Predicting Return of Fear Following Exposure Therapy

With an Implicit Measure of Attitudes

Michael W. Vasey*, Casaundra N. Harbaugh, Adam G. Buffington,

Christopher R. Jones, & Russell H. Fazio

Department of Psychology, The Ohio State University,

1835 Neil Avenue, Columbus, OH 43210-1222, USA

Running head: Predicting Return of Fear with an Implicit Measure of Attitudes

* Corresponding author. Department of Psychology, The Ohio State University, 1835 Neil
Avenue, Columbus, OH 43210-1222, USA. Tel.: +1 614 292 2951; fax: +1 614 688 8261
Phone: (614) 292-2951; email: vasey.1@osu.edu

Abstract

We sought to advance understanding of the processes underlying the efficacy of exposure therapy and particularly the phenomenon of return of fear (ROF) following treatment by drawing on a social psychological view of phobias as attitudes. Specifically, a dual process theory of attitude-related behavior predicts that a positive response to exposure therapy may reflect change in either the automatic (the attitude representation itself) or controlled (skills and confidence at coping with the fear) responses to the phobic stimulus, or both. However, if the attitude representation remains negative following treatment, ROF should be more likely. We tested this hypothesis in a clinical sample of individuals with public speaking phobia using a single-session exposure therapy protocol previously shown to be efficacious but also associated with some ROF. Consistent with predictions, a post-treatment implicit measure of attitudes toward public speaking (the Personalized Implicit Association Test [PIAT]) predicted ROF at 1-month follow-up. These results suggest that change in the automatically activated attitude toward the phobic stimulus is an important goal of exposure therapy and that an implicit measure like the PIAT can provide a useful measure of such change by which to gauge the adequacy of exposure treatment and predict its long-term efficacy.

Predicting Return of Fear Following Exposure Therapy

With an Implicit Measure of Attitudes

After decades of research documenting the efficacy of exposure therapy for phobic anxiety disorders (Choy, Fyer, & Lipsitz, 2007), much about the technique remains inadequately understood (Moscovitch, Antony, & Swinson, 2009). This is perhaps no more clearly demonstrated than by the fact that it remains difficult to determine when gains achieved during treatment will persist or fade over time (Craske et al., 2008). Despite showing excellent reductions in fear in the context of therapy, many treatment responders will, eventually, experience some degree of "return of fear" (ROF; Rachman, 1989). ROF is defined simply as the reappearance of a fear that was previously present but had undergone a decline. Although ROF typically does not amount to full clinical relapse, it is a significant problem in a substantial percentage of cases and the phenomenon is not well understood (Craske & Mystikowski, 2006). In particular, predictors of ROF's occurrence have proven elusive (e.g., see Craske et al., 2008).

Translational research drawing on advances in learning theory (see Craske, Hermans, & Vansteenwegen, 2006) and neuroscience (see McNally, 2007) has led to significant progress in understanding of ROF. In our view, work drawing on advances in social psychology has similar promise. From a social cognitive perspective on attitudes (see Fazio, 2007), phobias are essentially inappropriately negative and highly accessible attitudes toward objects (e.g., spiders) or situations (e.g., public speaking). Because the basic science of attitudes and attitude-related behavior is well developed, it has considerable potential to advance understanding of the processes underlying the efficacy of exposure therapy and the durability of its effects. Specifically, we believe conceptualizing phobias from the perspective of a dual process theory of attitude-related behavior, such as the MODE model (Fazio, 1990; Olson & Fazio, 2009), can help to explain ROF's occurrence.

The MODE model distinguishes two classes of attitude-to-behavior processes. The difference centers on the extent to which pursuing a particular course of action involves a spontaneous reaction to one's perception of the immediate situation versus deliberation regarding the behavioral alternatives. Thus, MODE is an acronym for Motivation and Opportunity as DEterminants – determinants of whether attitude-related behavior is primarily spontaneous versus controlled in nature. The model postulates that attitudes can guide behavior in a spontaneous manner, without the individual actively considering the relevant attitude and without the individual's necessary awareness of its influence. Instead, the attitude may be activated from memory automatically upon the individual's encountering the attitude object and influence how the person construes the object in the immediate situation. Ultimately, this construal will affect the person's behavioral response. Alternatively, attitudes may guide behavior through a more deliberative process. Individuals analyze the costs and benefits of a particular behavior and deliberately reflect upon the attitudes relevant to the behavioral decision so as to arrive at a behavior plan, which then may be enacted (see Ajzen, 1991). Given the effortful reflection required by the deliberative alternative, some motivating force is necessary to induce individuals to engage in such reasoning. The time and the resources to deliberate (i.e., the opportunity) also must exist.

In addition to delineating two distinct classes of attitude-behavior processes, the MODE model explicitly postulates the possibility of "mixed" processes that involve both automatic and controlled components. Any controlled component within a mixed sequence requires, once again, that the individual be both motivated to engage in the necessary effort and have the opportunity to do so. Thus, some individuals might be motivated to check on the appropriateness, or even counter the influence, of an automatically-activated attitude. That

motivation might stem from an enhanced desire for accuracy, a sense of accountability, a concern with social desirability, or, as in research concerning racial prejudice, a motivation to control for prejudiced reactions. Within the MODE model, opportunity is essentially a gating mechanism. One must have the self-regulatory resources and the time for such motivated processing. Provided the opportunity exists, an individual can counter or even correct for the influence of the automatically-activated attitude. So, for example, in the domain of racial attitudes, evidence shows that some individuals are troubled by the activation of a negative racial attitude and, provided the opportunity exists, will work very hard to correct for that unwanted influence. However, if the individual is fatigued or cognitively taxed in some way, or if the situation demands an immediate response, there will be little opportunity to engage in motivated deliberation. In that case, judgments or behavior tend to be influenced by the automatically activated attitude, regardless of any motivational concerns (see Olson & Fazio, 2009, for a review).

