Running head: EMOTIONAL REACTIVITY IN PSYCHOPATHOLOGY

Defining reactivity: How several methodological decisions can affect conclusions about emotional reactivity in
psychopathology

Brady D. Nelson

University of Illinois - Chicago

Stewart A. Shankman

University of Illinois - Chicago

Thomas M. Olin

Western Psychiatric Institute

Daniel N. Klein

Stony Brook University

Abstract

There are many important methodological decisions that need to be made when examining emotional reactivity in psychopathology. In the present study, we examined the effects of two such decisions in an investigation of emotional reactivity in depression – 1) which (if any) comparison condition to employ and 2) how to define change. Depressed ($N=69$) and control ($N=37$) participants viewed emotion inducing film clips while subjective and facial responses were measured. Emotional reactivity was defined using no comparison condition (i.e., raw scores), baseline comparison condition (i.e., no stimulus presented), and neutral comparison condition (i.e., neutral stimulus presented). Change in emotional reactivity was assessed using four analytic approaches: difference scores, percentage change, residualized change, and ANCOVA. Results differed among the three comparison conditions and among several of the analytic approaches. Overall, our investigation suggests that choosing a comparison condition and the definition of change can significantly influence the presence of group differences in emotional reactivity. Recommendations for studies of emotional reactivity in psychopathology are discussed.

Acknowledgement

This research was supported by the American Psychological Foundation and Council of Graduate Departments of Psychology Clarence J. Rosecrans Scholarship as well as National Institute of Mental Health Grant F31 MH67309, both awarded to Stewart A. Shankman. Additionally, we would like to thank Jonathan Rottenberg for his assistance and for allowing us to use his self-report emotions questionnaire and standardized movie clips.

Introduction

Aberrations in emotional reactivity have been identified in several psychopathologies, such as depression (Clark, Watson, & Mineka, 1994; Davidson, Pizzagalli, Nitschke, & Putnam, 2002; Rottenberg, Gross, & Gotlib, 2005), anxiety (Goldin, Manber, Hakimi, Canli, & Gross, 2009; McNeil, Vrana, Melamed, Cuthbert, & Lang, 1993) and schizophrenia (Herbener, Song, Khine, & Sweeney, 2008; Kring & Moran, 2008). However, for many of these conditions research on emotional reactivity has yielded inconsistent results. For example, research on negative emotional reactivity in depression has led to two contradictory hypotheses, predicting both facilitated (Beck, 1976; Beck, Rush, Shaw, Emery, 1979; Scher, Ingram, & Segal, 2005) and attenuated negative reactivity (Rottenberg, Kasch, Gross, & Gotlib, 2002). One explanation for these conflicting findings is important methodological differences between studies. There are many critical questions that need to be considered when designing an experiment to test individuals' reactions to an induction, treatment, or experimental manipulation (Kazdin, 2002; Keppel & Wickens, 2004). In the present study we will examine the implications of making two such decisions: choosing the comparison condition and analytic method for measuring change.

What Type of Comparison Condition to Use (If Any)?

The term *emotional reactivity* suggests that the acute emotional output being measured is a change from some preceding state. Indeed, many basic theories of emotion argue that emotional reactions do not occur in isolation, but rather are superimposed on prior affective states (Rosenberg, 1998). It therefore stands to reason that the term emotional reactivity implies not only a response to a stimulus, but also that a stimulus-induced emotional state changed (or varied) from the pre-stimulus state.

On the other hand, it may be appropriate to operationalize emotional reactivity as an individual's response to an emotional stimulus or event without referencing a pre-stimulus state or any other comparison condition. For example, experience sampling method is an approach that assesses in-the-moment emotional reactions to various daily events often without referencing a preceding emotional state (Csikszentmihalyi & Larson, 1987; Stone & Shiffman, 2002). In these types of studies, groups are often compared on their mean acute emotional response to daily events, rather than change in emotional state before and after the event. For

example, using this approach Peeters, Nicolson, Berkhoff, Delespaul, and deVries (2003) found that compared to healthy controls, depressed individuals reported reduced positive and negative emotional reactivity to daily negative events. Of course, not using a control condition raises the possibility that the emotional response may be due to the natural drift of the emotion (i.e., the emotion would have occurred regardless of the stimulus or situation) or the presentation of any stimulus rather than the specific event.

Baseline comparison condition. If a comparison condition is desired, a second issue is deciding what constitutes an appropriate condition. One option is to measure the baseline emotional level prior to an emotion induction and compare it to the level following the induction (e.g., Gotlib, Joormann, Minor, & Hallmayer, 2008). This method has the advantage of controlling for the natural drift in emotions prior to the induction. However, it is important to consider that groups may systematically differ in their emotional state at baseline. In their classic description of experimental design, Campbell and Stanley (1963) refer to this type of design as a nonequivalent control group design. Interpreting results from this type of design can be problematic if effects are misattributed to the induction, rather than the group differences present at baseline. Nonetheless, the presence of group differences at baseline should not necessarily be considered ‘noise’ that needs to be controlled, but may rather reflect an important substantive phenomenon.

Another important issue with baseline emotion measures is that they may actually reflect longer lasting mood states and be qualitatively different from more transient emotional responses elicited by an emotion induction (Rosenberg, 1998). Moods have been defined as diffuse, slow moving feeling states that are weakly tied to specific objects or situations (Rottenberg, 2005). For example, a person’s mood may be a down in the dumps feeling that somewhat waxes and wanes, but is pervasive across time. In contrast, emotions are quick-moving reactions that occur when encountering meaningful stimuli that call for adaptive responses (Rottenberg, 2005). An important distinction between mood and emotions is their temporal length. While moods can last for hours or days, emotions are typically much briefer, lasting only seconds or minutes. Therefore, baseline measures may better reflect current mood state as they are often assessed prior to encountering meaningful stimuli.

Neutral comparison condition. Another potential solution is to use neutral valenced stimuli that are

matched in modality, or some other feature, with the emotion inducing stimulus or event. Examples can include benign tasks/situations, such as reading neutral words (e.g., Mogg, Bradley, Williams, & Mathews, 1993), or viewing neutral valenced stimuli, such as pictures (Lang, Bradley, & Cuthbert, 2008), film clips (Gross & Levenson, 1995), or fixation crosses. Neutral conditions provide the benefit of controlling for stimulus presentation. They may also be advantageous because, compared to group differences present at baseline, which are uncontrolled, neutral conditions contain some situational demands and thus should attenuate baseline group differences.

However, similar to baseline measurements, groups may still differ during a neutral condition, as they may elicit some (albeit small) positive or negative emotion. For example, Rottenberg et al. (2002) presented depressed and control participants with a neutrally valenced film clip (coastal landscape scenery), yet still found group differences in subjective sadness and amusement during the film clip.

Neutral stimuli may actually be viewed as ambiguous, because the situational demands may be somewhat unclear. In studies of emotional reactivity, participants are often instructed that they will be presented with emotion inducing situations. Thus, when asked to rate their level of happiness during a situation intended to elicit minimal emotional reactions, participants may instead rely on other sources of information (i.e., episodic memory). This is particularly apparent in individuals with psychopathology. For example, individuals with and at risk for depression have been shown to have a negative bias in how they interpret ambiguous situations (Dearing & Gotlib, 2009; Lawson & MacLeod, 1999; Mogg, Bradbury, Bradley, 2006).

In other words, neutral stimuli may be better understood as ‘weak situations.’ Situational strength has long been identified as an important factor that moderates the relationship between individual or group differences and behavior (Caspi & Moffitt, 1993; Cooper & Withey, 2009; Mischel, 1977; Snyder & Ickes, 1985). Strong situations provide unambiguous stimuli that generally yield uniform reactions across individuals. In contrast, weak situations are more ambiguous events that attenuate the influence of the situation and increase the contribution of individual differences on subsequent responses. When comparing emotional reactivity between psychopathological and control populations, group differences will likely be more amplified during weak relative to strong situations (Lissek, Pine, & Grillon, 2006). Therefore, if neutral

comparison conditions are perceived as ambiguous, they may actually produce the most prominent group differences in emotional responding.

Another issue with neutral comparison conditions is that they are likely to produce poor response integration across different systems of emotion (i.e., subjective, behavioral, and physiological). Situations that elicit strong, intense emotional experiences produce greater coherence across emotion systems (Lang, Levin, Miller, & Kozak, 1983; Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005). Since neutral conditions are intended to elicit little to no emotional reactions, it may be problematic to compare one condition where coherence is low (i.e., neutral condition) to another where coherence is high (i.e., emotional condition).

