American Psychological Association

Psychological Bulletin

ISSN: 0033-2909

2020, Vol. 146, No. 4, 355–375
http://dx.doi.org/10.1037/bul0000224

The Manipulation of Affect: A Meta-Analysis of Affect Induction Procedures

Dana L. Joseph
University of Central Florida

Micaela Y. Chan
University of Texas at Dallas

Samantha J. Heintzelman
Rutgers University

Louis Tay
Purdue University

Ed Diener
University of Virginia and University of Utah

Victoria S. Scotney
Purdue University

Affect inductions have become essential for testing theories of affect and for conducting experimental research on the effects of mood and emotion. The current review takes stock of the vast body of existing literature on affect induction procedures (AIPs; also referred to as mood inductions) to evaluate the effectiveness of affect inductions as research tools and to test theories of affect (e.g., the bipolarity hypothesis, negativity bias, positivity offset, and theories of emotionality and gender) using meta-analytic data. In doing so, we seek to address whether AIPs are effective for inducing affective states, what conditions maximize their effectiveness, for which emotions they are most effective, for whom they are most effective, and whether affect induction findings can provide insight into theories of affect. A meta-analysis of 874 samples and 53,509 participants suggests that affect inductions are effective on average ($\delta = 1.32$), but this effectiveness varies with the type of affect induction, the emotion being induced, and the gender of the participants. Further, results indicate coupled activation where the induction of positive (negative) emotions leads to a corresponding reduction in negative (positive) emotions, which provides support for the bipolar continuum of positive and negative affect. Results also revealed a negativity bias in which individuals display stronger reactions to negative stimuli than positive stimuli. A practical guide in the choice of affect induction procedures for researchers is presented and implications for emotion theory are discussed.

Public Significance Statement
This meta-analysis illustrates the effectiveness of affect inductions as a tool for changing one's emotional state. However, the effectiveness of these procedures varies substantially across how, why, and to whom these inductions are administered, illustrating the importance of careful thought when choosing the best way to enhance (or reduce) one's emotional state.

Keywords: emotion, meta-analysis, mood induction, positive/negative affect

Supplemental materials: http://dx.doi.org/10.1037/bul0000224.supp

The influence of emotion spans across nearly all aspects of human psychology, including the study of memory (Bower, 1981; Isen, Shalker, Clark, & Karp, 1978), creativity (Baas, De Dreu, & Nijstad, 2008; Isen, Daubman, & Nowicki, 1987), problem solving (Gasper, 2003), decision making (Lerner, Li, Valdesolo, & Kassam, 2015), persuasion (Bless, Bohner, Schwarz, & Strack, 1990; Petty & Briñol, 2015), and stereotyping (Bodenhausen, Kramer, & Süsser, 1994; Bless, Schwarz, & Kemmelmeier, 1996; Park &

This article was published Online First January 23, 2020.
Dana L. Joseph, Department of Management, University of Central Florida; Micaela Y. Chan, School of Behavioral and Brain Sciences, University of Texas at Dallas; Samantha J. Heintzelman, Department of Psychology, Rutgers University; Louis Tay, Department of Psychological Sciences, Purdue University; Ed Diener, Department of Psychology, University of Virginia, and Department of Psychology, University of Utah; Victoria S. Scotney, Department of Psychological Sciences, Purdue University.

Correspondence concerning this article should be addressed to Dana L. Joseph, Department of Management, University of Central Florida, 12744 Pegasus Drive, Orlando, FL 32816. E-mail: dana.joseph@ucf.edu

Banaji, 2000). Historically, researchers have utilized two approaches to understanding the effects of emotion on cognition and behavior (Cronbach, 1957): examining the relationship between naturally occurring emotion and outcomes, or experimentally manipulating emotion to assess the effect of the manipulation on outcomes. The former correlational approach is limited in the extent to which one can infer causal relationships between emotion and cognition/behavior. In contrast, the latter experimental approach, in which an affective state is induced, allows for stronger causal inferences. Given broad interest in understanding the causal effects of affect, this approach to studying emotions has become widespread across the field of psychology. For example, Fredrickson and Branigan (2005) tested their broaden-and-build theory of positive emotions (Fredrickson, 1998, 2001) by inducing emotions such as amusement, contentment, anger, and anxiety. Similarly, affect inductions were central to Larsen and Ketelaar's (1989) test of Gray's model of personality functioning (Gray, 1981) in which extroverts' and neurotics' sensitivity to positive and negative affect inductions mapped on to emotional reactivity profiles associated with these personality traits (for additional evidence of affect induction influence, see early reviews by Clark, 1983 and Goodwin & Williams, 1982). Given the ubiquity of this experimental technique, understanding how to manipulate emotions, which manipulations are most effective, and for whom these manipulations are most effective is essential for advancing the science of emotions.

Therefore, the purpose of this paper is to review and evaluate the effectiveness of affect induction procedures (AIPs), or strategies used to experimentally manipulate a participant's affective state. In this effort, our first goal is to evaluate the effectiveness of AIPs as a research tool. Given that AIPs have been criticized for producing only minimal amounts of change in the intended emotional state (Marston, Hart, Hileman, & Faunce, 1984) and no existing meta-analysis has evaluated the entire body of AIP literature to determine the accuracy of these criticisms, this is an important step in evaluating whether AIPs deserve their current popularity as a research tool. Thus, we seek to evaluate the efficacy of AIPs and the features of AIPs that enhance their efficacy in research and practice. Although this has clear implications within the field of psychology, we argue that evaluating AIP effectiveness has implications for the study of emotion beyond psychology as well, in other fields such as marketing (e.g., the effect of emotion on purchasing decisions), engineering (e.g., the effect of emotion on human-robot interactions), healthcare (e.g., the effect of patient emotion on medical advice adherence), and political science (e.g., the effect of emotion on voter behavior).

Beyond investigating the practical utility of AIPs, a review of AIP effectiveness can also deepen our theoretical understanding of emotions. For instance, our review can aid in addressing the ongoing debate regarding the extent to which positive emotions and negative emotions are bipolar opposites or if they are independent and can be experienced simultaneously (Russell & Carroll, 1999). Thus, our second goal is to seek an answer for the questions, "What happens to positive emotions when negative emotions are induced?" and "What happens to negative emotions when positive emotions are induced?" As a third goal, the current review also seeks to understand the characteristics of individuals (e.g., male or female) who react most strongly to AIPs. Finally, we seek to provide comparisons across discrete emotions to contribute to literature on emotion regulation by noting which emotions are particularly sticky (and difficult to change via intervention) or mutable (and comparatively easier to change via intervention).

Affect Induction Procedures

Prior to introducing various AIPs, we would first like to clarify that although we use the term *affect induction* in the current paper, the more commonly used term for this procedure in the field is *mood induction*. To explain our decision to use a term that is inconsistent with existing scholarly work, we note that *emotions* involve short-duration states in response to a stimulus, whereas *moods* are longer-duration states that are not in response to a particular stimulus, and *affect* is an umbrella term that incorporates both mood and emotion (Ekman, 1994; Gross, 2010). Because AIPs are stimuli that may elicit both short-term and longer-term states (i.e., emotion and mood), we use the term *affect induction* in the current review. In Table 1, we present a summary of the AIPs that are commonly utilized to manipulate emotion in an experimental context (for an in-depth review of each AIP type, see Martin, 1990). These procedures range from those that ask participants to actively participate in the induction by imagining a scenario, recalling an event, or contracting a particular set of facial muscles, for example, to those that involve more passive participation where the individual is asked to listen to music, watch a film, or look at photographs, for example.

Given the popularity of AIPs in research, it comes as no surprise that the AIP literature has been previously reviewed. Early reviews of the AIP literature did not provide meta-analytic effect sizes but did provide qualitative evidence that AIPs are most effective when the intent of the induction is known to participants (Martin, 1990) and when the AIP is inducing a negative rather than positive emotion (Gerrards-Hesse, Spies, & Hesse, 1994). Other reviews sought to evaluate and/or compare specific types of affect induction procedures (Albersnagel, 1988; Brenner, 2000; Kenealy, 1986). Later reviews of the literature provided meta-analytic effect sizes that evaluated a single type of AIP (e.g., the Velten AIP; Larsen & Sinnett, 1991; Nummenmaa & Niemi, 2004) or focused on specific discrete emotions such as elation and depression (Westermann, Spies, Stahl, & Hesse, 1996). More recently, a meta-analysis by Lench, Flores, and Bench (2011) sought to examine the extent to which AIPs impact *non-affect* outcomes including physiological, behavioral, and cognitive outcomes (e.g., do happiness inductions enhance creativity?). Although Lench et al. report the effect of affect inductions on affect (i.e., the focus of the current paper), they only present effect sizes representing whether each of these AIPs change *any* self-report of emotion or arousal (e.g., whether a happiness AIP induces anxiety) as opposed to whether affect inductions induce the intended emotion (e.g., whether a happiness AIP induces happiness), which is the focus of the current paper.