In our view, the MODE model has important implications for understanding exposure therapy. As noted earlier, many individuals who exhibit improvement following treatment nonetheless experience ROF. The MODE model leads us to raise the question: what changed during treatment in such individuals – the automatic or the controlled component, or both? Did the individual's attitudinal representation of the feared object change, or just the individual's skills at controlling the automatically-activated fear? Although progress of the latter sort is certainly desirable, if the attitude remains intact, the negativity will continue to be activated whenever the feared object or situation is encountered. This may happen at a time when the individual is fatigued or in a context in which he/she believes him/herself incapable of effective control, in other words, when the opportunity for successful effortful control is lacking.

Moreover, the very experience of anxiety and the efforts to regulate it have the potential to exacerbate any such fatigue and interfere with performance (Jones, Fazio, & Vasey, in press). This may erode the individual's confidence regarding the management of the fear and foster avoidance and safety behaviors that promote relapse.

Interestingly, a similar implication can be derived from cognitive models that distinguish between the automatic and strategic processes that play a role in anxiety (e.g., Beck & Clark, 1997; Ouimet, Gawronski, and Dozois, 2009). Like the MODE model, these information processing models point to the possibility of the automatic component exerting an especially dominant role when the opportunity for strategic processing is lacking and, hence, imply that treatment must foster change in the automatic processes. As we have argued, however, the MODE model's grounding in the attitude literature highlights the significance of the attitudinal representation in memory and offers a clear prediction regarding the value of promoting change in the representation for limiting the potential for ROF and, importantly, of measuring the extent of such attitude change to predict its occurrence.

To summarize, the MODE model predicts that successful exposure therapy requires, not just the development of fear-management skills and the confidence that one can control the fear, but attitude change. Thus, what is needed is a way to measure the attitude representation of the phobic object following treatment. Drawing on findings from social psychology, a number of scholars have suggested that implicit measures of attitudes may prove useful in this regard (e.g., Hermans et al., 2002; Teachman & Woody, 2003).

Insofar as implicit measures tap the underlying attitude representation regarding the phobic stimulus, ROF should be likely when negative automatically activated attitudes toward the phobic stimulus persist following exposure therapy. In addition to such well-documented

problems with self-reported assessments as their sensitivity to demand characteristics and social

desirability concerns, people who have just undergone some treatment for an anxiety disorder

may not be very good at reporting any success that they have achieved. They simply may not be

very well-calibrated at estimating their improvement across time or the extent of their fear

relative to what is normative. Consistent with this view, Craske et al. (2008) note that client

judgments of their own competence during exposure therapy are not good predictors of actual

competence at a later re-testing. Indeed, they suggest that the typical approach to exposure

therapy (i.e., one involving massed training and constancy of training conditions) fosters

illusions of competence. For all these reasons, the estimates of automatically-activated attitudes

provided by an appropriately designed implicit measure should serve as valuable benchmarks of

treatment progress.

   A variety of evidence shows that implicit measures do indeed differentiate those with

phobias from non-phobic controls (see Roefs et al., 2011). Importantly, such measures typically

account for significant variance in behavioral approach task (BAT) performance above and

beyond that accounted for by explicit measures (e.g., Asendorpf, Banse, & Mucke, 2002;

Ellwart, Rinck, & Baker, 2006; Teachman & Woody, 2003; but see Huijding & de Jong, 2007).

However, although some evidence suggests that implicit measures may be sensitive to change

following exposure treatment (Gamer et al., 2008; Teachman & Woody, 2003), such effects have

not always been found (Huijding & de Jong, 2007; 2009; Teachman & Woody, 2003) and when

found the effect is typically small and may reflect practice rather than treatment effects (see

Roefs et al., 2011). Thus, it has been difficult to discern whether the treatment did not produce

attitude change or the implicit measure was insensitive to such change. Moreover, the only study

to test an implicit measure as a predictor of ROF failed to find a significant effect (Huijding & de

Jong, 2009). In that case, however, it should be noted that the implicit measure also did not reveal significant change following treatment, likely limiting its ability to predict ROF.

Although the failure to find consistent evidence of exposure effects using implicit measures may simply reflect their absence, it may also reflect limitations of the specific measures used. For example, the attribute response labels (e.g., pleasant/unpleasant) used in the traditional form of the most widely employed implicit measure, the Implicit Association Test (IAT), have been shown to be problematic when the IAT is intended to measure attitudes. The traditional IAT involves sorting four categories of stimuli (e.g., a race IAT may involve pictures of pleasant and unpleasant stimuli with pictures of White and Black faces) using only two keys. Performance depends on how easily the participant can cope with the dual meanings of the response keys (e.g., Black/pleasant and White/unpleasant in some blocks versus Black/unpleasant and White/pleasant in other blocks). For example, a negative attitude toward Blacks would be demonstrated by faster responses when Black and unpleasant are mapped onto the same key than when Black and pleasant are represented by the same key. However, because the typical evaluative labels are inherently ambiguous with respect to perspective (pleasant or unpleasant to whom?), IAT performance is easily affected by how the labels are interpreted (see Han, Czellar, Olson, & Fazio, 2010). Although the respondent's attitudes can affect the ease of accommodating to a particular response mapping, contextually salient information that does not form the basis for the individual's attitude also can do so, for example, by encouraging adoption of the perspective of society as whole (Olson & Fazio, 2004) or of some specific individual with an attitude that differs from one's own (Han, Olson, & Fazio, 2006). Therefore, traditional implementations of the IAT can be influenced by factors other than the attitudinal associations that the procedure aims to measure, resulting in mis-estimation of an individual's attitude.