Analytic Methods for Measuring Change

Once a comparison condition has been chosen, another decision is determining how to measure the change in emotional reactivity. Measuring change has been a widely debated topic for decades (Bereiter, 1963; Blomqvist, 1977; Cronbach & Furby, 1970; Gottman, 1995; Rogosa, 1995; Twisk & Proper, 2004), and while this may appear to be a rather straightforward decision, complications can arise as different methods can present statistical and interpretative hurdles. Studies examining emotional reactivity have used several different conventions for measuring change. Often the most basic approach is to calculate difference scores (i.e., absolute change scores), which entail subtracting the response during the comparison condition from the response during the emotion induction condition¹. A non-zero difference score indicates that there has been a change (either increase or decrease) in emotional reactivity.

However, the interpretation of difference scores can be problematic (Rogosa, Brandt, & Zimowski, 1982). To illustrate this point consider the following hypothetical situation (see Table 1) where two Groups (A and B) undergo an emotion induction procedure, and their happiness is assessed before and after the induction on a scale ranging from 0 (None) to 8 (Maximum). The difference scores indicate that Group B had greater happiness reactivity following the emotion induction compared to Group A. Yet, this is completely because of the differences in pre-induction happiness and not the responses to the emotion induction. In addition, this is not the product of a ceiling effect, as both groups scored a 6 out of a possible 8 following the induction. While difference scores can be informative when groups are equivalent during the comparison condition, they can be

problematic when group differences are present during the comparison condition. Difference scores are also susceptible to statistical biases, such as regression towards the mean, especially if differences are present during the comparison condition (Blomqvist, 1986; Hayes, 1988).

Another approach is to calculate a percentage change score (PCS; e.g., Blumenthal, Elden, & Flaten, 2004; Greenstein & Kassel, 2010). This is typically calculated by subtracting the response during the comparison condition from the response during the emotion induction condition, dividing this figure by the response during the comparison condition, and multiplying this value by 100 (i.e., $[\text{Induction} - \text{Comparison}] / \text{Comparison} \times 100$). Larger scores represent a greater percentage change in emotional reactivity from the comparison condition. Unlike difference scores, PCSs factor in differences during the comparison condition. However, PCSs can be problematic as the ceiling and floor of the scale can influence the outcome. Consider the previously mentioned example with a different set of results (see Table 2). In this case, both Groups A and B increase by 2 points on the scale, but Group A has a larger PCS. These group differences may be because Group B has reached the ceiling of the scale (8) during the post-induction measurement. Another possible problem with PCS is that it requires that the measure is on a ratio scale (Stevens, 1946). Therefore, while a PCS does factor in differences during the comparison condition, it too presents drawbacks.

A third approach is to statistically control for group differences during the comparison condition using residualized change scores (e.g., McFarland & Klein, 2009). Calculating residualized change requires two steps. First, a linear regression of the outcome measurement on the comparison measurement is performed for both groups pooled together, and the residuals from this regression are computed. Second, a between-subjects comparison is performed on the residuals to compare the mean value between groups. Residualized change is often preferred over difference scores because it more accurately removes the influence of baseline ratings (Cohen, Cohen, West, & Aiken, 2003).

Analysis of covariance (ANCOVA) is another method used to control for group differences during the comparison condition. In ANCOVA, variables that are correlated with the dependent measure, but not the independent variable, are statistically controlled for by being included as covariates². The primary purpose of ANCOVA is to reduce noise caused by covariates and thus increase power of the *F* test (Tabachnik & Fidell,

2007). In a simulation study, Forbes and Carlin (2005) showed that ANCOVA is generally preferred over residualized change, as the latter is biased by smaller sample sizes and larger group differences during the comparison condition. However, there are important limitations to ANCOVA. For example, ANCOVA should be avoided when random assignment is not possible and groups differ on the covariate (i.e., comparison condition). One of the assumptions of ANCOVA is that the covariates are unrelated to the independent variable. In true experiments, this issue can be addressed with random assignment to the independent variable; however, this is not an option in quasi-experimental designs. As discussed in the eloquent discourse by Miller and Chapman (2001), group differences in the covariate may actually reflect meaningful substantive differences related to group membership; therefore, including them as covariates can alter the meaning of group membership.

Study Aims

The purpose of this study is to highlight the importance of choosing a comparison condition and measuring change when examining emotional reactivity in psychopathology. Specifically, we intend to demonstrate how these decisions can potentially alter the pattern of results, and thus lead to different (and even opposing) conclusions. To illustrate these points we used data from an investigation of emotional reactivity in depression.

Depression is an interesting test case for these aims as emotional dysfunction is often considered the core deficit in the disorder. Indeed, there has been a great deal of interest in characterizing the pattern of emotional reactivity in depression (Clark et al., 1994; Davidson et al., 2002; Rottenberg et al., 2005). For example, Clark and Watson proposed the tripartite model, which hypothesizes that depression is characterized by high Negative Emotionality (i.e., a tendency to react to negative stimuli with fear, sadness, and/or guilt) and low Positive Emotionality (i.e., a tendency to react to positive stimuli with low joy, interest, and excitement; Clark & Watson, 1991; Clark et al., 1994; Watson & Tellegen, 1985). A second theoretical model is the Emotion Context Insensitivity (ECI) hypothesis, which states that depression has an inhibitory effect on reactivity to stimuli that ordinarily elicit emotional reactions (Rottenberg et al., 2005). Thus, the ECI hypothesis predicts that depression is characterized by diminished positive *and* negative emotional reactivity

(Bylsma, Morris, & Rottenberg, 2008).

While there has been a great deal of debate on these and other models (Davidson, 1998; Depue & Iacono, 1989; Heller & Nitschke, 1998; Shankman & Klein, 2003), the purpose of this paper is not to compare the different theoretical models, but rather to demonstrate how important methodological decisions in designing studies of emotional reactivity may affect whether the results support or refute different models. We chose to use an investigation of emotional reactivity in depression to exemplify these points because (1) emotional dysfunction is often considered a core deficit and (2) there are deficits in both positive and negative emotional reactivity.

In the present study we had depressed and healthy control participants view emotion inducing film clips and measured their subjective and facial responses to the clips. In order to highlight the importance of methodological decisions, the present study replicated the design and methods of Rottenberg et al. (2002), with slight modifications (i.e., included an additional positive film clip). Emotional reactivity was defined using three different comparison conditions: (1) No comparison condition, (2) Baseline comparison condition, and (3) Neutral comparison condition. In addition, we also examined four different analytic methods for measuring change: (1) Difference scores, (2) Percentage change scores (PCS), (3) Residualized change scores, and (4) ANCOVA.

Methods

Participants

The sample consisted of 69 individuals with current major depressive disorder (MDD), as defined by the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; American Psychiatric Association, 1994), and 37 control participants. The control group was required to have no lifetime diagnoses of mood or anxiety disorder, hard drug or alcohol dependence, or eating disorder. The control group was also required to have a 24-item Hamilton Rating Scale of Depression (HRSD; Hamilton, 1960) score of less than 8. All participants were recruited through advertising in the community and psychiatric and psychological clinics (see Shankman, Klein, Tenke, and Bruder [2007] for a more detailed description of the sample and recruitment). Participants were excluded from the study if they had a lifetime diagnosis of schizophrenia or other psychotic disorder,

bipolar disorder, or dementia; were unable to read and write English; or had a history of head trauma in which they lost consciousness (since participants also completed an EEG based reward processing task – see Shankman et al. [2007]). All participants gave informed consent and were paid for their participation.

Interview Measures

Diagnoses and clinical characteristics were made via face-to-face interviews using the Structured Clinical Interview for DSM–IV (SCID; First, Spitzer, Gibbon, & Williams, 1996). Severity of depressive symptomatology was assessed using the HRSD. S.A.S. and a master’s level diagnostician (Suzanne Rose) conducted the assessments. Ms. Rose has demonstrated high levels of inter-rater reliability in the past and has trained numerous diagnosticians on the SCID and HRSD for 15 years (Keller et al., 1995; Klein, Schwartz, Rose, & Leader, 2000). She and D.N.K. trained S.A.S. to criterion, and diagnoses were regularly discussed in best estimate meetings (Klein, Ouimette, Kelly, Ferro, & Riso, 1994).