Thus, although the AIP literature has been reviewed several times in the past, a meta-analytic evaluation of the effectiveness of the wide range of existing AIPs in inducing the emotion they are intended to induce does not exist. Thus, we seek to evaluate

Table 1
Affect Induction Procedures

Method	Description
Velten	• Participants read 60 elated or depressed self-referential statements, and are asked to feel the suggested affective state • Modified version includes shorter versions (e.g., Terhaar et al., 2010) or versions with specific content removed (e.g., Henderson & Lohr, 1982)
Imagination	• Participants are instructed to imagine a given scenario that involves a specific affective state • Participants may also be instructed to write down their thoughts
Autobiographical recall	• Participants are instructed to recall an autobiographical event that involves a specific affective state • Participants may also be instructed to write down the scenario and feelings they re-experienced
Film	• Participants watch a film that involves a specific affective state • Emotionally contagious films (ECOFs), or short films that do not contain a story but only show the head and shoulders of a person who expresses a particular emotion are sometimes used (Papousek, Schulter, & Lang, 2009)
Reading a story	• Participants read a story or vignette that involves a specific affective state
Music/sounds	• Participants listen to music or sounds that are meant to reflect a specific affective state • To induce sadness, the music is often manipulated to be slower than the original
Pictures	• Photographs of emotionally-charged situations or stimuli are used to induce a specific affective state • The International Affective Picture System (Lang, Bradley, & Cuthbert, 2008) is one of the most common stimuli • Pictures of faces can also serve as the stimuli (e.g., Schneider, Gur, Gur, & Muenz, 1994)
Feedback	• Participants are given a bogus test of ability (e.g., math), and are given predetermined positive or negative feedback regarding their performance to induce positive or negative emotion, respectively
Coping challenge	• Participants are put in a challenging situation that requires active coping (e.g., giving a public speech or performing a difficult math task in front of others) • Most often, this is used to induce anxiety
Manipulation of face/body	• Participants are instructed to contract specific muscles of the face or maintain a specific body posture that represents a particular affective state • Most often, participants are not told what emotion/posture/facial expression to make (e.g., they are told to show their teeth and contract the zygomaticus muscle instead of being told to smile)
Jokes/cartoons	• Participants are asked to read jokes or are shown cartoons, most often to induce a positive affective state
Odor	• Participants are exposed to a scent, most often to induce disgust
Video recording/social pressure	• Participants are told they will be video recorded or are put under some form of social pressure (in some cases they are recorded and in others the instructions are the only induction) • Most often, this is used to induce anxiety

the effectiveness of AIPs in the current study. In doing so, we also provide methodological advances over previous reviews by not restricting our search to specific journals (cf. Westermann et al., 1996), not excluding dissertations (a practice that can inflate effect sizes; cf. Lench et al., 2011), and using appropriate meta-analytic techniques to combine across between-person and within-person effects sizes (Morris & DeShon, 2002), which no prior review has done. Below, we discuss a series of six practical and theoretical questions that we seek to address in the current paper.

Are AIPs Effective and Which AIPs Are Most Effective?

Despite the ubiquitous use of AIPs, there appears to be an equally large body of criticism regarding whether AIPs induce the intended emotional state. Although induction procedures have evolved well beyond early elicitation methods (e.g., asking participants to decapitate a rat to induce disgust; Landis, 1924), some have suggested that modern AIPs only offer mild elicitations of the intended emotion (Marston et al., 1984; Philippot, 1993). Thus, the current paper seeks to examine whether AIPs are effective (i.e., do they result in a significant increase in the intended emotion) and the extent to which they are effective (i.e., is this change minimal, as previously suggested).

Equally important, we also ask "Which type of AIP is the most effective?". Prior reviews have produced equivocal results upon comparing AIP types. Whereas some have lauded the performance of music to elicit emotions (Albersnagel, 1988; Martin, 1990), others have found support for pictures (Lench et al., 2011) or film (Westermann et al., 1996) as the most effective type of AIP. Thus, it is unclear which type of AIP (see Table 1 for a summary of types) is most successful in eliciting the intended emotion.

Interestingly, prior reviews of the AIP literature have consistently referred to the Velten procedure, which involves reading elated or depressed self-referential statements and attempting to feel the emotion suggested by each (Velten, 1968), as "by far the most widely used AIP" (Westermann et al., 1996, p. 559; see also Kenealy, 1986). In this vein, we ask, "Which AIP is most commonly used?" in hopes of evaluating whether the most popular AIPs are also the most effective. If results suggest that popularly used AIPs are ineffective, this would have implications for the interpretation of a wide array of findings across multiple research areas.

What Enhances the Effectiveness of AIPs?

Instruction

AIPs can be distinguished on whether participants are explicitly instructed to feel a particular emotion (e.g., "Become fully involved and try as hard as possible to feel the mood suggested"; Sutherland, Newman, & Rachman, 1982, p. 129), or

whether the participants are not given instructions to feel a particular emotion (e.g., "Please watch the film carefully"; Rottenberg, Ray, & Gross, 2007, p. 12) and therefore do not know the intended effect of the stimuli. Although there is some evidence that deliberate instruction increases AIP effectiveness (e.g., Lenton & Martin, 1991), some argue that this may be an artificial inflation (i.e., not representative of real emotion change) that is due solely to demand characteristics (Buchwald, Strack, & Coyne, 1981; Clark, 1983; Goodwin & Williams, 1982; Westermann et al., 1996). Thus, the current analysis investigates AIP effectiveness with and without instruction to not only evaluate whether explicit instructions increase the effectiveness of AIPs, but also to evaluate whether AIPs are still effective even without instruction (and can be used with fewer concerns about demand characteristics).

Instructions may also be delivered truthfully (e.g., "Listen to the music and try to feel the emotions present within the music") or with deception (e.g., "Listen to the music. After the music has ended, you will be asked to evaluate how much you like/dislike the music."). Any positive change in the intended mood exhibited as a result of an AIP delivered with deception is less likely to be due to demand characteristics because participants are less likely to be aware of the purpose of the study. Therefore, we also investigate whether AIPs delivered truthfully exhibit larger effect sizes than those delivered with deception, and whether those delivered with deception are still effective, making them suitable for use while reducing concerns about demand characteristics.

Single Versus Combined AIPs

In addition to deciding whether one should use instructions and/or deception, the administrator of an AIP must also decide whether to deliver a single AIP or multiple AIPs in combination. Multiple AIPs are sometimes used in a single experiment with the intended goal of increasing the effect of the induced emotion. For example, music AIPs are commonly added to the background of other AIPs. Although combined AIPs tend to show larger effects than single AIPs when inducing a depressed state (Westermann et al., 1996), it is unclear whether this extends to other emotions. Thus, we evaluate the effectiveness of single AIPs versus combined AIPs, and we compare common combinations of AIPs to determine their relative effectiveness.

Which Emotions Are the Easiest to Elicit?

Positive Versus Negative Affect

The negativity bias refers to the human tendency to attend to and give greater weight to negative stimuli compared with positive stimuli (Cacioppo & Berntson, 1994). There are several ways in which the negativity bias can manifest, including negative potency, or the tendency for negative stimuli to be more salient than positive stimuli of equal magnitude (Rozin & Royzman, 2001). Some have argued that the negativity bias is evolutionarily adaptive because it predisposes our attentional processes to pay extreme attention to negative stimuli, allowing us to evade dangerous, life-threatening stimuli (Vaish, Grossmann, & Woodward, 2008). Extended to the AIP literature, the negativity bias would suggest that negative AIPs should induce negative emotion to a greater extent than positive AIPs induce positive emotion because humans are predisposed to attend to negative stimuli more than positive stimuli. Interestingly, no meta-analytic point estimate of the negativity bias exists. Therefore, although primary study evidence has found support for the negativity bias in affective judgments (Ito, Cacioppo, & Lang, 1998), social information processing (Fiske & Taylor, 1991), event-related brain potentials (Ito, Larsen, Smith, & Cacioppo, 1998), and even in infants and animals (Miller, 1961; Vaish et al., 2008), we do not have composite evidence regarding the strength of the negativity bias phenomenon in general. Thus, as a test of the negativity bias, we evaluate whether negative AIPs elicit negative emotion to a greater extent than positive AIPs elicit positive emotion.

Discrete Emotions

Although emotion states elicited by AIPs can be analyzed using broad categories such as positive and negative affect, it may be the case that all emotions within these broad categories are not uniformly affected by AIPs. Answering recent calls to move beyond studying emotional valence exclusively (Lerner & Keltner, 2000), we also examine the role of discrete emotions in AIP effectiveness. Specifically, we argue that it may be functionally adaptive to avoid attending to sadness stimuli (because attending to sadness stimuli makes one sad; Gross & Levenson, 1995) and therefore, sadness AIPs may exhibit weaker effect sizes compared with AIPs targeting other emotions. In contrast, anger, anxiety, and disgust stimuli may produce larger effect sizes because humans are evolutionarily predisposed to pay close attention to these stimuli (e.g., the fight-or-flight response arguably cannot exist without first detecting an anger/anxiety-provoking stimulus, and disgust similarly allows us to avoid life-threatening situations such as avoiding a contaminated environment or individuals with disease). Therefore, in addition to examining AIPs categorized as broadly inducing positive and negative affect, the present analysis also examines the effectiveness of AIPs designed to induce happiness, sadness, anxiety, anger, calm, and disgust (i.e., the most commonly induced emotional states).

For Whom Are AIPs Most Effective?

Gender

Gender is often inextricably linked to beliefs about emotionality. Women are often stereotyped as being more emotional than men (Barrett & Bliss-Moreau, 2009; Brody & Hall, 2008; Shields, 2002), suggesting that women may react more strongly to AIPs than men. However, results do not necessarily support the notion that women are more emotionally reactive than men (Barrett, Robin, Pietromonaco, & Eyssell, 1998; Grossman & Wood, 1993; Shields, 2002). In particular, recent meta-analytic evidence on gender differences in self-reports of shame, guilt, embarrassment, and pride concluded that "blanket stereotypes about women's greater emotionality are inaccurate" (Else-Quest, Higgins, Allison, & Morton, 2012, p. 947). Nevertheless, some empirical evidence still supports gender differences in the perception of and sensitivity to emotion, wherein women use more cues to detect emotional stimuli (i.e., both internal and external cues) than men (who are

less sensitive to external cues; Pennebaker & Roberts, 1992). Similarly, Westermann et al.'s (1996) review of elation/depression AIPs found evidence that women were more reactive to AIPs. Thus, we seek to meta-analytically test whether AIPs exhibit stronger effects in women than men.