Recent research has demonstrated that this malleability of the IAT labels can lead to the inference that attitude change has occurred even when there is very good reason to believe it has not, and conversely, can obscure attitude change when such change is indeed likely to have occurred (Han et al., 2010).

Recent research also suggests that focusing the IAT on personal attitudes by using more specific evaluative labels can obviate these concerns. Simply modifying the response labels from "pleasant/unpleasant" to "I like/I don't like," and thus personalizing the IAT, provides a more precise measure of individuals' personal evaluations. This personalized IAT (PIAT) has proven more sensitive to the presence or absence of attitude change (Han et al., 2010) and has yielded stronger correspondence with past behavior and behavioral intentions (Olson & Fazio, 2004) than the traditional implementation.

On this basis, we developed and refined a PIAT aimed at assessing attitudes toward public speaking and used it to test for change in the attitudinal representation following exposure therapy and to predict ROF in individuals with a phobia of public speaking. Specifically, this test was conducted in the context of a single session exposure treatment developed by Tsao and Craske (2000), who found it to produce significant reductions in fear of public speaking but also to be associated with significant ROF over a follow-up interval of 1-month. Our focus is on whether the snapshot of the attitudinal representation provided by a PIAT administered immediately following treatment proves predictive of any ROF evident at the 1-month follow-up session.

Method

Participants

Participants were 40 adults (60% female) ranging in age from 18-46 years (*M* = 22.4, *SD* = 5.7). Most members of the sample self-identified as Caucasian (70%), with 17.5% Asian-American, 7.5% Hispanic, and 5% Bi- or Multi-Racial. Based on administration of the Structured Clinical Interview for DSM Disorders (SCID; First, Spitzer, Gibbon, & Williams, 1995), all participants met diagnostic criteria at minimum for Social Anxiety Disorder (SAD) in the context of public speaking, with 20% also meeting criteria for SAD–Generalized Type. Exclusion criteria were: 1) Current major depressive disorder or suicidal ideation; 2) Current or past diagnosis of bipolar disorder, schizophrenia, other psychosis, or organic mental syndrome; or 3) Current psychosocial treatment. Participants could be taking anxiolytic or antidepressant drugs, provided dose was stable during the study.[1]

Participants were recruited in two ways. First, students in introductory psychology classes completed the Personal Report of Confidence as a Speaker (PRCS; Paul, 1966) as part of a screening battery. Those with scores ≥ 26 (94th percentile; Phillips, Jones, Rieger, & Snell, 1997) were telephoned and offered the opportunity to participate in a study in which they would receive treatment for their fear of public speaking. Those indicating interest completed a screening interview to assess likely eligibility. Second, fliers describing the study were posted around the campus of a large Midwestern university. Potential participants responded by phone and completed a screening interview designed to assess likely eligibility.

Measures

Self-Report.

Personal Report of Confidence as a Speaker (PRCS; Paul, 1966). The PRCS is a widely used measure of public speaking anxiety, which consists of 30 true/false items, some of which

---

[1] Three participants (7%) were doing so.

are reverse-scored. Scores can range from zero (no fear) to 30 (extreme fear). It includes such face-valid items as "I prefer to have notes on the platform in case I forget my speech," "I feel relaxed and comfortable while speaking," and "I look forward to the opportunity to speak in public." A large body of evidence supports the PRCS's reliability and validity (for a review see Phillips et al., 1997). In the current study, Cronbach's alpha ranged from .91-.94 across time points.

Brief Fear of Negative Evaluation Scale (BFNE; Leary, 1983). The BFNE is a 12-item scale measuring anxiety concerning social evaluation. Items are rated on a 5-point Likert scale ranging from 1 (not at all characteristic of me) to 5 (extremely characteristic of me). The scale has been shown to have good reliability and validity (Collins, Westra, Dozois, & Stewart, 2005; Leary, 1983). In the current study, Cronbach's alpha ranged from .94-.96 across time points.

State-Trait Anxiety Inventory – State Version (STAI [Form Y]; Spielberge, Gorsuch, Lushene, Vagg, & Jacobs, 1983). The STAI – State Version is a 20-item scale designed to measure anxiety at the present moment. Items are rated on a 4-point Likert scale ranging from 1 (not at all) to 4 (very much so). The scale has been shown to have excellent reliability and validity (Spielberger, 1989). In the current study, Cronbach's alpha ranged from .92-.97 across time points.

Behaviours Checklist (BCL; Mansell & Clark, 1999). The BCL is an 18-item self-report scale on which participants rate how anxious they think they appeared during a speech and the extent to which they would be judged by others to have shown various positive (e.g., confident, self-assured) and negative (e.g., awkward, voice quivering) characteristics. Ratings on each characteristic are made on a 9-point Likert scale ranging from 0 (not at all) to (extremely). Positive items are reversed and a total score is computed, which has been shown to have good

internal consistency and validity (e.g., Wild, Cark, Ehlers, & McManus, 2008). Cronbach's alpha

ranged from .93-.94 across time points.

Physiological.