Stimuli

Films were selected from the film bank of Gross and Levenson (1995) and used in previous studies on emotional reactivity in depression (e.g., Rottenberg et al., 2002). Films were high in intensity, complexity, and attentional capture while having good standardization. Five different emotion eliciting film clips were used: neutral, sad, fear, and two amusing. The neutral film (180s) depicted costal landscape scenery. The fear film (140s) was from the movie *Fearless* and depicted heavy turbulence in the cabin of a commercial airline as it was about to crash. The sad film (171s) was from the movie *The Champ* and depicted a boy who was distraught over the death of his father. The first amusing film (120s) depicted a British comedian, Mr. Bean, engaging in antic, slapstick-type comedy. The second amusing film (205s) was taken from the stand-up comedy movie *Robin Williams- Live at the Met*. We used two amusing films as there are large individual differences in what people find humorous (Wyer & Collins, 1992). Specifically, the Robin Williams clip depicted darker, more cynical humor than the Mr. Bean clip. The neutral film clip was always presented first to get the most pure neutral state, and to reduce the possibility of the emotion eliciting film clips contaminating responses to the neutral film clip. Fear and sad film clips were always presented second and third, counterbalanced across all subjects. Finally, the amusing film clips were always presented last so that

participants would leave the experiment in the best mood possible. Film stimuli were presented on a 15-inch television monitor at a viewing distance of approximately two meters. Participants were videotaped from the neck up and recording took place in low ambient light.

Measure of Subjective Emotion

Participants were given subjective emotions questionnaires immediately after each film asking them to rate the greatest amount of happiness (Subjective-Happiness), fear (Subjective-Fear), and sadness (Subjective-Sadness) they were feeling on a scale ranging from 0 (none) to 8 (an extreme amount). Additional emotions words were included to reduce demand characteristics. Participants also rated the overall intensity of their feelings during each film clip on a scale ranging from -4 (extremely mild) to 4 (extremely intense). These were the same questionnaires used by Rottenberg et al. (2002).

Measure of Facial Response

The Emotional Expressive Behavior (EEB) coding system is a modified version of the coding system developed by Gross and Levenson (1993) and contains 18 codes including both descriptions and judgments about behavior. It is a sensitive and specific measure of a broad range of emotional behaviors, with particular attention to emotional expressive behaviors (Gross, 1999). Trained coders made emotions ratings on a 7-point scale (ranging from 0-6) with values representing an aggregate of intensity (slight, moderate, strong) and duration (short vs. long) of response. Videotapes of participant's facial responses during the emotions films were divided into 30s epochs and an emotion rating for each facial code was made during each epoch. That is, for each epoch the following values were possible: no response (0), slight and short (1), slight and long (2), moderate and short (3), moderate and long (4), strong and short (5), strong and long (6). Each code includes a detailed description of the specific facial movements associated with an emotion, as well as guidelines for determining the strength of the response.

We collapsed across the duration codes and used the peak response (i.e., highest score during any epoch), rather than averaging epochs. This allowed us to look at each participant's *greatest* facial response, rather than the average response, which may have been diminished by including epochs with no response. This also corresponded with the instructions for the self-report measure which asked about the *greatest* emotion felt

during the clip. Thus, facial response variables were recoded to a 4-point scale with the following values: no response (0), slight response (1), moderate response (2), and strong response (3). We report the results for three of these facial codes - happiness (Facial-Happiness), sadness (Facial-Sadness), and fear (Facial-Fear). Total number of smiles (Facial-Smiles) was also coded during the amusing clips as a second positive affect code. As would be expected from a neutral stimulus, the neutral film clip elicited very little facial response (see Table 4).

Inter-rater reliability for the codes was assessed for a subset ($N = 15$) of participants yielding the following intra-class correlations (Shrout & Fleiss [1979]; Case 3) - .88 for Facial-Sadness during the sad film; .10 for Facial-Fear during the fear film, .89 for Facial-Happiness and .89 for Facial-Smiles during the Mr. Bean film, and .86 for Facial-Happiness and .95 for Facial-Smiles during the Robin Williams film. Our measure of Facial-Fear had poor reliability between coders. Despite the poor reliability, the fear film clip still elicited its intended response as the Facial-Fear scores were higher for the fear film than the other films. Therefore, we still included Facial-Fear in the analyses with the qualification that the measurement may not be adequately reliable.

Data Analyses

Analyses were conducted separately for subjective and facial responses. For each emotions film we only examined the dependent measure of the intended emotion (e.g. Subjective-Sadness and Facial-Sadness during the sad film), as films elicited little to none of the other measured emotions. For the amusing films we averaged participant's responses to the Mr. Bean and Robin Williams film for three reasons. First, Subjective-Happiness to the Mr. Bean and Robin Williams films were correlated, $r(106) = .43, p < .001$. Second, averaging responses across the two amusing films reduced the noise created by individual differences in interpreting humorous stimuli (Wyer & Collins, 1992), and gave us a better estimate of participants' positive emotional responses. Third, we found the same pattern of results when the Mr. Bean and Robin Williams films were looked at separately as when they were averaged together.

For subjective responses, we examined emotional reactivity using three different comparison conditions: (1) No comparison condition, (2) Baseline comparison condition, and (3) Neutral comparison

condition. In the no comparison condition we examined the raw scores to each emotion induction using a one-way analysis of variance (ANOVA) with depression status (depressed vs. control) entered as the between-subjects factor. For the analyses that involved a comparison condition (i.e., baseline and neutral), we first calculated emotional reactivity using four different analytic strategies: (1) Difference scores (i.e., Induction – Comparison), (2) Percentage change scores (PCS: $100 \times [\text{Induction} - \text{Comparison}] / \text{Comparison}$), (3) Residualized change scores (i.e., linear regression of the induction variable on the comparison variable [baseline or neutral] was performed for both groups pooled together, yielding residualized change scores), and (4) ANCOVA (i.e., comparison condition [centered] entered as the covariate). Second, each measurement of emotional reactivity was then entered as the dependent variable in a one-way ANOVA with depression status as the between-subjects factor. The N 's for Subjective-Sadness slightly differed from the other subjective measurements as two participants erroneously skipped the item.

For the facial response, we examined emotional reactivity using two different comparison conditions: (1) No comparison condition and (2) Neutral comparison condition. Baseline facial activity was not measured as we anticipated that there would likely be few codeable behaviors during a period when there was no film presented. Similar to the subjective measures, we also calculated emotional reactivity from neutral (which again, had very little facial responses) using four different analytic strategies (see above). Each measurement of emotional reactivity was then entered as the dependent variable in a one-way ANOVA with depression status as the between-subjects factor. The N 's for facial response also slightly varied because of missing data due to equipment failure and obstruction of participant's faces during recording.

Results for both subjective and facial response were nearly identical when gender³ was added as a covariate; therefore, we collapsed across gender for all analyses.

Results

Demographic and Clinical Characteristics

As shown in Table 3, depressed and control participants did not differ on sex, age, ethnicity, or employment status. As expected, depressed participants had lower GAF, $F(1, 104) = 504.29$, $MSE = 47.01$, $p < .001$, and higher HRSD scores compared to controls, $F(1, 104) = 342.00$, $MSE = 41.53$, $p < .001$. Within the

depressed participants 50.7% were currently taking medication, 29.0% had a current comorbid anxiety disorder and 42.0% had a lifetime comorbid anxiety disorder⁴.

Coherence between Subjective and Facial Response

Subjective-Happiness, Sadness, and Fear, and the associated facial responses were highest during the respective film clips compared to the other film clips (all p 's < .05). Self-reported emotional intensity was significantly greater during the amusing, sad, and fear films compared to the neutral film clip (p 's < .001). Among the non-neutral films, emotional intensity was greater during the sad and fear film, compared to the amusing films (p 's < .001). Emotional intensity did not differ between the sad and fear clips.

Consistent with previous reports (Lang et al., 1983; Mauss et al., 2005), there was greater coherence between subjective and facial responses during the emotion inducing films compared to the neutral film. Specifically, there were significant correlations between Subjective-Happiness and Facial-Happiness, $r(105) = .57, p < .001$, and Subjective-Happiness and Facial-Smiles during the amusing films, $r(105) = .50, p < .001$; Subjective-Sadness and Facial-Sadness during the sad film, $r(104) = .24, p < .05$; and Subjective-Fear and Facial-Fear during the fear film, $r(104) = .35, p < .001$. In contrast, subjective and facial responses were not correlated during the neutral film (all p 's > .43). Poor response coherence during the neutral film was largely the result of restricted variance in facial responses, as less than 5% of participants exhibited any Facial-Happiness or Facial-Smiles and no participant exhibited Facial-Sadness or Facial-Fear.