Cross-Cultural Differences

Cross-cultural differences in the experience of emotions have been a prominent area of emotion research (e.g., Russell, 1983, 1991), and findings indicate a potential for systematic cross-cultural differences in AIP susceptibility. First, it appears that East Asian cultures experience fewer positive emotions than Americans (Kitayama, Markus, & Kurokawa, 2000), whereas negative emotions are more similarly experienced across cultures, suggesting that some cultural differences in emotion experiences after a positive AIP may emerge. Second, there are important cultural differences in expressiveness norms that relate to how one self-reports his or her emotional experience after an AIP. Namely, East Asian cultures tend to encourage less public expression of emotion than Western cultures (Bond, 1993), which may reduce the effectiveness of AIPs in East Asian cultures compared with Western cultures. Additionally, most self-report emotion assessments employ Likert-type scales, which have been found to have confounding effects across cultures due to the reference group effect wherein participants base their ratings on comparisons to a hypothetical reference group that differs across cultural contexts (Heine, Lehman, Peng, & Greenholtz, 2002). Given that cultural emotion standards vary widely, it is important to examine AIPs in a manner that is sensitive to these differences. In the present analysis we aim to test whether cultural differences impact AIP susceptibility. Studies from meaningfully different geographic regions will be examined separately, including North America, Europe, Australia/New Zealand, Central and South America, Middle East, and Asia.

Clinical Diagnosis

When AIPs were first introduced, some researchers (Goodwin & Williams, 1982) reviewed the possibility of using induced depression as a tool to study clinical depression. However, little is known about the effects of AIPs within clinical populations. Prior reviews have even excluded clinical patients from results to draw conclusions about nonclinical adults (Westermann et al., 1996). One could argue that clinical samples will be less sensitive to negative emotion AIPs because of a ceiling effect wherein they are already in a heightened negative emotional state (e.g., high in depression, anxiety) and may have less room to increase their negative emotions than their non-clinical counterparts. However, research on clinical populations suggests that these individuals attend to negative stimuli and/or have difficulty disengaging from negative stimuli more often than non-clinical counterparts (Joormann & Gotlib, 2007; MacLeod, Mathews, & Tata, 1986; Mathews & MacLeod, 2005; Olatunji & Wolitzky-Taylor, 2009; Pineles & Mineka, 2005). The present analysis, therefore, examines whether clinical populations are more susceptible to the influence of negative stimuli by examining whether negative emotion AIPs show greater effect sizes for negative emotions in clinical populations than in non-clinical populations. In addition, we also examine clinical and non-clinical differences in positive emotion AIPs to shed light on whether these populations react differently to positive stimuli as well.

Testing the Coupling and Decoupling of Positive and Negative Emotions: Do AIP Results Support the Bipolarity Hypothesis?

In this review, we use AIPs to test emotion theory in two ways. We first seek to provide a test of the bipolarity of positive and negative emotions. The notion that positive and negative emotion (i.e., happy vs. sad) are opposite ends of a bipolar continuum has been succinctly described by Russell and Carroll (1999), who state, "Bipolarity says that when you are happy, you are not sad and that when you are sad, you are not happy" (p. 25). Stated differently, the *bipolarity hypothesis* (Barrett & Russell, 2003) suggests that positive and negative affect are nonindependent, mutually exclusive, and exist on a bipolar, positive-negative continuum. The idea that positive and negative emotions lie on a continuum implies *coupled activation*: positive emotion AIPs should increase positive emotion and also decrease negative emotion, and negative emotion AIPs should increase negative emotion and also decrease positive emotion.

Alternatively, some emotion theorists suggest that positive and negative emotions are independent of each other (Cacioppo & Berntson, 1994), such that individuals can experience positive and negative emotions at the same time (Larsen & Mcgraw, 2011; Larsen, McGraw, & Cacioppo, 2001). Proponents of the independence of positive and negative emotions argue that positive and negative affect appear to have different physiological substrates that can have uncoupled activation (Cacioppo & Berntson, 1994). If positive and negative emotions are indeed independent and lie on two different continua, it implies *decoupled activation*: a positive AIP should result in an increase in positive emotion but a nonsignificant change in negative emotion, and a negative AIP should result in an increase in negative emotion and a nonsignificant change in positive emotion. Therefore, through our AIP review, we can examine whether AIPs generally lead to coupled or decoupled activation of positive and negative emotions, which can lend support to the bipolarity hypothesis or independence hypothesis, respectively.

Do Negative AIPs Overcome the Positivity Offset?

The positivity offset refers to the phenomenon in which individuals tend to be in positive moods in the absence of threatening stimuli (Cacioppo & Berntson, 1994, 1999), which may have evolved to promote behavior that increases chances of reproductive success during times when it is safe to do so (Diener, Kanazawa, Suh, & Oishi, 2015). Thus, prior to an AIP, the average participant should exhibit modest levels of positive emotions and low levels of negative emotions. This presents a problem for the effectiveness of negative AIPs that begs the question, "On average, are negative AIPs strong enough to overcome the positivity offset and induce true, negative affect?". Notably, prior reviews of the AIP literature have only examined change scores in emotion (e.g., Lench et al., 2011; Westermann et al., 1996). Although it is important to understand the extent to which negative AIPs *change* negative emotion, it is equally important to understand whether the

mean level of negative emotion after an induction indicates a negative emotional state. This is particularly important for negative AIPs given the positivity offset, because negative AIPs are tasked with changing the average individual's mood from positive to negative, whereas positive AIPs are tasked with changing the average individual's mood from positive to more positive.

Notably, the positivity offset and negativity bias can, and do, co-occur because the positivity offset refers to mean levels of emotion in the absence of stimuli and the negativity bias refers to what happens in the presence of negative versus positive stimuli. Thus, we ask whether the overattendance to negative stimuli that should occur with a negative AIP is enough to overcome the positivity offset. The answers to these questions are critical for the large body of research that relies on AIPs to study the effect of negative emotion on a construct or phenomenon; if negative AIPs do not induce a true negative emotion, then any empirical finding involving the effect of negative emotion on an outcome (for a full list, see Lench et al.'s 2011 meta-analysis of the effect of AIPs on outcomes) may be questioned because negative emotion may not have actually been induced in these studies.

Method

Literature Search

Studies were identified for inclusion by conducting a search in PsycINFO (1806–2018), PsycARTICLES (1894–2018), PubMed (1948–2018), and ProQuest Dissertation and Theses Global (1743–2018) for the keywords *mood induction, affect induction,* and *emotion induction*. The end date for the search was December 31, 2018. The search returned 3,871 studies (retracted papers were excluded [e.g., Avramova & Stapel, 2008]), which were then reviewed for the following inclusion criteria. First, the study had to be written in English. Second, the study had to involve an induction of affect (i.e., state positive affect, state negative affect, happiness, elation, cheerfulness, sadness, depression, anxiety, stress, worry, fear, anger, disgust, guilt, shame, calm). Inductions of cognitive states that were not explicitly affect-related (e.g., an induction of success that involved pre- and post-induction measures of feelings of success) were not included. Third, the study had to include a self-report measure of affect that was employed to assess the strength of the affect induction. Effect sizes were only included if the affect scale measured the same emotion or mood targeted by the AIP. For example, if a study induced happiness and measured disgust pre- and post-affect induction, it was excluded, but if a study induced happiness and measured happiness, it was included. The only exceptions to this involved the inclusion of all emotion scales for neutral AIP effect sizes (because neutral AIPs were never assessed with a self-report of neutral emotion) and the inclusion of discrete emotion AIPs that were evaluated with a broader positive/negative affect scale (e.g., a sadness AIP evaluated with a negative affect scale); however, these effect sizes were only included in the calculation of the broader positive/negative affect AIP meta-analytic effect sizes and were not included in meta-analyses of discrete emotion AIPs (e.g., the sadness AIP effect size only included AIPs that were evaluated with a sadness scale). The self-report measure was required to have been administered either (a) before and after the affect induction (i.e., a pre–post, repeated measures, within-person design), or (b) to a treatment and control group (i.e., an experimental, independent groups, between-persons design). Studies that involved both a treatment and control group as well as a pre–post administration of a self-report mood measure were also included. Fourth, studies were only included if they reported enough information to calculate an effect size (see Morris & DeShon, 2002). Finally, studies that only reported effect sizes involving participants for whom the AIP was successful were excluded. These inclusion rules reduced the database of studies to 529 studies, 874 independent samples, and a total sample size of 53,509. The mean proportion of females in the final set of samples was 65.30% and the mean age was 26.05. The entire search and inclusion process is displayed in Figure 1.

Data Coding

Studies that met the aforementioned inclusion criteria were then coded for several elements (all studies included in the meta-analysis and the coded information from each are provided in online supplemental material). One author coded each study first, and a second author checked these coding decisions. The coders agreed on 99% of the data coded (few, if any of the coding decisions involved a subjective judgment). First, each study was coded as either a repeated measures design (i.e., a single group received an AIP and responded to pre- and post-induction self-reports of affect), an independent groups design (i.e., an experimental group received an AIP and a control group received a neutral AIP and both groups responded to a post-induction self-report of affect), or an independent groups with repeated measures design (i.e., an experimental and control group responded to pre- and post-induction self-reports of affect). Next, the affective state being induced was coded for the specific type of emotion that was induced (e.g., happiness, sadness,[1] disgust), and for whether it was a positive affect induction (i.e., happiness, elation, cheerfulness, calm, or general positive affect) or a negative affect induction (i.e., sadness, depression, disgust, anxiety, anger, stress, fear, or general negative affect).

For the neutral AIP meta-analysis, any neutral induction that resulted in a negative effect size was reversed in sign so that when these effect sizes were combined into a meta-analytic effect size, the resulting *d* value indicated the absolute value of change in emotion, which addresses the general question of whether neutral AIPs are truly neutral. In addition, we also provide a meta-analytic effect size of neutral AIPs with the sign of the *d* values from primary studies coded to reflect the valence of the self-report emotion being assessed to determine whether neutral AIPs tend to induce more positive or negative emotions. For example, neutral AIPs that increased a self-report of positive emotion were coded as positive and neutral AIPs that increased a self-report of negative emotion were coded as negative.

In addition, studies were coded to indicate the type of affect induction procedure (e.g., imagination, film, music). Further information was coded to indicate whether the participants were told that a mood or emotion was being induced or whether the participants were deceived. The type of scale used to measure self-reported emotion or mood was also coded. If a study involved

[1] Elation AIPs were coded as happiness AIPs, and depression AIPs were coded as sadness AIPs.