Heart rate (HR) was recorded continuously during each BAT using a Polar ambulatory

heart-rate monitor attached to a standard chest belt. Data were transmitted to a model RS800

Polar watch. Mean HR was computed for each BAT. Due to equipment malfunction, HR data

were lost for two participants at pre- and post-treatment and 4 participants at follow-up.

Behavioral.

In each BAT, participants were given 3-minutes to prepare a 5-minute speech on two

topics selected at random without replacement from a set of 16 topics modeled after those used

by Tsao and Craske (2000). All speeches were delivered without notes in the presence of the

experimenter and while facing a video camera. The second of the two BATs at follow-up also

included a live audience of three research assistants. Immediately before and at 1-minute

intervals during each BAT, participants provided Subjective Units of Distress (SUDS; Wolpe,

1969) ratings on a scale ranging from 0 (no anxiety) to 100 (extreme anxiety).

Five undergraduate research assistants independently rated the overall quality and

effectiveness of the speeches. A video recording of each speech was viewed by a random set of

three of these judges. Raters were kept blind to participants' scores on other measures and the

point in the sequence of sessions when a given speech occurred. Each speech was rated on two

questions: (1) "How would you rate the overall quality of the speech?" and (2) "How would you

rate the presenter's effectiveness as a speaker?" Ratings were made on an 11-point scale ranging

from 0 (very poor) to 10 (very high). Because the correlation between the two ratings was very

high (r = .89), they were averaged to create a single index of speech quality. ICCs for these

average ratings ranged from .82 to .91. Jones, Fazio, and Vasey (in press) found this score to correlate significantly with speakers' self-ratings of fear of public speaking and speech performance. Similarly, in the current study, for the pre-treatment BAT, the average speech quality rating correlated significantly with participants' self-ratings on the BCL (r = -.63) and their maximum SUDS ratings during each BAT (r = -.48).

PIAT. The PIAT focused on the strength of personal associations between "public speaking" and "I like" versus "I don't like." Because there is no obvious contrast category to public speaking, the implementation involved the single category version of the IAT (Karpinski & Steinman, 2006; Wigboldus, Holland, & van Knippenberg, 2005), as well as personalization. During the critical blocks, participants were presented with three classes of images – public speaking (e.g., a lectern, a microphone, an audience), positively-valued objects (e.g., an ice cream sundae, a gift box, picturesque scenery), and negatively-valued (e.g., polluting smokestacks, burnt toast) and classified them using two response keys. In some blocks, one key was labeled "I Like" and "Public Speaking" and the other key "I Don't Like" whereas other blocks involved "I Don't Like" and "Public Speaking" sharing a response key. The task consisted of 10 blocks of 30 trials each. The first two blocks were for practice. In the remaining eight blocks, public speaking was mapped onto the "I like" key in four blocks and onto the "I don't like" key in the remainder. The two types of blocks were administered in counterbalanced order. Instructions regarding the meaning of the two keys were presented at the beginning of each block. The critical comparison involves the difference in latencies for the two response mappings, with higher scores indicating a more negative attitude, as shown by faster latencies when "Public Speaking" and "I Don't Like" were mapped on to the same response key than when "Public Speaking" and "I Like" were mapped on to the same key. The eight blocks of

critical trials thus produced four scores representing this critical difference, which were averaged to yield a total PIAT score. Reliability of this score was evaluated by computing Cronbach's alpha for the four difference scores. Alpha averaged .66 across the three PIAT tasks.

Procedure

The study involved three sessions. In Session 1 potentially eligible individuals provided informed consent and completed the PIAT followed by the SCID, which was administered by two doctoral students in clinical psychology trained to a minimum of 80% agreement. Inter-rater agreement in this sample was very good (Kappa = .82). Those meeting eligibility requirements were invited to attend Session 2 (typically within 2-weeks of Session 1), in which the exposure treatment was administered by two clinical psychology doctoral students. Session 2 typically lasted about 2-hours. During that session, participants completed the PRCS and BFNE, and immediately before completing the pre-treatment speech BAT, the STAI. Immediately following the speech BAT, they completed the BCL. They then completed the exposure treatment. Following treatment, participants again completed the PRCS, BFNE, the STAI, the post-treatment speech BAT, the BCL, and finally, the PIAT. After 1-month, participants returned for Session 3, at which they again completed the PIAT, followed by the PRCS, BFNE, and two BATs, each preceded by the STAI and each followed by the BCL. The 1-month follow-up interval was chosen because Tsao and Craske (2000) found significant ROF over that interval using the same treatment.

Exposure Treatment: The exposure treatment was conducted according to a detailed manual based on the massed exposure condition used by Tsao and Craske (2000). After receiving didactic information about public speaking anxiety and a treatment rationale emphasizing the importance of exposure, participants completed four exposure trials during

which they were given 2-minutes to prepare prior to delivering 5-minute speeches. For each speech, participants drew two topics randomly and without replacement from the same set of topics used for the BATs. SUDs ratings were obtained pre-speech and at 1-minute intervals thereafter. Each speech was presented without the use of notes before the therapist and an audience of three research assistants (trained to remain impassive and refrain from offering encouragement [e.g., nodding], or discouragement [e.g., frowning]). Therapists intervened only to obtain SUDs ratings and to encourage the participant to continue speaking if he/she stopped prematurely. All treatment was supervised by a licensed clinical psychologist and treatment sessions were observed either live or via video recording by the first author to monitor treatment fidelity.

Following Session 3, participants were offered an opportunity to receive further exposure therapy if desired. Those who did not complete the study (e.g., due to being deemed ineligible or declining treatment) were offered assistance in finding appropriate services.