Emotional Reactions during the Comparison Conditions

Table 4 shows the means and standard deviations of subjective responses during the baseline and neutral comparison conditions for depressed and control participants. Before examining emotional reactivity, we first looked at whether there were group differences during the comparison conditions. At baseline, depressed and control participants differed in their subjective emotions, with depressed participants reporting less Subjective-Happiness, $F(1, 104) = 35.00, MSE = 2.06, p < .001$, and more Subjective-Sadness, $F(1, 104) = 39.38, MSE = 2.81, p < .001$, and Subjective-Fear, $F(1, 104) = 8.64, MSE = 1.40, p < .01$, compared to controls.

Depressed and control participants also differed in subjective responses during the neutral film clip, with depressed participants reporting less Subjective-Happiness, $F(1, 104) = 4.34$, $MSE = 4.25$, $p < .05$, more Subjective-Sadness, $F(1, 104) = 16.65$, $MSE = 2.34$, $p < .001$, and a trend for more Subjective-Fear, $F(1, 104) = 3.87$, $MSE = 0.64$, $p < .06$, compared to controls. The neutral film clip elicited minimal facial responses, and depressed and control participants did not differ on any of these variables (p 's $> .43$).

Emotional Reactivity

Emotional reactivity was assessed using three different comparison conditions: (1) No comparison condition, (2) Baseline comparison condition, and (3) Neutral comparison condition. In the analyses containing no comparison condition, we simply compared depressed and control participants' raw emotional response to each film. For analyses utilizing a comparison condition (i.e., baseline and neutral) we used four different analytic approaches for measuring change: (1) Difference scores, (2) PCS, (3) Residualized change, and (4) ANCOVA.

No comparison condition. Table 4 shows the means and standard deviations of the raw subjective and facial responses to the emotion inducing film clips for depressed and control participants. Results indicated that depressed participants reported less Subjective-Happiness during the amusing films relative to controls, $F(1, 104) = 4.09$, $MSE = 3.49$, $p < .05$, but the groups did not differ in Subjective-Sadness during the sad film or Subjective-Fear during the fear film (p 's $> .27$). Depressed and control participants also did not differ in their facial response during any of the emotion inducing film clips (p 's $> .28$).

Baseline comparison condition. When using a baseline comparison condition, depressed and control participants differed in both their positive and negative subjective emotional reactivity (see Table 5). Interestingly, this differed among the various analytic approaches for measuring change. Specifically, when using difference scores or PCSs, depressed participants had *decreased* Subjective-Sadness and Fear reactivity, and *increased* Subjective-Happiness reactivity relative to controls in the sad, fear, and amusing film clips, respectively. However, there were no group differences in emotional reactivity when using residualized

change or ANCOVA. As mentioned above, there was no baseline measure of facial response, which would serve as the baseline comparison condition for this measure.

Neutral comparison condition. Compared to baseline comparison results, there were fewer group differences in emotional reactivity when using the neutral comparison condition (see Table 4). Similar to the baseline comparison condition, depressed participants reported *less* Subjective-Sadness reactivity relative to controls. However, group differences in Subjective-Fear reactivity were reduced to a trend level, and were eliminated for Subjective-Happiness. In addition, group differences in Subjective-Sadness and Subjective-Fear reactivity were only found when change was defined using PCSs, and not when using difference scores, residualized change or ANCOVA. As previously mentioned, the neutral film clip elicited minimal facial response, and there were no differences in facial response for the no comparison condition analyses (i.e., raw scores). Therefore, as expected depressed and control participants did not differ in their facial reactivity when using a neutral comparison condition (p 's > .32).

Discussion

The present study aimed to highlight the importance of choosing the comparison condition and analytic method for assessing change when investigating emotional reactivity in psychopathology. We intended to show how these decisions can potentially alter the pattern of results, and thus differentiate support for the various theories of emotional reactivity in conditions such as depression, anxiety and schizophrenia. In the present study we compared the pattern of subjective and facial emotional reactivity between depressed and control participants while they viewed emotion inducing film clips. Emotional reactivity was analyzed using three different comparison conditions: (1) No comparison condition, (2) Baseline comparison condition, and (3) Neutral comparison condition. In addition, we also examined four different analytic methods for assessing change: (1) Difference scores, (2) Percentage change scores (PCS), (3) Residualized change, and (4) ANCOVA.

Effects of Choosing the Comparison Condition on Emotional Reactivity

Subjective reactivity. Overall, our results demonstrate that choosing the comparison condition can significantly influence whether (and how) depressed and control participants differ in subjective emotional

reactivity. Specifically, for positive subjective emotional reactivity we found that all three comparison conditions produced different results. When emotional reactivity was defined using no comparison condition (i.e., raw scores), the results indicated that depressed participants reported *less* Subjective-Happiness during the amusing films relative to controls. These results support other findings of diminished positive emotional reactivity in depression (Bylsma et al., 2008) and are consistent with several extant models of the emotional dysfunction in depression (e.g., tripartite model - Clark & Watson, 1991; ECI hypothesis - Rottenberg et al., 2005). In contrast, when emotional reactivity was defined using a baseline comparison condition, the results indicated that depressed participants reported *greater* Subjective-Happiness reactivity relative to controls. However, this was predominately because of the group differences present at baseline, where depressed participants reported significantly lower Subjective-Happiness compared to controls. This finding is not surprising, as nonequivalent baseline scores between groups have long been known to cause statistical and interpretative problems (Campbell & Stanley, 1963). Finally, when emotional reactivity was defined using the neutral comparison condition (where the groups also differed), depressed and control participants did not differ in their Subjective-Happiness reactivity. In sum, all three comparison conditions produced a different set of results when comparing depressed and control participants in their pattern of positive emotional reactivity.

Similarly, for negative subjective emotional reactivity the results also differed among the three comparison conditions. In general, the pattern of negative emotional reactivity was similar for Subjective-Sadness and Subjective-Fear; a finding which supports other studies that combine the two negative affects together (e.g., Dunn, Dalgleish, Lawrence, Cusack, & Ogilvie, 2004). When negative emotional reactivity was defined using no comparison condition, depressed and control participants did not differ in Subjective-Sadness or Subjective-Fear. However, when using a baseline comparison condition the results indicated that depressed participants reported *less* Subjective-Sadness and Subjective-Fear reactivity relative to controls, a finding which would support the ECI hypothesis (Rottenberg et al., 2002; 2005). Nonetheless, similar to the results for Subjective-Happiness, this was completely because of group differences at baseline, where depressed participants reported significantly greater Subjective-Sadness and Subjective-Fear, a finding which would support the tripartite model (Clark & Watson, 1991, Clark et al., 1994). Thus, depending on the comparison

condition, the data can provide support for either of these two major theories of emotional reactivity in depression (ECI and tripartite). Finally, when emotional reactivity was defined using a neutral comparison condition, as with the baseline comparison results, depressed participants reported less Subjective-Sadness reactivity relative to controls, but the groups no longer differed in Subjective-Fear reactivity.

As illustrated in the present study, groups may differ in their emotional state during the comparison condition. In true experiments this is rarely an issue, as participants are randomly assigned to experimental groups; therefore, groups will likely not differ during the comparison condition (Campbell & Stanley, 1963). However, this is often not an option in quasi-experimental designs, such as the present study, where participants cannot be randomly assigned to have major depressive disorder. The potential problem is that, even if there are differences in emotional reactivity between groups, those differences may actually be attributable to differences present at baseline rather than to the induction.

Another issue is that the psychological variable being measured may not retain the same meaning over multiple observations (Bereiter, 1963; Cronbach & Furby, 1970; Lord, 1958). That is, the baseline (or comparison) observation may have a qualitatively different meaning than that taken during an emotion induction procedure. For example, baseline emotion ratings may actually reflect current mood state rather than the more transient emotional responses elicited by an induction (Rosenberg, 1998). Therefore, calculating a difference score between qualitatively different affective states (i.e., mood – emotion) may be difficult to interpret.