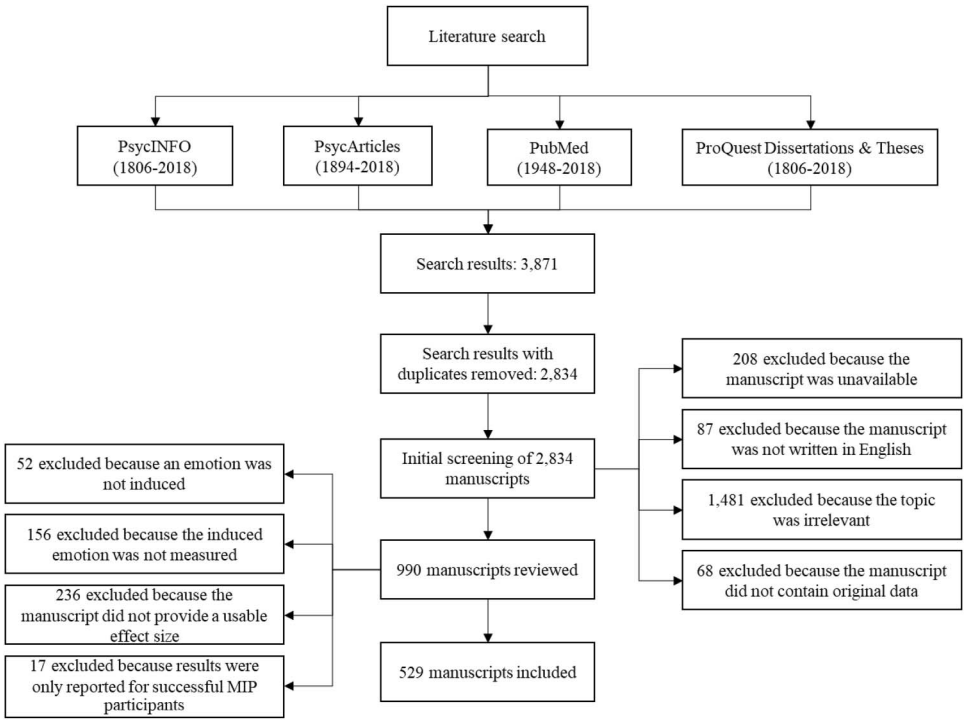

Figure 1. Flowchart of the search and selection procedure.

more than one self-report of emotion or mood (e.g., depression was induced and the study measured depression with two depression measures), the effect sizes for each scale were averaged. For studies that involved several different AIPs at multiple time points (e.g., a happiness AIP was followed by a separate disgust AIP), only the information from the first AIP was included to remove AIP effects that were likely to have been confounded by prior AIPs (studies in which the order of AIPs was counterbalanced were naturally excluded because an effect size for the first AIP was typically unavailable). To compare single and combined AIPs, we coded an effect size as combined if multiple AIPs were delivered to induce the same emotion and a single effect size was reported for the effects of both. Lastly, studies were also coded for several characteristics of the sample, including gender (if the entire sample was comprised of men or women or if the authors reported separate effects sizes for each gender), national origin (if the authors gave an explicit statement of the country in which the sample was collected), and clinical disorder (if the entire sample met criteria for a mental health diagnosis such as depression or the authors explicitly stated that the sample excluded those with mental health diagnoses).

Testing the Coupling and Decoupling of Positive and Negative Emotions

To test whether positive and negative emotions are coupled or decoupled, we examined whether positive (negative) emotion AIPs were effective in changing both positive and negative emotion. In addition to analyzing change in emotional states, we also sought to analyze mean levels of pre- and post-AIP positive and negative emotion to understand not only the direction of change, but also the magnitude of emotion these AIPs produce, on average. For these analyses, we only included studies using the most commonly used measure of self-reported positive and negative affect, the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988). Focusing on a single scale allows us to report mean levels of pre- and post-AIP emotion on that scale, which can inform us whether a positive emotion AIP, for example, not only increases positive emotion, but also whether the resulting emotional state represents low, medium, or high levels of positive emotion on the PANAS. For example, there may be positive change, but if this change only moves participants from a 2.0 to a 3.0 on the PANAS' 5-point Likert positive affect scale, one could argue that the AIP is not producing high levels of positive emotion. Arguably, our choice to focus on the PANAS is a very conservative test of the bipolarity hypothesis, given that the PANAS represents positive and negative affect as independent constructs rather than bipolar constructs (i.e., the positive affect items on the PANAS are not semantically bipolar to the negative affect items; Feldman-Barrett & Russell, 1999; Watson & Tellegen, 1985). Therefore, positive and negative affect measured with the PANAS are conceptually orthogonal within the affect space and should produce decoupled change.

Analyses

Because the current meta-analysis involved combing effect sizes from three types of primary studies (i.e., repeated measures, independent groups, and independent groups with repeated measures), we used the procedures described in Morris and DeShon (2002) to conduct a random effects meta-analysis (analyses were conducted in SAS Version 9.4). As described by Morris and DeShon, the

three types of effect sizes included in our database of primary studies must be transformed into a common metric to compute a single meta-analytic effect size. All original effect sizes were transformed into a repeated measures metric because the current meta-analysis was meant to investigate within-person change in emotion as a result of an affect induction. Standardized mean differences in repeated measures studies were calculated using Equation 4 from Morris and DeShon (2002). This equation requires knowledge of $r_{Pre.Post}$, or the correlation between pre- and post-affect induction scores. If a study did not directly report $r_{Pre.Post}$, the inverse sampling variance-weighted average $r_{Pre.Post}$ across all repeated measures studies was substituted into the equation. To obtain this value, a meta-analysis using Hedges and Olkin (1985) procedures was conducted, which produced a meta-analytic variance-weighted $r_{Pre.Post}$ of .61 ($k = 26$, $N = 971$). [To test the sensitivity of our results, analyses were run with an $r_{Pre.Post}$ of .40 and an $r_{Pre.Post}$ of .80 and the overall meta-analytic effect size changed no more than .001.]

Standardized mean differences in independent groups studies were calculated using Equation 2 from Morris and DeShon (2002), which were then transformed into the repeated measures metric via Equation 12 from Morris and DeShon (2002). Following recommendations by Morris (2008), effect sizes from independent groups with repeated measures studies were calculated in the repeated measures metric with Equation 8 from Morris (2008). When means and standard deviations were not available, this effect size was calculated by subtracting the repeated measures effect size of the control group from the repeated measures effect size of the experimental group (Morris & DeShon, 2002). Sampling variance for independent groups with repeated measures studies was calculated by summing the sampling variance for the experimental group and control group (i.e., Equation 25 from Morris, 2008).

Following Morris and DeShon (2002) procedures, all effect sizes were weighted by the reciprocal of the sampling variances (Hedges & Olkin, 1985), using the formulas provided by Morris and Deshon (and Morris [2008] for independent groups with repeated measures designs). Recent work has shown that weighting by the inverse of the sampling variance produces the most accurate estimates for meta-analytic *d* values (Brannick, Yang, & Cafri, 2011; Marín-Martínez & Sánchez-Meca, 2010) and is therefore the most appropriate weighting scheme for the current meta-analysis. Once the meta-analytic *d* values were calculated, corrected *d* values were estimated by applying Hunter and Schmidt's (2004) meta-analytic correction procedures. Specifically, because the reliability of the self-report emotion scale was not reported in all studies, an artifact distribution was used to correct for unreliability in self-reported emotion (the mean reliability of self-reported emotion was .92). Any meta-analytic effect size containing fewer than five independent samples was not listed in the results because most of these low-*k* results were based on extremely low sample sizes and were therefore quite unstable estimates.

Publication bias analyses were conducted on the overall meta-analytic effect size using Duval and Tweedie's (2000) trim-and-fill procedure, following Kepes, Banks, McDaniel, and Whetzel's (2012) recommendations. No publication bias was found (i.e., zero studies were imputed). In addition, Begg and Mazumdar's (1994) rank correlation test showed a nonsignificant rank order correlation between the effect size and the standard error ($\tau = .03$, $p > .05$) and Egger's intercept test showed a nonsignificant intercept when the effect size was regressed onto the inverse of the standard error ($\beta_0 = .10$, $p > .05$), also supporting a lack of publication bias.

Results

Are AIPs Effective and Which AIPs Are Most Effective?

Results of the current meta-analysis are presented in Tables 2–6. The overall corrected standardized mean difference across all AIPs was 1.32 ($k = 872$, $N = 53{,}509$), indicating that AIPs, on average, change an individual's affective state more than one standard deviation in the expected direction. Furthermore, the confidence interval of this effect size excluded zero (95% CI [1.20, 1.44]), indicating that the mean effect of AIPs is significantly different from zero. This overall meta-analytic effect is similar to previous meta-analytic estimates (current study: $d_{\text{uncorrected}} = 1.18$, $k = 874$; Westermann et al., 1996, p. 569: $d_{\text{uncorrected}} = 1.10$, $k = 250$), although it is based on more than three times the number of independent samples as previous meta-analytic work and is not restricted to a particular type of AIP. Perhaps unsurprisingly, there was substantial variance in this overall corrected effect size ($SD_{\delta} = 1.40$) and the lower 80% credibility interval (−.47) suggests that the least effective AIPs may induce emotion in the opposite direction of what was intended. Stated differently, AIPs appear to be effective, on average, but the extent to which they are effective varies greatly across AIPs, supporting the presence of moderators.[2]

AIP type. An examination of Table 2 showed that the most widely used AIP was film ($k = 182$), followed by autobiographical recall ($k = 137$), and the Velten procedure ($k = 77$), which contrasts with earlier work suggesting the Velten procedure is the most common AIP (Westermann et al., 1996). Upon comparing the effect sizes of the various AIP types, film administered with instructions exhibited the strongest effect size ($\delta = 2.62$), followed by pictures of facial expressions ($\delta = 2.06$), and reading with instruction ($\delta = 1.94$). Notably missing from this list of the strongest AIPs were several very popular AIPs including the Velten procedure ($\delta = 1.22$), autobiographical recall ($\delta = 1.36$), and film without instruction ($\delta = 1.40$). These results suggest the popularity of these particular AIPs are perhaps unwarranted, as other AIPs appear to produce much stronger induction effects.