<div align="center">Results</div>

Did Fear Decline Significantly Following Treatment?

Paired *t*-tests revealed significant improvement from pre- to post-treatment on all measures (Table 1). Most notably, PIAT scores were significantly reduced, indicating a less negative attitude representation following treatment.

Did ROF Occur at Follow-up?

ROF was assessed at both the average and the individual levels. As shown in Table 1, on average, significant increases from post-treatment to follow-up were seen in anticipatory STAI scores for BAT 2 but not BAT 1. A similar pattern emerged for pre-speech SUDs ratings although only at a trend level. Similarly, ROF on average was seen for maximum SUDs rating

during BAT 2 but not BAT 1. Finally, ROF was seen for mean heart rate during both BATs. In contrast, gains were maintained on average from post-treatment to follow-up on the PIAT and observer ratings of speech quality, while further improvement was seen in BFNE scores.

At the individual level, the percentage of participants whose scores at follow-up reflected an increase in fear (or a decrease in speech quality) from post-treatment levels ranged from 25% to 66%, with an average of 49.2% across measures. Thus, there was sufficient variability in ROF to support regression analyses designed to test the PIAT as a predictor of ROF.

Predictors of ROF at Follow-Up

We computed ROF scores for each measure at follow-up by subtracting the post-treatment score from the follow-up score. Thus, positive values on these ROF scores indicated an increase from post-treatment to follow-up. To determine if post-treatment PIAT scores predicted outcomes at follow-up above and beyond explicit measures of public speaking anxiety at post-treatment, each of the ROF scores at follow-up served as the dependent variable (DV) in a regression analysis in which the predictors were the PIAT, the pre- and post-treatment versions of the DV and the post-treatment PRCS and BFNE scores. The latter two measures were included to clarify if the implicit measure accounted for variance above and beyond explicit measures of social anxiety. Each of the predictor variables was standardized to ease interpretation of the results.

Because the pattern of results for these regression analyses did not differ for the two BATs at follow-up, to simplify presentation, all variables were averaged across those BATs when computing the ROF scores. Similarly, because the anticipatory STAI state anxiety scores and SUDs ratings were strongly correlated across BATs (average $r = .67$), we created an anticipatory anxiety index by combining those scores for each BAT. For pre- and post-treatment

these scores were simply standardized and averaged. However, to create an interpretable ROF score at follow-up, scores were standardized using the post-treatment mean and standard deviation. Thus, a positive score on this index at follow-up indicates an increase from post-treatment to follow-up in post-treatment SD units.

Results of the regression analyses are summarized in Table 2. Unsurprisingly, in most cases the best predictor of a given variable at follow-up was the same variable at either pre- or post-treatment. However, as shown in Table 2, post-treatment PIAT scores accounted for significant variance in the case of three DVs: Anticipatory Anxiety, BCL, and Mean HR. In each case, the regression coefficient makes it clear that, for an individual scoring at the average on the other predictor variables (i.e., z-scores equaling zero) PIAT scores that are above average (i.e., more negative) predict ROF scores greater than zero. Because the intercept in each case is also above zero, these positive ROF scores indicate increases in fear from post-treatment to follow-up. For example, in the case of Mean HR, the predicted ROF score for an individual scoring at the average for all predictors, including the PIAT is simply the intercept (i.e., 4.23 beats per minute [BPM]). Thus, for Mean HR, the average individual is predicted to show significant ROF because the intercept is significantly greater than zero. But for the same individual with a PIAT score of +1 SD (i.e., +55.4 ms) the predicted ROF score would be 4.81 BPM higher. Similarly, for the BCL a PIAT score of +1 SD predicts an increase of 5.32 points beyond the average level of 3.62 indexed by the intercept. For the Anticipatory Anxiety Index, a PIAT score of +1 SD predicts an increase of .20 SDs beyond an average ROF of .14 SDs.

The PIAT also approached significance for the BFNE score, with a PIAT score of +1 SD predicting an increase of 1.76 points but on average ROF was -2.28. Thus, whereas the average

individual's BFNE score declined significantly from post-treatment to follow-up, there was a trend for that to be less likely for those with higher PIAT scores.

Correlates of Post-Treatment PIAT Scores

To investigate how change on the PIAT following treatment related to change on other outcome measures, following Teachman and Woody (2003), we examined correlations among residualized change scores on the PIAT and other dependent measures. Change on the PIAT was not significantly predicted by change on any other measure following treatment (*r*s ranged from -.11 to .14, *p*s > .39).

Discussion

In this study we sought to advance understanding of the processes underlying the efficacy of exposure therapy and particularly the phenomenon of ROF by drawing on a social psychological view of phobias as attitudes. Specifically, based on the MODE model (Fazio, 1990), we predicted that individuals completing exposure therapy would vary in the degree to which their attitude representation of the phobic situation was successfully modified as a result of the learning opportunities afforded by exposure. To the extent that the attitude representation remains predominantly negative following treatment, ROF should be more likely. We tested this hypothesis in a sample of individuals with public speaking phobia using an implicit measure of attitudes toward public speaking (i.e., the PIAT) designed to minimize the influence of extrapersonal factors and thereby enhance its sensitivity to attitude change.

