunter, S. B. (2002). Challenge and threat during social interactions with White and Black men. *Personality and Social Psychology Bulletin, 28,* 939–952.

Mendes, W. B., Blascovich, J., Major, B., & Seery, M. D. (2001). Challenge and threat responses during downward and upward social comparisons. *European Journal of Social Psychology, 31,* 477–497.

Papillo, J. F., & Shapiro, D. (1990). The cardiovascular system. In J. T. Cacioppo & L. G. Tassinary (Eds.), *Principles of psychophysiology: Physical, social, and inferential elements* (pp. 456–512). New York: Cambridge University Press.

Sherwood, A., Allen, M. T., Fahrenberg, J., Kelsey, R. M., Lovallo, W. R., & van Dooren, L. J. P. (1990). Methodological guidelines for impedance cardiography. *Psychophysiology, 27,* 1–23.

Williams, R. B., Barefoot, J. C., & Shekelle, R. B. (1985). The health consequences of hostility. In M. A. Chesney & R. H. Rosenman (Eds.), *Anger and hostility in cardiovascular and behavioral disorders* (pp. 173–185). Washington, DC: Hemisphere.

Received May 28, 2003
Revision received February 4, 2004
Accepted February 16, 2004 ■