Results suggest the least effective AIPs were jokes/cartoons ($\delta = .33$) and video recording/social pressure ($\delta = .42$). How-

[2] To examine whether study quality was one of the sources of variability in AIP effectiveness, each independent sample was coded for all 12 of Chacón-Moscoso, Sanduvete-Chaves, and Sánchez-Martín's (2016) indicators of study quality. AIP effect sizes were then regressed onto each of the indicators in twelve separate regressions (following Chacón-Moscoso et al.'s recommendation to avoid creating a summed score of quality) using Wilson's meta-analytic regression macro for SPSS (Lipsey & Wilson, 2001; Wilson, 2019). Results indicate 10 of the 12 indicators were not significantly related to effectiveness, and the two indicators that were significantly related (Quality Indicator #7 and #8) only explained 1.78% of the variance in effectiveness, suggesting quality was not a meaningful moderator of effectiveness.

Table 2
Meta-Analytic Results for Overall AIP and Research Design Moderators

Measure	*k*	*N*	*d*	δ	SD_δ	% Var	95% CI-L	95% CI-U	80% CR-L
All AIPs	874	53,509	1.18	1.32	1.40	66.55	1.20	1.44	−.47
Instructions									
Truthful AIP	686	39,914	1.25	1.41	1.51	65.98	1.26	1.57	−.51
Deceptive AIP	185	13,251	.89	.99	.84	100.00	.81	1.16	−.09
Type of AIP									
Single AIP technique	626	40,145	1.16	1.30	1.49	54.87	1.17	1.44	−.61
Film	182	13,367	1.38	1.57	1.78	52.99	1.30	1.84	−.70
Film without instruction	153	10,800	1.24	1.40	1.47	69.44	1.13	1.66	−.48
Film with instruction	29	2,567	2.15	2.62	2.84	31.15	1.48	3.77	−1.01
Autobiographical recall	137	8,479	1.21	1.36	1.34	75.65	1.07	1.66	−.35
Autobiographical recall without writing	88	5,102	1.13	1.27	1.16	77.75	.93	1.61	−.22
Autobiographical recall with writing	49	3,377	1.34	1.52	1.60	75.32	.88	2.16	−.53
Velten	77	4,503	1.09	1.22	.86	100.00	.88	1.55	.11
Music/sounds	54	2,846	.74	.82	.87	100.00	.43	1.21	−.29
Music without instruction	42	1,785	.75	.83	.88	100.00	.35	1.30	−.30
Music with instruction	21	1,013	.74	.82	.85	100.00	.06	1.58	−.28
Pictures	45	1,993	1.44	1.64	2.43	31.75	1.09	2.19	−1.47
International Affective Picture System	22	1,313	.73	.80	.85	93.16	.39	1.20	−.28
Pictures of facial expressions	14	267	1.75	2.06	1.23	100.00	1.23	2.89	.47
Imagination	32	2,302	.98	1.09	.90	100.00	.70	1.47	−.07
Feedback	23	1,322	.83	.92	.61	100.00	.50	1.34	.15
Coping challenge	17	1,099	.68	.74	.42	100.00	.26	1.23	.21
Reading a story	16	1,794	1.56	1.80	2.25	15.02	.61	2.98	−1.08
Reading without instruction	8	764	1.45	1.65	1.19	40.72	.48	2.83	.13
Reading with instruction	8	1,030	1.67	1.94	2.93	10.34	−.19	4.08	−1.81
Manipulation of face/body	11	268	.66	.73	.87	100.00	.09	1.36	−.38
Jokes/cartoons	8	624	.31	.33	.41	70.41	.01	.65	−.19
Odor	5	86	1.24	1.40	.26	100.00	.94	1.86	1.06
Video recording/social pressure	5	183	.39	.42	.18	100.00	.15	.70	.19
Multiple AIP techniques	248	13,364	1.20	1.34	1.10	100.00	1.03	1.65	−.07
Music & autobiographical recall	79	3,781	1.43	1.64	.94	100.00	1.06	2.21	.43
Music & Velten	31	1,233	1.15	1.29	.95	100.00	.67	1.91	.07
Music & pictures	21	1,613	.94	1.04	.99	100.00	.32	1.75	−.22
Music & imagination	24	1,639	.89	.99	.70	100.00	.50	1.47	.09

Note. AIP = affect induction procedures; *k* = number of effect sizes; *N* = sample size; *d* = inverse sampling variance-weighted mean standardized difference in a repeated measures metric; δ = *d* value corrected for measurement error in a repeated measures metric; SD_δ = corrected standard deviation of *d* value; % Var = percent of variance accounted for by sampling error; 95% CI-L = lower bound of 95% confidence interval around δ; 95% CI-U = upper bound of 95% confidence interval around δ; 80% CR-L = lower bound of 80% credibility interval around δ. Effects from neutral AIPs were not included in any meta-analytic effect size in this table.

ever, it should be noted that the confidence intervals for these weaker effect sizes still excluded zero, suggesting they were effective at inducing their intended emotion on average, albeit to a lesser extent than alternate AIP strategies. Some caution is warranted in interpreting these estimates, however, as they are both based on fewer than 10 independent samples and may be somewhat unstable, requiring reevaluation following additional research. In sum, the results of AIP type as a moderator indicated that the choice of AIP matters for the effectiveness of an affect induction, as the estimated effect size for the most effective AIP (film with instruction; δ = 2.62) was approximately eight times as large as the least effective AIP (jokes/cartoons; δ = .33).

Positive and negative AIPs. Results comparing various AIP types for positive and negative affect, separately, are shown in Table 3. Given that negative AIPs were substantially more frequent (*k* = 665) in the current meta-analysis than positive AIPs (*k* = 346), it may come as no surprise that the results of negative AIPs alone are similar to those of the full sample of studies (i.e., the overall effect sizes were largely driven by the negative AIP results). That is, film with instruction exhibited the largest effect size for negative AIPs (δ = 2.89), followed by pictures of facial expressions (δ = 2.59). Upon comparing these two strongest AIP types for negative affect, we note that film with instruction exhibited a substantial amount of heterogeneity in effect sizes (i.e., the credibility interval was wide), whereas pictures of facial expression exhibited less heterogeneity, suggesting that pictures of facial expression may be a better choice because it is both effective, on average, and displays a narrower range of effectiveness than film with instruction.

In comparison, results for positive AIPs indicated the most effective positive AIP was reading a story (δ = 2.08) followed by pictures of facial expressions (δ = 1.53). However, both of these effect sizes were based on a small number of primary studies and fewer than 600 participants, suggesting that additional research is needed to confirm the strength of these AIPs. Given that the most effective positive AIPs were based on a relatively small amount of data, the best choice for positive induction may be to use a film. Film was only the fifth most-effective positive AIP in our meta-

Table 3
Meta-Analytic Results for Affect Induced and Type of AIP as a Moderator

Measure	*k*	*N*	*d*	δ	SD_δ	% Var	95% CI-L	95% CI-U	80% CR-L
Positive affect	346	19,994	.82	.91	.88	80.20	.78	1.04	−.22
Film	79	5,245	1.00	1.11	.70	100.00	.83	1.40	.22
Film without instruction	71	4,812	.98	1.09	.63	100.00	.79	1.39	.28
Film with instruction	7	363	.84	.93	.64	100.00	.13	1.74	.11
Autobiographical recall	50	2,882	.95	1.06	1.27	73.43	.67	1.45	−.56
Autobiographical recall without writing	36	2,078	.82	.91	.53	100.00	.55	1.26	.22
Autobiographical recall with writing	14	804	1.28	1.44	2.18	57.26	.35	2.53	−1.35
Velten	27	1,502	.80	.88	1.17	50.94	.32	1.45	−.61
Music/sounds	34	1,430	.55	.60	.57	100.00	.33	.87	−.13
Music without instruction	25	969	.37	.41	.43	100.00	.19	.62	−.14
Music with instruction	9	461	1.06	1.18	.54	100.00	.07	2.29	.49
Pictures	26	893	.75	.83	1.12	23.55	.40	1.26	−.60
Pictures of facial expressions	14	254	1.35	1.53	1.13	100.00	.74	2.33	.09
International Affective Picture System	9	481	.17	.18	.54	34.17	−.17	.54	−.50
Imagination	12	688	.43	.47	.63	79.31	.04	.90	−.34
Reading a story	6	555	1.77	2.08	1.43	18.00	.31	3.85	.25
Feedback	10	463	.87	.96	.54	100.00	.29	1.63	.27
Jokes/cartoons	8	624	.31	.33	.41	70.41	.01	.65	−.19
Music & autobiographical recall	16	706	.79	.87	.48	100.00	.30	1.43	.25
Music & imagination	13	923	.63	.69	.44	100.00	.23	1.15	.13
Music & pictures	11	909	.50	.54	.49	100.00	.11	.98	−.08
Music & Velten	5	122	.97	1.08	.23	100.00	.61	1.55	.79
Negative affect	665	38,259	1.40	1.59	1.55	73.80	1.42	1.76	−.40
Film	135	9,157	1.61	1.86	2.01	53.44	1.50	2.23	−.71
Film without instruction	107	6,732	1.43	1.63	1.70	66.83	1.27	2.00	−.54
Film with instruction	27	2,355	2.31	2.89	2.88	32.30	1.61	4.16	−.80
Autobiographical recall	105	6,538	1.41	1.61	1.56	70.92	1.20	2.01	−.39
Autobiographical recall without writing	66	3,814	1.34	1.52	1.36	72.54	1.02	2.02	−.22
Autobiographical recall with writing	39	2,724	1.50	1.73	1.85	69.71	.94	2.51	−.64
Velten	68	3,545	1.20	1.36	.77	100.00	.99	1.72	.38
Music	43	1,783	.88	.97	.99	94.28	.42	1.51	−.30
Music without instruction	25	998	.99	1.10	1.00	98.90	.34	1.86	−.18
Music with instruction	17	737	.72	.80	.97	94.41	−.09	1.68	−.44
Pictures	38	1,517	2.01	2.41	2.63	54.39	1.55	3.27	−.96
Pictures of facial expressions	14	254	2.12	2.59	1.63	100.00	1.63	3.54	.49
International Affective Picture System	18	1,008	1.18	1.33	1.15	100.00	.46	2.20	−.14
Imagination	25	1,816	1.17	1.31	.97	100.00	.84	1.78	.07
Coping challenge	17	1,099	.68	.74	.42	100.00	.26	1.23	.21
Reading a story	13	1,345	2.04	2.47	1.81	37.59	1.06	3.87	.14
Feedback	17	905	.84	.92	.64	100.00	.45	1.40	.11
Manipulation of face/body	6	123	.98	1.09	.82	100.00	.75	1.42	.03
Odor	5	86	1.24	1.40	.26	100.00	.94	1.86	1.06
Video recording/social pressure	5	183	.39	.43	.18	100.00	.15	.70	.19
Music & autobiographical recall	64	3,075	1.60	1.86	.97	100.00	1.06	2.65	.62
Music & Velten	30	1,159	1.16	1.30	.97	100.00	.67	1.94	.06
Music & imagination	11	683	1.37	1.56	.78	100.00	.06	3.06	.56
Music & pictures	14	959	1.20	1.34	1.01	100.00	.05	2.65	.06
Neutral	170	6,048	.33	.36	.36	100.00	.29	.43	−.10
Film	35	1,287	.31	.34	.28	100.00	.20	.47	−.02
Autobiographical recall	16	819	.22	.23	.18	100.00	.13	.34	.01
Music	12	339	.38	.41	.41	100.00	.12	.70	−.12
Velten	16	571	.47	.52	.76	33.95	.15	.89	−.46
Feedback	6	113	.24	.26	.37	100.00	−.04	.56	−.21
Imagination	7	404	.29	.32	.41	62.50	−.03	.68	−.21
Reading a story	11	531	.40	.44	.38	100.00	.13	.75	−.05
Music & autobiographical recall	9	218	.27	.30	.21	100.00	.10	.50	.03
Music & Velten	11	279	.30	.32	.21	100.00	.13	.52	.06