Consistent with the findings of Tsao and Craske (2000) using the same treatment that we employed, participants showed significant improvement following treatment on a wide range of measures. It is notable that participants, on average, showed improvement not only on self-report measures, but also on a measure of heart rate during a speech, observer ratings of speech quality,

and especially on the PIAT. Furthermore, whereas the reduction in PIAT scores was maintained at follow-up, ROF was evident on a variety of other measures. Thus, this study provided an excellent context for testing our hypotheses regarding the predictive value of the PIAT. Consistent with expectations, in the case of three DVs, the post-treatment PIAT accounted for significant variance in ROF even after controlling for pre- and post-treatment scores on the DV as well as two explicit measures of public speaking fear. Specifically, this was true for pre-speech anticipatory anxiety, post-speech beliefs about observers' evaluations (i.e., BCL scores) and mean HR during the speeches at follow-up. That this was true of an objective physiological measure of anxiety is particularly noteworthy. In contrast, scores on explicit outcome measures at follow-up were generally not significantly related to post-treatment scores on explicit measures of fear of public speaking (i.e., the PRCS and BFNE).

In sum, these results provide clear support for the predictions derived from the MODE model. In the month following successful exposure treatment some individuals experienced ROF and a "snapshot" of the relevant attitudinal representation (the PIAT) taken immediately after treatment was able to predict significant variance in its occurrence. This implies that improvement during treatment is accompanied by attitude change for some individuals. Others may have learned to control their fear but not have experienced a change in the attitudinal representation and, hence, continue to have negativity automatically activated when encountering the fearful situation.

It is important to note that these results can also be understood within the context of a model of exposure therapy that emphasizes the centrality of inhibitory learning to the process of extinction (Bouton, Woods, Moody, Sunsay, & Garcia-Gutierrez, 2006). However, as noted by Craske et al. (2008), the strength of such inhibitory associations acquired is not well indexed by

fear levels expressed during or immediately following extinction. Consequently, a mismatch is often seen between the level of fear immediately following exposure and that seen in a different context or at a later time. From this perspective, the PIAT appears to provide a better measure of the strength and/or breadth of the inhibitory associations acquired through exposure than do explicit measures. In keeping with past research seeking to predict ROF (see Craske et al., 2008), measures of fear reduction following treatment were unrelated to change on the PIAT.

Whereas a variety of processes may weaken inhibitory learning over time, it is important to note that the PIAT predicted the occurrence of ROF before any time had passed. Thus, ROF in this study appears to have been a function of inadequate inhibitory learning during exposure that was not detected by explicit outcome measures or by a measure of physiological arousal or objective ratings of speech quality. It will be important for future research to identify predictors of, and more generally strategies for promoting, attitude change (i.e., inhibitory learning). For example, inhibitory learning is likely to be enhanced when the exposure treatment context maximizes client expectancies for negative outcomes so as to create the greatest possible mismatch between expected and actual outcomes (Craske et al., 2008). This is most likely when exposure occurs in the presence of as many predictors of the expected aversive outcome as possible. It may be that those showing the greatest attitude change were those for whom the standardized exposure context made the best contact with the complex of stimuli associated with their expectancy of negative outcomes. Similarly, those showing the least attitude change may be those for whom the standard exposure context may have included one or more safety signals or those whose safety behaviors were insufficiently limited during the standardized exposure format (Craske et al., 2008). Thus, the PIAT appears to provide a useful metric by which to gauge the

adequacy of inhibitory learning achieved through exposure and to predict long-term treatment efficacy.

It is noteworthy that the PIAT proved to be most predictive of ROF in variables related to anticipatory anxiety (i.e., pre-speech SUDS and STAI State Anxiety) and post-event processing (i.e., beliefs endorsed on the BCL that one will be judged to have appeared very nervous and to have delivered a poor quality speech). As emphasized by authors like Wells and Clark (1997) and Hofmann (2007), anticipatory anxiety and post-event processing are central contributors to the maintenance of social anxiety. Anticipatory anxiety drives avoidance and, if avoidance is impossible, safety behaviors are used to minimize the threat posed by the situation. This suggests the possibility that incomplete attitude change (i.e., inadequate inhibitory learning) may lead to ROF because anticipatory anxiety leads to avoidance of and/or failure to seek out further exposure experiences during the follow-up interval. This and similar mediators of the link between post-treatment PIAT scores and ROF should be considered in future research.

It is important to consider several limitations of the study. First, the sample was relatively small and therefore caution is warranted in generalizing to the broader population. Second, although participants met DSM-IV criteria for social anxiety disorder, severity in this sample was generally moderate. Only about 20% met the more stringent criteria for generalized social anxiety disorder. Obviously, a more severe group would likely show less response to a single session of treatment. However, we see no reason to believe that the PIAT would have less value as an index of attitude change in such a sample.

In conclusion, results of this study suggest not only that change in the attitude representation regarding the phobic stimulus is an important goal of exposure therapy but that the PIAT appears to provide an efficient means of measuring such change. However, it should be

noted that this does not imply that change on explicit measures is unimportant. Rather in our

view the MODE model implies that such change is likely to be insufficient by itself. That view is

confirmed by the current study.

Acknowledgments

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179-211.

Asendorpf, J.B., Banse, R., & Mucke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83*, 380-393.

Beck, A. T., & Clark, D. A. (1997). An information processing model of anxiety: Automatic and strategic processes. *Behaviour Research and Therapy, 3*, 49-58.

Bouton, M. E., Woods, A. M., Moody, E. W., Sunsay, C., & Garcia-Gutierrez, A. (2006). Counteracting the context-dependence of extinction: Relapse and tests of some relapse prevention methods. In M. G. Craske, D. Hermans, & D. Vansteenwegen (Eds.) *Fear and learning: From basic processes to clinical implications*. Washington, DC: American Psychological Association.