There are also issues with using a neutral comparison condition. Neutral ‘control’ conditions may be better understood as weak situations. In the present study, the neutral condition depicted coastal landscape scenery that, despite its intended valence, elicited mild amounts of both positive and negative emotions. Prior investigations have found that group differences in emotional reactivity are more amplified during weak relative to strong situations (Lissek, Pine, & Grillon, 2006). Indeed, we found that during the neutral condition depressed participants reported significantly greater Subjective-Sadness and Subjective-Fear and less Subjective-Happiness relative to controls. Moreover, these group differences were larger than that observed during the other film clips. Therefore, our neutral film clip (and neutral conditions in general) may better be

viewed as a weak situation, because during the film clip the situational demands were relatively low allowing individual difference factors, such as depression, to play a larger role.

Our results also exemplify how two commonly used comparison conditions (i.e., a pre-induction baseline and neutral stimulus condition) can produce different patterns of results. Specifically, when utilizing the baseline comparison condition, depressed participants had reduced Subjective-Sadness and Subjective-Fear reactivity and increased Subjective-Happiness reactivity relative to healthy controls. In contrast, when using the neutral comparison condition depressed participants only had reduced Subjective-Sadness reactivity relative to controls. While both the baseline and neutral conditions can be considered weak situations, the baseline condition provided even less situational demands compared to the neutral condition. This may explain why group differences were most prominent when the baseline measurement was used as the comparison condition.

Facial response. Depressed and control participants did not differ in their facial response during any of the film clips. Expectedly, the films elicited their intended facial responses (i.e., Facial-Happiness during the amusing films). However, despite differences in their subjective emotional reactivity, depressed and control participants did not differ in any of the measured facial responses. Participants also exhibited minimal facial responses during the comparison conditions. For the baseline comparison condition, we anticipated that there would likely be few codeable behaviors (as participants did not view any stimuli) and thus their facial response was not coded. The absence of codeable behaviors would have made difference scores using the baseline comparison condition nearly identical to the ‘no comparison condition’ analyses (i.e., raw scores). This point was exemplified in the analyses that used the neutral film (where participants had negligible facial responding) as the comparison condition. Facial reactivity difference scores for these analyses were nearly identical to the raw scores.

One reason depressed and control participants may not have differed in their facial response is because the coding system we chose to use (i.e., Emotional Expressive Behavior [EEB] coding system) was not sensitive enough. Prior research has shown differences in facial responding between depressed and non-depressed individuals (Gehricke & Shapiro, 2000; Renneberg, Heyn, Gebhard, & Backmann, 2005; Sloan,

Strauss, Quirk, & Stajatovic, 1997). To our knowledge, only two studies (Chentsova-Dutton, Chu, Tsai, Rottenberg, Gross, & Gotlib, 2007; Rottenberg, Gross, Wilhelm, Najmi, & Gotlib, 2002) using the EEB coding system have found group differences in facial expression between depressed and control participants. Additionally, both studies only found differences using the facial code for crying during a sad film clip, a code for which we had too few incidents to analyze. Therefore, the general lack of variance across studies using the EEB coding system may indicate it is not a sensitive enough coding system to differentiate diagnostic groups.

Response Coherence

Consistent with the weak situation hypothesis (Mischel, 1977), as expected, there was stronger coherence between subjective and facial responses during the emotional films compared to the neutral film. Participants also rated their feelings as more intense during the emotional film clips compared to the neutral film clip. This coincides with prior research, which has found greater response coherence across different emotion systems (i.e., subjective, behavioral, physiological) as emotional intensity increases (Lang et al., 1983; Mauss et al., 2005). The lack of response coherence during the neutral film clip may have been due to it being a weak situation; although it is also likely the result of restricted variance in facial response, as participants elicited little to no facial response while viewing the neutral film clip.

While researchers aim to have subjective reports of emotional experience *only* reflect the emotion experienced during the induction, there are several sources of information that can contribute to these reports. Robinson and Clore (2002) proposed an accessibility model of emotional self-report, which states that people can access four types of knowledge when self-reporting their emotions, ranging from most specific to most general. First, people can access their *experiential knowledge*, which constitutes an online, in-the-moment report of their current feelings. In order to access this information, the subjective report must be temporally close to the actual experience. Second, people can retrieve information from their *episodic memory*, which involves reconstructing specific moments from the past by recalling thoughts and details from the event. Third, people can access *situation-specific beliefs*, which are beliefs about emotions that occur during specific situations (i.e., ‘comedians are funny’). Finally, people can access *identity-related beliefs*, which involve beliefs about specific traits and stereotypes (i.e., ‘depressed people are sad’). The relative contribution of each

source of knowledge is dependent on the accessibility of that information, and when multiple sources are accessible there is a general preference to use the more specific information. What is important though is that each source of information can lead to a potentially different subjective emotional response.

In the present study, participants likely used varying degrees of each source of information while reporting their subjective reactivity during the different conditions. During the baseline measurement, participants were asked how they felt at that given moment, without referencing any specific context or situation. The generality of the baseline measurement suggests that there was a relatively greater contribution from *identity-related beliefs* (e.g., ‘depressed people [like me] are sad and unhappy’), and explains why group differences in positive and negative reactivity were greatest during the baseline assessment. Following the neutral film, participants were asked to rate their peak emotional reaction during the film clip. However, as previously mentioned, the neutral film clip was a relatively weak situation, and thus provided less direction for how to respond. The subjective emotion ratings were also administered immediately following the presentation of the neutral film. Therefore, there was likely a contribution from multiple sources, including *situation-specific beliefs* (e.g., ‘coastal landscapes are pleasant places’), *episodic memory* (e.g., ‘when I last went to the coast, I had a pleasant time’), and *experiential knowledge* (e.g., ‘right now I am feeling pleasant’), while rating emotional reactions to the neutral film clip. For the emotional films, participants again rated their peak emotional reaction immediately following the film clip. The higher emotional intensity of the films and close proximity between the film clip and subsequent self-report suggests a relatively greater contribution from *experiential knowledge* and *situation-specific beliefs*. We want to emphasize that the proposed contribution of each source of information above is merely a hypothesis, and we are sure that strong arguments could be made in favor of a different pattern. The important point we intend to make is that the relative contribution of each source of information likely differed amongst the conditions (i.e., baseline, neutral, emotional), and helps explain how depressed and control participants’ subjective emotional reactions differed during some, but not other, conditions.

Measuring Change in Emotional Reactivity

We also found different results using the various analytic approaches for measuring change. Depressed and control participants only differed in subjective emotional reactivity when using difference scores and PCSs. When the baseline measurement was used as the comparison condition, there were no differences between the difference scores and PCSs. However, when the neutral film clip was used as the comparison condition, depressed and control participants only differed in Subjective-Sadness using PCSs. The interpretation of these results is somewhat problematic though because the groups differed during the comparison conditions. For example, for both difference scores and PCSs, depressed participants had reduced Subjective-Sadness reactivity relative to controls. However, this was completely because of group differences during the comparison conditions, as depressed and control participants reported similar levels of Subjective-Sadness during the sad film clip. Considering just the difference scores alone suggests that depressed participants responded with less reactivity to the sad film clip, which is misleading (and possibly untrue).

While depressed and control participants did differ in subjective emotional reactivity using PCSs, it was unfortunately an inappropriate method to use with the subjective data. Calculating PCSs are only permissible using data measured on a ratio scale (Stevens, 1946) and our measure of subjective emotional response was assessed on an interval scale with an arbitrary zero point. Despite being an inappropriate method for the present study, we chose to include PCS because it is one of the approaches for measuring change used in the literature. Of all our measurements, the only one for which it was appropriate to calculate PCSs was Facial-Smiles (measured on a ratio scale). Therefore, while PCSs are a viable option toward measuring change, they should only be considered for data measured on a ratio scale.

Depressed and control participants did not differ in subjective emotional reactivity for either residualized change or ANCOVA. While we chose to use multiple analytic approaches for measuring change for illustrative purposes, ANCOVA was also an inappropriate method for our design. ANCOVA's primary purpose is to reduce noise caused by covariates, such as group differences present at baseline (Tabachnik & Fidell, 2007). However, ANCOVA should be avoided when random assignment is not possible and differences at baseline are systematically related to group affiliation (Chapman & Miller, 2001).

There are other potential ways to address the many issues with measuring change that were not covered in the present study. One option is to conduct simulation studies (e.g., Monte Carlo simulation) comparing the different analytic methods for measuring change that could provide additional leverage for identifying how the various approaches would influence results. Another option is to consider alternative experimental designs. The present study utilized a basic pretest-posttest design, where both groups received the manipulation (i.e., watching the film clips). However, there are other experimental designs that warrant consideration. For example, one option is to use a variant of the pretest-posttest control group design (Campbell & Stanley, 1963) where participants in both groups are randomized into a manipulation group and a non-manipulation control group.