Note. AIP = affect induction procedures; *k* = number of effect sizes; *N* = sample size; *d* = inverse sampling variance-weighted mean standardized difference in a repeated measures metric; δ = *d* value corrected for measurement error in a repeated measures metric; SD_δ = corrected standard deviation of *d* value; % Var = percent of variance accounted for by sampling error; 95% CI-L = lower bound of 95% confidence interval around δ; 95% CI-U = upper bound of 95% confidence interval around δ; 80% CR-L = lower bound of 80% credibility interval around δ. All effect sizes from neutral AIPs were positively scored. Positive moods included happiness, elation, cheerfulness, and calm. Negative moods included sadness, depression, disgust, anxiety/stress, anger, guilt/shame, and fear.

Table 4
Meta-Analytic Results for Discrete Emotion Induced and Type of AIP as a Moderator

Measure	*k*	*N*	*d*	δ	SD_{δ}	% Var	95% CI-L	95% CI-U	80% CR-L
Happiness	84	3,991	1.02	1.13	1.18	72.98	.82	1.45	−.37
Film	13	815	.79	.87	.49	100.00	.38	1.36	.24
Autobiographical recall	18	934	1.36	1.54	2.02	81.61	.57	2.52	−1.04
Velten	10	792	.64	.70	.44	100.00	.03	1.37	.14
Music	12	348	.57	.62	.49	100.00	.12	1.13	.00
Pictures	9	178	1.84	2.18	.99	100.00	1.29	3.07	.91
Sadness	223	12,653	1.71	2.00	1.66	84.21	1.70	2.30	−.12
Film	63	4,285	1.74	2.04	2.18	48.79	1.44	2.64	−.75
Autobiographical recall	20	1,479	2.27	2.81	2.18	77.02	1.57	4.05	.02
Velten	43	2,310	1.31	1.49	.86	100.00	1.00	1.98	.39
Music	12	469	1.55	1.79	1.24	100.00	.01	3.56	.20
Pictures	10	208	2.98	4.20	1.57	100.00	2.15	6.25	2.19
Music & autobiographical recall	35	1,756	1.70	1.99	.99	100.00	.92	3.06	.72
Music & Velten	14	451	.95	1.05	.98	96.17	.35	1.75	−.21
Anxiety	72	4,369	.96	1.07	.83	100.00	.77	1.37	.01
Film	8	579	1.35	1.53	.76	100.00	−.34	3.40	.56
Autobiographical recall	9	747	1.50	1.72	1.01	100.00	.66	2.78	.42
Coping challenge	9	790	.62	.68	.45	100.00	.05	1.32	.11
Video recording/social pressure	5	183	.39	.43	.18	100.00	.15	.70	.19
Music & autobiographical recall	6	129	1.58	1.83	.86	100.00	−.30	3.95	.73
Anger	41	2,757	1.54	1.78	2.31	48.41	.86	2.69	−1.19
Autobiographical recall	19	1,688	2.07	2.50	3.25	33.93	.97	1.03	−1.66
Velten	6	395	1.13	1.27	.46	100.00	.58	1.95	.68
Disgust	16	734	4.16	9.75	3.49	100.00	5.95	13.54	5.27
Calm	7	225	.71	.78	.31	100.00	.29	1.27	.38

Note. AIP = affect induction procedures; *k* = number of effect sizes; *N* = sample size; *d* = inverse sampling variance-weighted mean standardized difference in a repeated measures metric; δ = *d* value corrected for measurement error in a repeated measures metric; SD_{δ} = corrected standard deviation of *d* value; % Var = percent of variance accounted for by sampling error; 95% CI-L = lower bound of 95% confidence interval around δ; 95% CI-U = upper bound of 95% confidence interval around δ; 80% CR-L = lower bound of 80% credibility interval around δ. Happiness effect sizes included happiness and elation AIPs. Sadness effect sizes included sadness and depression AIPs. Anxiety effect sizes included anxiety, stress, and worry AIPs.

analysis (δ = 1.11), but evidence of its effectiveness was based on the largest amount of data (*k* = 79) and exhibited less heterogeneity than some of the more effective positive AIPs.

Finally, an examination of neutral AIPs showed an unexpected finding in that neutral AIPs do not appear to be neutral. That is, analyses utilizing the absolute value of emotional change in any direction indicated neutral AIPs induced nonzero emotional change (δ = .36, 95% CI [.29, .43]). Results suggest that the extent to which a neutral AIP was neutral varied somewhat across the types of AIPs, ranging from δ = .23 (autobiographical recall) to δ = .52 (Velten). Only two types of neutral AIPs were indeed neutral, as evidenced by effect sizes in which the confidence interval included zero (indicating a nonsignificant emotion induction, as intended): feedback (δ = .26) and imagination (δ = .32), but these wide confidence intervals may be due to low power for these effect sizes. Thus, it appears that neutral AIPs are producing unintended emotional change, on average.

To examine the direction of this nonzero change found in the absolute value analyses, the meta-analytic effect size of neutral AIPs was recalculated such that a positive effect size represents an increase in positive affect and a negative effect size represents an increase in negative affect. This recalculation of the effectiveness of neutral AIPs indicated that neutral AIPs are slightly negative (δ = −.08, 95% CI [−.15, −.03]), again suggesting that neutral AIPs are not neutral, and instead seem to be producing mild negative emotions, on average.

What Enhances the Effectiveness of AIPs?

Instruction. Results comparing AIPs delivered with instruction versus those delivered without instruction indicated mixed results (see Table 2): although film delivered with instruction (δ = 2.62) was significantly more effective than film delivered without instruction (δ = 1.40; $t = 3.27$, $p < .05$), music delivered with instruction (δ = .82) was not significantly more effective than music delivered without instruction (δ = .83; $t = .04$, $p > .05$), nor was reading with instruction (δ = 1.94) more effective than reading without instruction (δ = 1.65; $t = .28$, $p > .05$). Results comparing AIPs delivered with honest instructions regarding the intent of the affect induction and AIPs delivered with deception (see Table 2) indicated that truthful AIPs were significantly more effective (δ = 1.41) than deceptive AIPs (δ = .99; $t = 3.85$, $p < .05$). This echoes the findings of previous meta-analytic work indicating that honest instructions result in a more effective AIP (Westermann et al., 1996), although we note that honest and deceptive AIPs exhibited substantial variability, suggesting the presence of moderators.