Choy, Y., Fyer, A. J., & Lipsitz, J. D. (2007). Treatment of specific phobia in adults. *Clinical Psychology Review, 27*, 266-286.

Collins, K. A., Westra, H. A., Dozois, D. J. A., & Stewart, S. H. (2005). The validity of the brief version of the Fear of Negative Evaluation Scale. *Journal of Anxiety Disorders, 19*, 345-359.

Craske, M. G., Hermans, D., & Vansteenwegen, D. (2006). *Fear and learning: From basic processes to clinical implications*. Washington, DC: American Psychological Association.

Craske, M. G., Kircanski, K., Zelikowsky, M., Mystikowski, J., Chowdhury, N., & Baker, A.
(2008). Optimizing inhibitory learning during exposure therapy. *Behaviour Research and
Therapy, 46*, 5-27.

Craske, M. G., & Mystkowski, J. (2006). Exposure therapy and extinction: Clinical studies. In
M. G. Craaske, D. Hermans, & D. Vansteenwegen (Eds.), *Fear and learning: From basic
processes to clinical implications* (pp. 217-233). Washington, DC: American
Psychological Association.

Ellwart, T., Rinck, M., & Baker, E. S. (2006). From fear to love: Individual differences in
implicit spider associations. *Emotion, 6*, 18-27.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as
an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social
psychology* (Vol. 23, pp. 75-109). New York: Academic Press.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social
Cognition, 25*, 603-637.

First, M. B., Spitzer, R. L., Gibbon, M., Williams, J. B. W. (1995). *Structured Clinical Interview
for DSM-IV Axis I Disorders—Patient Edition*. New York: Biometrics Research
Department, New York Psychiatric Institute.

Gamer, J., Schmukle, S. C., Luka-Krausgrill, U., & Egloff, B. (2008). Examining the dynamics
of the implicit and the explicit self-concept in social anxiety: changes in the implicit
association test–anxiety and the social phobia anxiety inventory following treatment.
Journal of Personality Assessment, 90, 476-480.

Han, H. A., Czellar, S., Olson, M. A., & Fazio, R. H. (2010). Malleability of attitudes or
malleability of the IAT? *Journal of Experimental Social Psychology, 46*, 286-298.

Han, H. A., Olson, M. A., & Fazio, R. H. (2006). The influence of experimentally-created extrapersonal associations on the Implicit Association Test. *Journal of Experimental Social Psychology, 42*, 259-272.

Hermans, D., Vansteenwegen, D., Crombez, G., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Affective priming as an indirect and unobtrusive measure of conditioned valence. *Behaviour Research and Therapy, 40*, 217-234.

Hofmann, S. G. (2007). Cognitive factors that maintain social anxiety disorder: A comprehensive model and its treatment implications. *Cognitive Behaviour Therapy, 36*, 193-209.

Huijding, J., & de Jong, P. J. (2007). Beyond fear and disgust: The role of (automatic) contamination-related associations in spider phobia. *Journal of Behavior Therapy and Experimental Psychiatry, 38,* 200-211.

Huijding, J., & de Jong, P. J. (2009). Implicit and explicit attitudes toward spiders: Sensitivity to treatment and predictive value for generalization of treatment effects. *Cognitive Therapy and Research, 33,* 211-220.

Jones, C. R., Fazio, R. H., Vasey, M. W. (in press). Attentional control buffers the effect of public-speaking anxiety on performance. *Social Psychological and Personality Science*.

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test (SC-IAT) as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*, 16- 32.

Leary, M. R. (1983a). A brief version of the Fear of Negative Evaluation Scale. *Personality and Social Psychology Bulletin, 9*, 371–375.

Mansell, W., & Clark, D. M. (1999). How do I appear to others? Social anxiety and processing of the observable self. *Behaviour Research and Therapy, 37*, 419-434.

McNally, R. J. (2007). Mechanisms of exposure therapy: How neuroscience can improve psychological treatments for anxiety disorders. *Clinical Psychology Review, 27*, 750-759.

Moscovitch, D. A., Antony, M. M., & Swinson, R. P. (2009). Exposure-based treatments for anxiety disorders: Theory and process. In M. M. Antony & M. B. Stein (Eds.), *Oxford handbook of anxiety and related disorders* (pp. 461-475). New York: Oxford University Press.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*, 653-667.

Olson, M. A., & Fazio, R. H. (2009). Implicit and explicit measures of attitudes: The perspective of the MODE model. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 19-63). New York, NY: Psychology Press.

Ouimet, A. J., Gawronski, B., & Dozois, D. J. A. (2009). Cognitive vulnerability to anxiety: A review and an integrative model. *Clinical Psychology Review, 29,* 459-470.

Paul. G. L. (1966). *Insight versus desensitization in psychotherapy.* Stanford, CA: Stanford University Press.

Philips, G. C., Jones, G. E., Rieger, E. J., & Snell, J. B. (1997). Normative data for the Personal Report of Confidence as a Speaker. *Journal of Anxiety Disorders, 11*, 215-220.

Rachman, S. J. (1989). The return of fear: Review and prospect. *Clinical Psychology Review, 9*, 147-168.

Roefs, A., Huijding, J., Smulders, F. T. Y., MacLeod, C. M., de Jong, P., Wiers, R. W., & Jansen, A. T. M. (2011). Implicit measures of association in psychopathology research. *Psychological Bulletin, 137*, 149-193.

Spielberger, C. D. (1989). *State-Trait Anxiety Inventory: Bibliography* (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto: Consulting Psychologists Press.