Additional Methodological Issues to Consider

While the present study was not intended to address every important methodological issue related to measuring emotional reactivity, there are several others that warrant brief discussion. One issue is that subjective measures of emotional experience are not, by nature, on a ratio scale. This is problematic because the definition of the response anchors may have very different meanings across individuals. For example, a depressed individual may anchor their sadness ratings to where a '0' reflects the sadness experienced while in remission and an '8' reflects the sadness they experienced during their worst depressive episode. On the other hand, a control participant may anchor their sadness ratings to where a '0' reflects the sadness experienced on a typical day and an '8' reflects the sadness experienced after ending a romantic relationship. Additionally, the intervals between ratings would likely be qualitatively different. In sum, groups may differ in how they define and interpret the anchors and intervals of subjective measures of emotional experience. As previously mentioned, another problem with subjective responses is that, since they are not on a ratio scale, they are restricted to certain numerical operations.

Another issue is determining when to measure the emotional experience. Emotional reactions are not static, but rather dynamic events that unfold over time. As a result, there are different parameters of emotional responding (i.e., affective chronometry; Davidson, 1998) that can be assessed. One way to measure emotional experience is to collect retrospective emotion ratings after the induction period ends, similar to the present

study. However, even with this option there are further decisions to be made. For example, should participants rate their emotional response right now (i.e., after the induction) or should they rate their emotional experience during the induction? Should participants rate their peak or average response? However, as discussed above, retrospectively collected emotion ratings (even those collected immediately after an induction) are susceptible to multiple sources of influences compared to in-the-moment assessments (Robinson & Clore, 2002). Thus, another way to assess emotional reactivity is to measure it online as the response unfolds using methods such as affect rating dials (Ruef & Levenson, 2007). Online ratings are advantageous as they decrease the influence of other sources of information that may contaminate ratings. However, online subjective measurements also have the potential to interfere with other measurements, such as facial response, as participants may be tempted to look away while making their rating. In the end, no matter what method of measuring subjective reactivity is chosen, something is always sacrificed (Coan & Allen, 2007; Gray & Watson, 2007).

Inducing Positive Compared to Negative Emotions

Several researchers have argued that positive emotion induction is often more challenging than negative emotion induction in laboratory based paradigms (Gerrards-Hesse, Spies, & Hesse, 1994; Westermann, Spies, Stahl, & Friedrich, 1996). Consequently, the difficulty in inducing positive emotions will likely lead to reduced variance in measures of positive compared to negative emotional reactivity. Thus, when examining emotional reactivity, the relative difference between positive and neutral conditions will be smaller than the difference between negative and neutral conditions, and it may be mistakenly concluded that group differences are specific to negative emotions. This problem is analogous to Chapman and Chapman's (1973, 1978) argument regarding psychometrically matched tasks.

Normative vs. Idiographic Stimuli

For the present study we used standardized/normative and not idiographic emotional stimuli. Standardized stimuli are selected to elicit different emotions based on the fact that they previously elicited their targeted emotion(s) in other studies. Idiographic stimuli on the other hand are personally meaningful, individualized stimuli (e.g., personal narratives) intended to elicit a specified level of emotion reactivity (e.g., think of the saddest event you have ever experienced). Prior research comparing the pattern of emotional

reactivity between depressed and control participants has found differences between normative and idiographic stimuli (Rottenberg et al., 2005). Therefore, a future direction may be to examine whether idiographic stimuli present the same issues regarding measuring and defining reactivity. There are issues though with using idiographic stimuli. First, almost by definition, subjects (and thus between-subjects variables, such as depressed vs. controls) will not report different subjective emotional responses for idiographic stimuli as they are intended to elicit analogous levels of emotional responses. One potential solution to this challenge is to not measure subjective responses and only measure response in another system, such as physiological or behavioral. Second, groups are likely to differ in the idiographic stimuli they select (e.g., depressives are likely to have sadder narratives than non-depressed individuals).

Limitations

Our investigation included several limitations. First, we only assessed emotional reactivity via subjective and facial responses. Future studies may want to include physiological measures to determine whether similar issues exist when measuring heart rate, skin conductance, EMG, etc. Second, we did not include a baseline comparison measurement for facial response. However, as discussed above, it can be challenging to measure a person's baseline facial reactivity in which there are no situational demands and it may be questionable whether this would be a true measure of their baseline facial activity. Finally, the fear film clip elicited minimal Facial-Fear (although it elicited more Facial-Fear than any other facial response) and the inter-rater reliability was very low (.10). We recognize this is far below acceptable standards, and suggest that our findings for Facial-Fear should be interpreted with caution.

Recommendations for 'Better Practices'

As illustrated throughout this paper, there are many important methodological decisions that need to be made when designing studies of emotional reactivity in psychopathology. While we believe that there are no 'best practices', as every decision contains strengths and weaknesses, we think it is important to briefly make recommendations for what we consider 'better practices' in the two domains that we examined (comparison condition and analytic approach). We first want to emphasize that results which replicate across comparison conditions and analytic approaches for measuring change should be considered the strongest (and

potentially most valid). However, since this outcome may be quite rare, we make the following suggestions. First, we recommend that studies of emotional reactivity use more than one comparison condition, preferably of varying situational strength. While this may raise concerns that a pluralistic approach may increase the risk of spurious findings, we believe that this risk can be limited by discussing any differences between comparison conditions in terms of reflecting important substantive phenomena. Second, given the statistical limitations with PCS, ANCOVA, and residualized change, and the interpretive problems with a ‘no comparison’ condition, we recommend using change scores when measuring emotional reactivity (especially for subjective responses). However, it is imperative to examine each source of the change score (i.e., comparison and induction conditions) independently and thus determine whether differences in change scores are due to differences during the comparison condition, induction condition, or both.

Summary

In the present study, we examined the effects of choosing different comparison conditions and analytic approaches for measuring change. We found that making such decisions can alter the pattern of results and, more importantly, conclusions drawn regarding various theories of emotional reactivity in depression. Overall, we hope that raising these issues will further the discussion of what constitutes ideal methods for comparing group differences in emotional reactivity. We also want to re-iterate that the points we raised are not specific to examining emotional reactivity in depression, but apply to other psychological constructs as well, such as schizophrenia (Kring & Moran, 2008), anxiety (McNeil et al., 1993), borderline personality disorder (Kuo & Linehan, 2009), and smoking (Kassel, Evatt, Greenstein, Wardle, Yates, & Veilleux, 2007).

References

- American Psychiatric Association. (1994). *Diagnostic and Statistic Manual of Mental Disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- Blomqvist, N. (1977). Relation between change and initial value. *Journal of the American Statistical Association*, 72, 746-749.
- Blumenthal, T. D., Elden, A., & Flaten, M. A. (2004). A comparison of several methods used to quantify prepulse inhibition of eyeblink responding. *Psychophysiology*, 41, 326-332.
- Beck, A. T. (1976). *Cognitive therapy and the emotional disorders*. Madison, CT: International Universities Press.
- Beck, A. T., Rush, J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy for depression*. New York: The Guilford Press.
- Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 3-20). Madison, WI: University of Wisconsin Press.
- Bylsma, L. M., Morris, B. H., & Rottenberg, J. (2008). A meta-analysis of emotional reactivity in major depressive disorder. *Clinical Psychology Review*, 28, 676-691.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, 101, 147-158.
- Campbell, D. T. & Stanely, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.
- Caspi, A., & Moffitt, T. E. (1993). When do individual differences matter? A paradoxical theory of personality coherence. *Psychological Inquiry*, 4, 247-271.
- Chapman, L. J. & Chapman J. P. (1973). Problems in the measurement of cognitive deficits. *Psychological Bulletin*, 79, 380-395.
- Chapman, L. J., & Chapman, J. P. (1978). Measurement of differential deficit. *Journal of Psychiatric Research*, 14, 303-311.