Single versus multiple AIPs. Upon comparing the results of single AIPs and affect inductions involving multiple procedures (see Table 2), the current meta-analysis counterintuitively suggests multiple AIPs do not offer an advantage over single AIPs: the corrected effect size of AIPs involving multiple procedures (δ = 1.34) was not significantly larger than the effect size for single AIPs (δ = 1.30; $t = .47$, $p > .05$). Similarly, when examining the

Table 5
Meta-Analytic Results for Gender, Country, and Clinical Diagnosis of the Sample as a Moderator

Variable	*k*	*N*	*d*	δ	SD_δ	% Var	95% CI-L	95% CI-U	80% CR-L
Women									
All AIPs	115	6,755	1.57	1.80	1.73	50.23	1.43	2.18	−.41
Positive affect	25	972	1.04	1.16	1.22	77.15	.47	1.86	−.40
Negative affect	104	6,112	1.67	1.94	1.77	50.57	1.54	2.34	−.33
Men									
All AIPs	37	1,347	.97	1.09	.89	100.00	.60	1.58	−.05
Positive affect	17	544	.72	.79	.77	100.00	.24	1.35	−.19
Negative affect	29	996	1.20	1.35	1.14	100.00	.68	2.02	−.11
Geographic region of the sample									
North America									
All AIPs	423	28,098	1.25	1.41	1.58	53.90	1.23	1.59	−.60
Positive affect	158	10,432	.90	1.00	1.06	69.53	.78	1.21	−.36
Negative affect	327	19,701	1.46	1.67	1.72	60.51	1.41	1.92	−.53
Europe									
All AIPs	287	14,238	1.13	1.27	1.39	69.45	1.03	1.50	−.51
Positive affect	131	5,921	.75	.83	.71	98.36	.66	1.00	−.08
Negative affect	208	9,653	1.39	1.58	1.61	79.13	1.22	1.93	−.49
Australia/New Zealand									
All AIPs	37	2,539	1.16	1.30	.71	100.00	.79	1.81	.39
Positive affect	18	1,272	1.04	1.16	.58	100.00	.39	1.93	.42
Negative affect	27	1,504	1.31	1.48	.79	100.00	.85	2.11	.47
Asia									
All AIPs	18	1,513	1.11	1.25	1.01	100.00	.41	2.08	−.05
Positive affect	9	570	.63	.69	.95	44.14	.01	1.36	−.53
Negative affect	12	1,018	1.64	1.90	1.02	100.00	.18	3.63	.59
Middle East									
All AIPs	13	663	1.19	1.33	1.02	100.00	.71	1.96	.02
Positive affect	5	367	.73	.80	.13	100.00	.51	1.10	.64
Negative affect	10	522	1.68	1.95	1.27	100.00	1.47	2.44	.32
Central/South America									
All AIPs	6	200	.69	.76	.60	100.00	−.09	1.61	−.01
Clinical diagnoses of the sample									
Participants without a diagnosed clinical disorder	106	4,193	1.39	1.59	1.73	69.56	1.20	1.98	−.62
Positive affect	33	1,186	.91	1.01	1.32	81.66	.54	1.48	−.68
Negative affect	82	3,319	1.59	1.84	1.94	71.58	1.31	2.37	−.64
Participants diagnosed with a clinical disorder	52	2,012	1.00	1.11	1.46	46.39	.66	1.55	−.77
Positive affect	14	457	.73	.81	.55	100.00	.10	1.51	.10
Negative affect	39	1,541	1.09	1.22	1.65	42.46	.67	1.77	−.89

Note. AIP = affect induction procedures; *k* = number of effect sizes; *N* = sample size; *d* = inverse sampling variance-weighted mean standardized difference in a repeated measures metric; δ = *d* value corrected for measurement error in a repeated measures metric; SD_δ = corrected standard deviation of *d* value; % Var = percent of variance accounted for by sampling error; 95% CI-L = lower bound of 95% confidence interval around δ; 95% CI-U = upper bound of 95% confidence interval around δ; 80% CR-L = lower bound of 80% credibility interval around δ. Effects from neutral AIPs were not included in any effect sizes in this table.

effectiveness of single versus multiple AIPs for positive AIPs alone (see Table 3), it does not appear that multiple AIPs are more effective than single AIPs for producing positive emotions (i.e., the strongest effect sizes for positive AIPs were from single AIPs). The same was also true for negative AIPs, where the most effective AIPs were single AIPs.

Which Emotions Are Easiest to Elicit?

Positive versus negative affect. Results comparing positive and negative emotion AIPs (see Table 3) are consistent with a negativity bias in that the effect size for negative emotion AIPs (δ = 1.59) was significantly stronger than that for positive emotion AIPs (δ = .91; $t = 9.08$, $p < .05$).

Discrete emotions. A comparison of happiness, sadness, anxiety, anger, and disgust AIPs (see Table 4) revealed that disgust AIPs (δ = 9.75) exhibited significantly stronger effect sizes than all other discrete emotions (happiness: δ = 1.13, $t = 11.43$, $p < .05$; sadness: δ = 2.00; $t = 13.64$, $p < .05$; anxiety: δ = 1.07; $t = 15.06$, $p < .05$; anger: δ = 1.78; $t = 4.18$, $p < .05$; calm: δ = .78; $t = 12.97$, $p < .05$), suggesting disgust is substantially easier to induce than other emotions. When combined with our results from Table 3, it appears that although negative inductions are more effective than positive inductions, on average, not all negative inductions are created equal, wherein disgust is substantially easier to induce than other negative emotions. Upon examining positive inductions, happiness inductions were significantly stronger than calm inductions, ($t = 2.02$, $p < .05$). Beyond comparing the effect sizes of AIPs for various discrete emotions, we also provide Table 4, which presents effect sizes for the effectiveness of various AIP types for each discrete emotion to be used by researchers in guiding their choice of AIP.

Table 6
Meta-Analytic Results of Pre- and Post-AIP PANAS Scores

Type of AIP and measure	*k*	*N*	Sample-weighted mean PANAS score	*d*	δ	SD_δ	% Var	95% CI-L	95% CI-U	80% CR-L
Positive affect induced										
Positive Affect Scale	41	1,251		.53	.58	.61	90.89	.33	.83	−.20
Pre-AIP	41	1,251	2.72							
Post-AIP	41	1,251	3.01							
Negative Affect Scale	29	665		−.32	−.35	.38	100.00	−.53	−.18	−.84
Pre-AIP	29	665	1.32							
Post-AIP	29	665	1.22							
Negative affect induced										
Positive Affect Scale	64	2,296		−.66	−.73	.75	70.17	−.99	−.47	−1.69
Pre-AIP	64	2,296	2.73							
Post-AIP	64	2,296	2.37							
Negative Affect Scale	84	4,444		.76	.84	.54	100.00	.62	1.05	.15
Pre-AIP	84	4,444	1.46							
Post-AIP	84	4,444	1.91							

Note. AIP = affect induction procedures; PANAS = Positive and Negative Affect Schedule; *k* = number of effect sizes; *N* = sample size; *d* = inverse sampling variance-weighted mean standardized difference in a repeated measures metric; δ = *d* value corrected for measurement error in a repeated measures metric; SD_δ = corrected standard deviation of *d* value; % Var = percent of variance accounted for by sampling error; 95% CI-L = lower bound of 95% confidence interval around δ; 95% CI-U = upper bound of 95% confidence interval around δ; 80% CR-L = lower bound of 80% credibility interval around δ. Only studies using the PANAS (Response Scale: 1–5) were included.

For Whom Are AIPs Most Effective?

Gender. An examination of the results comparing AIPs in women and men (see Table 5) suggests AIPs exhibit significantly stronger inductions in women (δ = 1.80) than men (δ = 1.09; $t = 3.20$, $p < .05$). When the effect sizes for men and women were separated into positive AIPs (women δ = 1.16; men δ = .79) and negative AIPs (women δ = 1.94; men δ = 1.35), samples of women only showed significantly larger effects than samples of men for negative AIPs (positive AIPs $t = 1.23$, $p > .05$; negative AIPs $t = 2.00$, $p < .05$). Notably, however, there existed a substantial amount of heterogeneity within these effect sizes, indicating the presence of moderators that suggest AIPs administered to women will not always be more effective than those administered to men.

Cross-cultural differences. Effect sizes representing AIP effectiveness across geographic region are displayed in Table 5. AIPs exhibited the strongest effects in North American samples (δ = 1.41), followed by Middle Eastern samples (δ = 1.33), Australian/New Zealand samples (δ = 1.30), European samples (δ = 1.27), Asian samples (δ = 1.25), and finally, Central/South American samples (δ = .76). Because the Central/South American effect size was based on a small amount of primary data, we do not interpret it further here and call for additional research in this region. Upon comparing the remaining geographic regions, there were no significant differences in AIP effectiveness across regional comparisons. When comparing positive AIPs and negative AIPs within each geographic region, all regions showed a significant negativity bias (i.e., negative AIPs were stronger than positive AIPs; $t_{North\ America} = 5.22$, $p < .05$; $t_{Europe} = 6.50$, $p < .05$; $t_{Asia} = 2.50$, $p < .05$; $t_{Middle\ East} = 4.27$, $p < .05$) except in Australia, where negative and positive AIPs were not significantly different ($t = 1.57$, $p > .05$).

Clinical diagnosis. Moderator analyses of AIPs in nonclinical samples compared with AIPs administered in samples in which participants had been diagnosed with a clinical disorder (see Table 5) revealed AIPs in non-clinical samples (δ = 1.59) were not significantly more effective than in clinical samples (δ = 1.11; $t = 1.74$, $p > .05$). Similarly, neither positive AIPs, ($t = .75$, $p > .05$) nor negative AIPs, ($t = 1.69$, $p > .05$), differed in effectiveness across clinical diagnosis. To evaluate whether a ceiling effect may have stunted the effectiveness of AIPs in clinical samples, we examined the pre-AIP mean levels of negative affect in clinical samples. Because we were interested in means, we restricted this analyses to a single scale—a visual analog scale (0–100 response scale) of sadness, which is the only scale for which we had enough data on clinical samples. Also, because this is a sadness scale, if there is a ceiling effect, one would expect clinically depressed samples to exhibit pre-AIP sadness levels that are close to the maximum range of the scale. This was not the case—the pre-AIP mean for depressed samples was 32.65 ($k = 5$), which is far below the maximum value of 100, suggesting no ceiling effect. In contrast, the pre-AIP mean for non-clinical samples was 21.27 ($k = 9$), which is substantially lower than the mean for depressed samples. Thus, initial evidence suggests that although clinical samples are not more emotionally reactive than non-clinical samples, they enter into AIPs with more negative affect, resulting in more negative end-states (mean clinical post-AIP sadness = 51.88; mean nonclinical post-AIP sadness = 42.50).

Do AIP Results Support the Coupling of Positive and Negative Emotions?