Teachman, B. A., & Woody, S. R. (2003). Automatic processing in spider phobia: Implicit fear associations over the course of treatment. *Journal of Abnormal Psychology, 112*, 100 – 109.

Tsao, J. C. I., & Craske, M. G. (2000). Timing of treatment and return of fear: Effects of massed, uniform, and expanding-spaced exposure schedules. *Behavior Therapy, 31*, 479-497.

Wells, A. & Clark, D. M. (1997). Social phobia: A cognitive approach. In G. C. L. Davey (Ed.), *Phobias: A handbook of description, treatment and theory* (pp. 3–26). Chichester, Sussex: Wiley.

Wigboldus, D.H.J., Holland, R.W., & van Knippenberg, A. (2005). *Single target implicit associations*. Unpublished manuscript, Radboud University Nijmegen.

Wild, J., Clark, D. M., Ehlers, A., & McManus, F. (2008). Perception of arousal in social anxiety: Effects of false feedback during social interaction. *Journal of Behavior Therapy and Experimental Psychiatry, 39*, 102-116.

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

Table 1. Means and (SDs) of outcome measures at pre-treatment, post-treatment, and follow-up.

Measure	Pre-Treatment		Post-Treatment			Follow-up	
	Mean	(SD)	Mean	(SD)		Mean	(SD)
Personalized IAT	46.02_a	(73.13)	- 2.59_b	(58.02)		0.20_b	(52.22)
PRCS	23.27_a	(5.78)	19.62_b	(7.65)		20.42_b	(7.64)
BFNE	41.72_a	(11.57)	40.10_a	(11.32)		37.82_b	(12.22)
					Speech to Camera		Speech to Audience
					Mean	(SD)	Mean	(SD)
Pre-BAT STAI-State Anxiety	53.05_a	(12.44)	44.32_b	(9.56)	45.05_b	(13.35)	47.80_c	(14.29)
SUDs - Anticipatory	50.22_a	(21.63)	40.21_b	(17.76)	39.68_b	(23.16)	$43.18_b{}^{1}$	(22.25)
SUDs - Maximum	62.05_a	(23.66)	41.20_b	(19.04)	43.18_b	(24.16)	45.92_c	(21.61)
BCL Total	83.30_a	(26.78)	58.64_b	(23.38)	$63.65_b{}^{2}$	(27.71)	61.25_b	(27.04)
Mean Heart Rate During BAT	$91.05_a{}^{3}$	(13.56)	$87.79_b{}^{3}$	(11.60)	$92.53_a{}^{4}$	(14.88)	$91.60_a{}^{4}$	(15.31)
Observer Rating of Speech Quality	5.24_a	(1.61)	5.88_b	(1.27)	5.84_b	(1.38)	6.05_b	(1.28)

Note: Means with different subscripts differ at $p < .05$. However, the two speeches at follow-up were not compared to one another. $N = 40$ except where noted. [1] Difference from post-treatment: $p = .062$; [2] Difference from post-treatment: $p = .073$; [3] $N = 38$; [4] $N = 36$

Table 2. Summary of regression analyses for PIAT scores predicting change from post-treatment to 1-month follow-up.

Dependent Variable (Follow-up minus Post)	R^2	Intercept B (SE)	DV at Pre-Treatment B (SE)	sr	DV at Post-Treatment. B (SE)	sr	PRCS at Post-Treatment B (SE)	sr	BFNE at Post-Treatment B (SE)	sr	PIAT at Post-Treatment B (SE)	sr
PRCS Change	.15	.81† (.45)	1.29† (.74)	.27†	-1.71* (.77)	-.34*	--	--	-.22 (.53)	-.06	.60 (.52)	.18
BFNE Change	.27*	-2.28* (.89)	5.31* (2.09)	.37*	-5.98* (.2.09)	-.42*	-.62 (1.04)	-.08	--	--	1.76† (1.05)	.24†
Mean Anticipatory Anxiety Change	.43*	.14† (.08)	.30* (.12)	.32*	-.16 (.13)	-.16	.06 (.11)	.06	.06 (.09)	.09	.20* (.09)	.28*
Mean SUDs Maximum Change	.32*	3.35* (1.56)	4.37† (2.29)	.27†	-4.01† (2.37)	-.24†	.04 (2.16)	.00	3.12 (1.89)	.23	2.53 (1.89)	.20
BCL Change	.44*	3.62† (2.04)	11.06* (3.7)	.39*	-14.01* (3.3)	-.55*	2.48 (3.01)	.11	.21 (2.47)	.01	5.32* (2.37)	.29*
Mean Heart Rate Change	.26†	4.23* (1.60)	4.36 (3.67)	.19	-6.73† (3.73)	-.28†	2.08 (1.82)	.18	-3.75† (.1.91)	-.31†	4.81* (2.18)	.35*
Observer Rating of Speech Quality Change	.34*	.07 (.08)	.49* (.21)	.32*	-.63* (.21)	-.42*	-.24* (.10)	-.33*	.00 (.10)	.00	.06 (.10)	.08

Note: All predictors z-transformed; B = unstandardized regression coefficient; *sr* = semi-partial correlation coefficient. Change scores reflect follow-up minus post-treatment (i.e., positive change scores indicate an increase from post-treatment to follow-up [i.e., ROF]). The Anticipatory Anxiety Composite score was the average of the standardized anticipatory SUDS and State Anxiety scores for each follow-up BAT, each standardized based on the post-treatment mean and *SD*. Follow-up scores were standardized based on the post-treatment mean and *SD* so that a positive change score indicates an increase in *SD* units from post-treatment to follow-up).
* $p < .05$; † $p < .10$