- Chentsova-Dutton, Y. E., Chu, J. P., Tsai, J. L., Rottenberg, J., Gross, J. J., & Gotlib, I. H. (2007). Depression and emotional reactivity: Variation among Asian Americans of East Asian descent and European Americans. *Journal of Abnormal Psychology, 116*, 776-785.
- Clark, L. A. & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100*, 316-336.
- Clark, L. A., Watson, D., & Mineka, S. (1994). Temperament, personality, and the mood and anxiety disorders. *Journal of Abnormal Psychology, 103*, 103-116.
- Coan, J. A., & Allen, J. J. B. (2007). Introduction: Organizing the tools and methods of affective science. In J. A. Coan, & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 3-8). New York, NY: Oxford University Press.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cooper, W. H., & Withey, M. J. (2009). The strong situation hypothesis. *Personality and Social Psychology Review, 13*, 62-72.
- Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological Bulletin, 74*, 68-80.
- Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease. Special Issue: Mental Disorders in their Natural Settings: The Application of Time Allocation and Experience-Sampling Techniques in Psychiatry, 175*, 526-536.
- Davidson, R. J. (1998). Affective style and affective disorders: Perspectives from affective neuroscience. *Cognition and Emotion, 12*, 307-330.
- Davidson, R. J., Pizzagalli, D., Nitschke, J. B., & Putnam, K. (2002). Depression: Perspectives from affective neuroscience. *Annual Review of Psychology, 53*, 545-574.
- Dearing, K. F., & Gotlib, I. H. (2009). Interpretation of ambiguous information in girls at risk for depression. *Journal of Abnormal Child Psychology: An Official Publication of the International Society for Research in Child and Adolescent Psychopathology, 37*, 79-91.

- Depue, R. A., & Iacono, W. G. (1989). Neurobehavioral aspects of affective disorders. *Annual Review of Psychology*, 40, 457-492.
- Dunn, B. D., Dalgleish, T., Lawrence, A. D., Cusack, R., & Ogilvie, A. D. (2004). Categorical and dimensional reports of experienced affect to emotion-inducing pictures in depression. *Journal of Abnormal Psychology*, 113, 654-660.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1996). Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV). Washington, D.C.: American Psychiatric Press.
- Forbes, A. B., & Carlin, J. B. (2005). "Residual change" analysis is not equivalent to analysis of covariance. *Journal of Clinical Epidemiology*, 58, 540-541.
- Gerrards-Hesse, A., Spies, K., & Hesse, F. W. (1994). Experimental inductions of emotional states and their effectiveness: A review. *British Journal of Psychology*, 85, 55-78.
- Goldin, P. R., Manber, T., Hakimi, S., Canli, T., & Gross, J. J. (2009). Neural bases of social anxiety disorder: Emotional reactivity and cognitive regulation during social and physical threat. *Archives of General Psychiatry*, 66, 170-180.
- Gotlib, I. H., Joormann, J., Minor, K. L., & Hallmayer, J. (2008). HPA axis reactivity: A mechanism underlying the associations among 5-HTTLPR, stress, and depression. *Biological Psychiatry*, 63, 847-851.
- Gray, E. K., & Watson, D. (2007). Assessing positive and negative affect via self-report. In J. A. Coan, & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 171-183). New York, NY: Oxford University Press.
- Greenstein, J. E., & Kassel, J. D. (2010). The effects of smoking on selective attention as measured by startle reflex, skin conductance, and heart rate responses to auditory startle stimuli. *Psychophysiology*, 47, 15-24.
- Gross, J. J. (1999). Emotion and emotion regulation. In Pervin, L. A. & John, O. P. (Eds.), *Handbook of personality: Theory and research* (2nd ed.) (pp. 525-552). New York, NY: Guilford Press.

- Gross, J. J. & Levenson, R. W. (1993). Emotional suppression: Physiology, self-report, and expressive behavior. *Journal of Personality and Social Psychology*, 64, 970-986.
- Gross, J. J. & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion*, 9, 87-108.
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, 23, 56-62.
- Harris, C. W. (Ed.) (1963). *Problems in measuring change*. Madison, WI: University of Wisconsin Press.
- Hayes, R. J. (1988). Methods for assessing whether change depends on initial-value. *Statistics in Medicine*, 7, 915-927.
- Heller, W., & Nitschke, J. B. (1997). Regional brain activity in emotion: A framework for understanding cognition in depression. *Cognition and Emotion. Special Issue: The Cognitive Psychology of Depression*, 11, 637-661.
- Herbener, E. S., Song, W., Khine, T. T., & Sweeney, J. A. (2008). What aspects of emotional functioning are impaired in schizophrenia? *Schizophrenia Research*, 98, 239-246.
- Kassel, J. D., Evatt, D. P., Greenstein, J. E., Wardle, M. C., Yates, M. C., & Veilleux, J. C. (2007). The acute effects of nicotine on positive and negative affect in adolescent smokers. *Journal of Abnormal Psychology*, 116, 543-553.
- Kazdin, A. E. (2002). *Research design in clinical psychology* (4th ed.). Boston: Allyn and Bacon.
- Keppel, G. & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). New Jersey: Prentice Hall.
- Keller, M. B., Klein, D. N., Hirschfeld, R. M. A., Kocsis, J. H., McCullough, J. P., Miller, I., et al. (1995). Results of the DSM-IV Mood Disorders Field Trial. *American Journal of Psychiatry*, 152, 843-849.
- Klein, D. N., Ouimette, P. C., Kelly, H. S., Ferro, T., & Riso, L. P. (1994). Test-retest reliability of team consensus best-estimate diagnoses of Axis I and II disorders in a family study. *American Journal of Psychiatry*, 151, 1043-1047.
- Klein, D. N., Schwartz, J. E., Rose, S., & Leader, J. B. (2000). Five-year course and outcome of dysthymic disorder: A prospective, naturalistic follow-up study.

- American Journal of Psychiatry*, 157, 931-939.
- Kring, A. M., & Moran, E. K. (2008). Emotional response deficits in schizophrenia: Insights from affective science. *Schizophrenia Bulletin*, 34, 819-834.
- Kuo, J. R. & Linehan, M. M. (2009). Disentangling emotion processes in borderline personality disorder: Physiological and self-reported assessment of biological vulnerability, baseline intensity, and reactivity to emotionally evocative stimuli. *Journal of Abnormal Psychology*, 117, 531-544.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL.
- Lang, P. J., Levin, D. N., Miller, G. A., & Kozak, M. J. (1983). Fear behavior, fear imagery, and the psychophysiology of emotion: The problem of affective response integration. *Journal of Abnormal Psychology*, 92, 276-306.
- Lawson, C., & MacLeod, C. (1999). Depression and the interpretation of ambiguity. *Behaviour Research and Therapy*, 37, 463-474.
- Lissek, S., Pine, D. S., & Grillon, C. (2006). The strong situation: A potential impediment to studying the psychobiology and pharmacology of anxiety disorders. *Biological Psychology*, 72, 265-270.
- Lord, F. M. (1958). Further problems in the measurement of growth. *Educational and Psychological Measurement*, 18, 437-451.
- Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion*, 5, 175-190.
- McFarland, B. R., & Klein, D. N. (2009). Emotional reactivity in depression: Diminished responsiveness to anticipated reward but not to anticipated punishment or to nonreward or avoidance. *Depression and Anxiety*, 26, 117-122.
- McNeil, D. W., Vrana, S. R., Melamed, B. G., Cuthbert, B. N., & Lang, P. J. (1993). Emotional imagery in simple and social phobia: Fear versus anxiety. *Journal of Abnormal Psychology*, 102, 212-225.

- Miller, G. M., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110*, 40-48.
- Mischel, W. (1977). The interaction of person and situation. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology* (pp. 333-352). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Mogg, K., Bradbury, K. E., & Bradley, B. P. (2006). Interpretation of ambiguous information in clinical depression. *Behaviour Research and Therapy, 44*, 1411-1419.
- Mogg, K., Bradley, B. P., Williams, R., & Mathews, A. (1993). Subliminal processing of emotional information in anxiety and depression. *Journal of Abnormal Psychology, 102*, 304-311.
- Peeters, F., Nicolson, N. A., Berkhof, J., Delespaul, P., & deVries, M. (2003). Effects of daily events on mood states in major depressive disorder. *Journal of Abnormal Psychology, 112*, 203-211.
- Ruef, A. M., & Levenson, R. W. (2007). Continuous measurement of emotion: The affect rating dial. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 286-297). New York: Oxford University Press.
- Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin, 128*, 934-960.
- Rogosa, D. R. (1995). Myths and methods: Myths about longitudinal research plus supplemental questions. In J. M. Gottman (Ed.), *The analysis of change* (pp. 3-66). Mahwah, NJ: Lawrence Erlbaum.
- Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 92*, 726-748.
- Rosenberg, E. L. (1998). Levels of analysis and the organization of affect. *Review of General Psychology, 2*, 247-270.
- Rottenberg, J. (2005). Mood and emotion in major depression. *Current Directions in Psychological Science, 14*, 167-170.
- Rottenberg, J., Gross, J. J., & Gotlib, I. H. (2005). Emotion context insensitivity in major depressive disorder. *Journal of Abnormal Psychology, 114*, 627-639.

- Rottenberg, J., Gross, J. J., Wilhelm, F. H., Najmi, S., & Gotlib, I. H. (2002). Crying threshold and intensity in major depressive disorder. *Journal of Abnormal Psychology, 111*, 302-312.
- Rottenberg, J., Kasch, K. L., Gross, J. J., & Gotlib, I. H. (2002). Sadness and amusement reactivity differentially predict concurrent and prospective functioning in major depressive disorder. *Emotion, 2*, 135-146.
- Ruef, A. M., & Levenson, R. W. (2007). Continuous measurement of emotion: The affect rating dial. In J. A. Coan, & J. J. B. Allen (Eds.), *Handbook of emotional elicitation and assessment* (pp. 286-297). New York, NY: Oxford University Press.
- Scher, C. D., Ingram, R. F., & Segal, Z. V. (2005). Cognitive reactivity and vulnerability: Empirical evaluation of construct activation and cognitive diatheses in unipolar depression. *Clinical Psychology Review, 25*, 487-510.
- Shankman, S. A., & Klein, D. N. (2003). The relation between depression and anxiety: An evaluation of the tripartite, approach-withdrawal and valence-arousal models. *Clinical Psychology Review, 23*, 605-637.
- Shankman, S. A., Klein, D. N., Tenke, C. E., & Bruder, G. E. (2007). Reward sensitivity in depression: A biobehavioral study. *Journal of Abnormal Psychology, 116*, 95-104.
- Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.
- Snyder, M., & Ickes, W. (1985). Personality and social behavior. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 2, pp. 883-947). New York: Random House.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677-680.
- Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine, 24*, 236-243.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn and Bacon.
- Twisk, J., & Proper, K. (2004). Evaluation of the results of a randomized controlled trial: How to define changes between baseline and follow-up. *Journal of Clinical Epidemiology, 57*, 223-228.

- Watson, D. & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98, 219-235.
- Westermann, R., Spies, K., Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: A meta-analysis. *European Journal of Social Psychology*, 26, 557-580.
- Wyer Jr., R. S. & Collins II, J. E. (1992). A theory of humor elicitation. *Psychological Review*, 99, 663-688.

Footnotes

¹ A statistically identical approach that is often used is a mixed-design ANOVA with one within-subjects factor (i.e., baseline vs. emotion induction) and one between-subjects factor (i.e., depressed vs. controls).

² A statistically identical approach that is often used is to conduct an ANCOVA with difference scores as the dependent variable and the comparison condition score as the covariate.

³ We also examined whether there were any main effects of gender for both subjective and facial responses. Results indicated a main effect of gender for Subjective-Sadness PCS using the baseline comparison condition, $F(1, 100) = 4.13, p < .05$, with women reporting greater sadness reactivity relative to men. In addition, there was also a main effect of gender for Facial-Fear during the fear film, $F(1, 100) = 6.70, p < .05$, with men exhibiting more facial fear relative to women.

⁴ We conducted additional analyses examining whether depressed participants with comorbid anxiety differed from those with depression only. In general, the results were nearly identical between the two depressed groups, with the only significant difference occurring for Subjective-Sadness difference scores using the neutral comparison condition. Results indicated that participants with depression and a current anxiety disorder had reduced Subjective-Sadness reactivity relative to controls ($p < .05$), but they did not differ from participants with depression only. We also examined whether depressed participants currently taking medication differed from depressed participants who were medication free. Results indicated that depressed participants currently taking medication had greater Subjective-Happiness reactivity when using the baseline comparison condition for both difference scores ($p < .05$) and PCSs at a trend level, ($p < .07$). The groups also differed in their raw Subjective-Fear during the neutral film clip, with medicated depressed participants reporting more Subjective-Fear than non-medicated depressed participants at a trend level ($p < .06$).

Table 1. Example data demonstrating issues with using difference scores.

Group	Pre-induction	Post-induction	Difference score
A	4	6	2
B	1	6	5

Table 2. Example data demonstrating issues with using percentage change scores.

Group	Pre-induction	Post-induction	Difference score	PCS
A	4	6	2	50.0%
B	6	8	2	33.3%

Note. PCS = percentage change score.

Table 3. Demographic and clinical characteristics of sample

	Control (<i>N</i> = 37)	MDD (<i>N</i> = 69)
Demographic variables		
Female	73.0%	66.7%
Age	33.4 yrs	34.4 yrs
Caucasian	73.0%	79.7%
Employed	81.1%	78.3%
Clinical variables		
Global Assessment of Function (GAF)	85.1	53.7
Hamilton Rating Scale of Depression (HRSD; SD)	1.7 (1.8)	26.0 (7.9)
Currently taking medication	-	50.7%

Note. MDD = major depressive disorder, SD = standard deviation.

Table 4. Means (and standard deviations) of emotional response by condition, diagnostic group, and response type

	Baseline condition			Neutral condition			Emotional condition		
	MDD	Control	<i>p-value</i>	MDD	Control	<i>p-value</i>	MDD	Control	<i>p-value</i>
Subjective-Sadness	2.49 (1.91)	0.35 (1.11)	< .001	1.41 (1.87)	0.14 (0.35)	< .001	4.81 (2.11)	4.31 (2.36)	<i>ns</i>
Subjective-Happiness	2.57 (1.41)	4.30 (1.49)	< .001	2.26 (1.95)	3.14 (2.26)	< .05	3.20 (1.83)	3.97 (1.93)	< .05
Subjective-Fear	0.90 (1.41)	0.19 (0.57)	< .01	0.35 (0.98)	0.03 (0.16)	= .05	2.83 (2.12)	3.14 (2.50)	<i>ns</i>
Facial-Sadness				0.00 (0.00)	0.00 (0.00)	<i>ns</i>	1.19 (1.46)	1.16 (1.39)	<i>ns</i>
Facial-Happiness				0.09 (0.41)	0.03 (0.17)	<i>ns</i>	2.15 (1.56)	2.28 (1.61)	<i>ns</i>
Facial-Smiles				0.06 (0.24)	0.12 (0.54)	<i>ns</i>	3.44 (3.30)	4.29 (4.07)	<i>ns</i>
Facial-Fear				0.00 (0.00)	0.00 (0.00)	<i>ns</i>	0.52 (0.99)	0.54 (1.02)	<i>ns</i>

Note. MDD = major depressive disorder.

Table 5. Comparison of subjective emotional reactivity between depressed and control participants evaluated with different comparison conditions and different analytic approaches for defining change

Film Clip	Baseline comparison			Neutral comparison		
	<i>F</i> statistic	<i>P</i> value	MDD vs. Controls	<i>F</i> statistic	<i>P</i> value	MDD vs. Controls
Amusing Film - Happiness						
Difference score	6.33	<.05	MDD > Controls	0.06	.81	<i>ns</i>
PCS	6.17	<.05	MDD > Controls	0.16	.70	<i>ns</i>
Residualized change	0.03	.86	<i>ns</i>	1.40	.24	<i>ns</i>
ANCOVA	0.04	.84	<i>ns</i>	1.44	.23	<i>ns</i>
Sad Film - Sadness						
Difference score	8.79	<.01	Controls > MDD	2.39	.13	<i>ns</i>
PCS	24.22	<.001	Controls > MDD	7.97	<.01	Controls > MDD
Residualized change	0.45	.51	<i>ns</i>	0.14	.71	<i>ns</i>
ANCOVA	0.61	.44	<i>ns</i>	0.16	.69	<i>ns</i>
Fear Film - Fear						
Difference score	4.92	<.05	Controls > MDD	1.89	.17	<i>ns</i>
PCS	6.13	<.05	Controls > MDD	3.27	.07	<i>ns</i>
Residualized change	1.93	.17	<i>ns</i>	1.06	.31	<i>ns</i>
ANCOVA	2.07	.15	<i>ns</i>	1.09	.30	<i>ns</i>

Note. MDD = major depressive disorder, PCS = percentage change score, ANCOVA = analysis of covariance.