Table 6 presents an analysis of AIPs that have been evaluated with the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) to test the bipolarity hypothesis. The results indicate that positive emotion AIPs increase positive emotion significantly more than zero (δ = .58, 95% CI [.33, .83]) and also decrease negative emotion significantly more than zero (δ = −.35, 95% CI [−.53, −.18]). However, these positive emotion AIPs induced positive emotions to a significantly greater degree than they decreased negative emotions, ($t = 1.96$, $p < .05$). Similarly, negative

emotion AIPs increased negative emotion significantly more than zero (δ = .84, 95% CI [.62, 1.05]) and decreased positive emotion significantly more than zero (δ = −.73, 95% CI [−.99, −.47]). Negative AIPs did not induce negative emotions significantly more than they decreased positive emotions (t = .98, $p > .05$). These results provide evidence of a coupling of positive and negative emotions which points toward support of the bipolarity hypothesis.

Do Negative AIPs Overcome the Positivity Offset?

Table 6 presents the sample-weighted mean scores on the PANAS (Watson et al., 1988) from participants before and after an AIP. As shown in Table 6, the mean pre-induction positive affect scores for positive and negative AIPs were 2.72 and 2.73, respectively, on a 1–5 Likert scale. In contrast, mean levels of negative affect prior to an AIP were 1.32 and 1.46 for positive and negative AIPs, respectively. These results are consistent with a positivity offset in that prior to any stimuli being administered, participants appeared to be in a mildly positive state (the level of positive affect compared with negative affect exhibited an approximate 2:1 ratio on average). Notably, after a negative AIP, levels of negative affect still remained quite low—a 1.91 on a 5-point scale, on average—suggesting that negative AIPs are not successfully inducing a negative affective state (i.e., they do not induce a negative state beyond the midpoint of a negative affect scale). Instead, they appear to be decreasing one's positive affect as much as they are inducing mild negative affect. In contrast, positive AIPs do appear to induce a positive affective state (i.e., positive affect after a positive AIP is on average a 3.01 on a 5-point scale). Thus, it appears that although positive AIPs are successful in inducing positive emotional states, negative emotion AIPs are not successful in overcoming the positivity offset and only induce mildly negative emotional states.

Discussion

Within psychology and beyond, AIPs have become essential tools for conducting experimental research to understand emotions and their effects across a variety of contexts. The goal of the current systematic review was, first, to take stock of AIP research and evaluate the effectiveness of these procedures as a research tool, and second, to use meta-analytic data to test theories of emotion (e.g., the bipolarity hypothesis, negativity bias, positivity offset, and theories of emotionality and gender). More specifically, we sought to address whether AIPs are effective for changing intended emotions, what conditions maximize their effectiveness, for which emotions they are most effective, for whom they are most effective, and whether aggregated affect induction findings provide insight into theories of emotion.

Are AIPs Effective and Which AIPs Are Most Effective?

Our results reveal that AIPs are generally effective at inducing affect roughly one standard deviation in the expected direction. However, we note that the most effective AIP (film with instruction) was approximately eight times more effective than the least effective AIP (jokes/cartoons), and of the 874 effect sizes analyzed, 45 were negative, suggesting that the least effective AIPs change emotion in the *opposite direction* of what was intended. Thus, we conclude that although the average AIP is effective, given the variability of effect sizes in our meta-analysis, not all AIPs are created equal, and care should be taken when choosing an AIP.

How should one choose an AIP? To answer this question, we begin by noting that the most effective stand-alone AIP was film with instruction. Interestingly, this AIP was not particularly popular (k = 29), leading us to highlight this AIP as one that appears to warrant additional use in future work. In contrast, the most popularly employed AIPs (film without instruction, autobiographical recall, and the Velten procedure) were not always the most effective, suggesting that these commonly used AIPs should only be used when other methods are not appropriate. For inducing positive and negative affect specifically, our results suggest two AIPs were the most effective: film and pictures of facial expressions. Although both performed well, we recommend pictures of facial expressions for inducing negative affect, and film for inducing positive affect, given the strength, stability, and lower relative heterogeneity of these meta-analytic effect sizes.

In addition, we also identified the most effective AIPs for the discrete emotions of happiness, sadness, anxiety, and anger (see Table 4). Results suggest that when one is interested in inducing happiness or sadness, pictures (e.g., pictures of facial expressions) should be used, whereas autobiographical recall should be used for inducing anxiety or anger. We propose this is consistent with existing work on empathic processes showing that pictures of happiness and sadness activate mirror systems in the brain, causing one to feel happiness and sadness, respectively (Chakrabarti, Bullmore, & Baron-Cohen, 2006). This contrasts with anxiety, which exhibits a reduced empathic response to pictures compared with happiness and sadness (Morelli & Lieberman, 2013), and anger, which shows activation in areas of the brain that aid in evaluating a threat (rather than only activating areas of the brain associated with a mirrored response; Chakrabarti et al., 2006). This suggests that while pictures of *others* displaying happiness and sadness may activate an empathic, mirrored response that translates into higher self-reports of happiness/sadness, the stimuli for anxiety and anger need to be *personal* (e.g., autobiographical recall) to induce the intended emotion.

Altogether, the results of these analyses suggest that (a) the average AIP is effective, (b) although most AIPs were effective, some were more effective than others, and (c) AIP effectiveness varies with the emotion induced. With appropriate consideration of the sample sizes and heterogeneity of each effect size estimate displayed in Tables 2–4, researchers can use this information to guide their choice of AIP. To enhance the practical applicability of these results, we also took note of which film selections and music selections were most popularly used. For example, the most commonly used film was the father's death scene from the film *The Champ* (k = 17, N = 982, δ = 3.53), which was used to induce sadness (other films used across multiple studies included the suicide scene from *Dead Poets' Society,* the mother's death scene from *Stepmom,* and the young woman's death from *Steel Magnolias,* which were all used to induce sadness or negative affect). Although there was no commonly used film to induce positive affect, most of the film-based positive affect inductions were clips from a comedic movie or TV show (k = 25) or stand-up comedy routine (k = 12). For those interested in using a music-based AIP, two

music selections were commonly used across the studies in our meta-analytic database: "Russia Under the Mongolian Yoke" by Prokofiev (played at half speed) was most commonly used to induce negative affect ($k = 16$, $N = 730$, $\delta = 1.34$), and "Coppelia" by Delibes was most commonly used to induce positive affect ($k = 7$, $N = 377$, $\delta = .95$).

What Enhances the Effectiveness of AIPs?

Instruction. First, we highlight that AIPs in which the participations were given instructions were not always more effective than AIPs without instructions, suggesting that merely adding instructions to the protocol will not necessarily increase effectiveness. To maximize effectiveness, our results suggest that truthful instructions should be administered. Specifically, the emotion eliciting effects of AIPs utilizing truthful instruction were substantially larger than the effects of AIPs using deceptive instructions, replicating Martin (1990). Of course, this finding could indicate that truthful instructions, in which the aim of the manipulation is made known to participants, increase demand characteristic effects and thus their stronger effectiveness could be due to artificial inflation of true emotional change. As such, researchers ought to consider the implications of this possibility when deciding how to present AIPs in their own work. Fortunately for research in which demand characteristics are a concern, our results suggest AIPs with deceptive instructions were still effective at eliciting their target emotions, on average, with nearly a one standard deviation change in emotions. Thus, although the most effective AIPs involve truthful instructions, AIPs were still effective, on average, with deceptive instructions.

Single versus combined AIPs. Our results surprisingly contrast with previous reviews that indicated combined AIPs are more effective at eliciting targeted emotions than single AIPs (Westermann et al., 1996). Instead, our results show that multiple AIPs do not outperform the use of single, stand-alone AIPs, on average. Although some combined AIPs (e.g., music with autobiographical recall) outperformed some stand-alone AIPs (e.g., music), none of the strongest effect sizes were from combined AIPs. Thus, the benefit of adding another layer of induction to a study appears to be minimal. We recommend that researchers avoid the assumption that more is better when administering AIPs.

Which Emotions Are Easiest to Elicit?

Positive versus negative affect. Another important finding from our meta-analysis is that the effect size of negative AIPs for eliciting negative emotions is nearly twice as large as the effect size of positive AIPs for eliciting positive emotions. This not only contributes to evidence supporting the negativity bias (Rozin & Royzman, 2001) and the "bad is stronger than good" phenomenon (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001, p. 323), but it also provides the first meta-analytic estimate of just how strong this phenomenon is—specifically, bad is nearly twice as strong as good (with regard to AIPs). Overall, these findings emphasize the difficulty of emotion regulation in response to negative emotional stimuli, which highlights the importance of research on emotion regulation and the critical role of clinical psychologists and counselors in applying this research to those who struggle with overcoming the negativity bias.

Discrete emotions. Our results also suggest that there is variability in the specific emotions that are most strongly elicited with experimental manipulations. In particular, our findings indicate that disgust is the most easily manipulated discrete emotion, with AIPs targeting disgust producing effects approximately six times as large as the effects of sadness, anger, anxiety, and happiness AIPs. The susceptibility of disgust to experimental manipulation fits with the developmental and evolutionary origins of this emotion. Namely, humans have a sensitivity to stimuli that have the potential for bodily harm, which is necessary for protection from toxic substances, disease, and dangerous animals/insects (Chapman, Kim, Susskind, & Anderson, 2009; Davey, 1994). Although all emotions have been thought to serve an evolutionary function (Scherer, 1984), disgust is unique in that it is a threat-based emotion, and therefore should be a fast-acting, strong response to serve its protective function, which appears to be supported by the evidence presented in the current review.

Although we expected all discrete emotions associated with fight-or-flight responses (e.g., disgust, anxiety, anger) to exhibit strong AIP effects, only disgust exhibited uniquely strong AIP effectiveness. This may suggest that anxiety and anger, despite their threat-based functions, are difficult to create in an artificial environment that involves artificial threats. In addition, although we expected sadness to display weak AIP effects, results suggested that sadness inductions are stronger than previously thought—so strong that even looking at pictures of other sad individuals can increase one's sadness by over four standard deviations, on average. Ultimately, these results indicate that although all emotions can be successfully induced, some emotions are sticky and more difficult to change (e.g., calm, happiness, anxiety, anger), whereas others are mutable and easier to catch from environmental stimuli (e.g., disgust, sadness).

For Whom Are AIPs Most Effective?

The current analysis also examined the effects of AIPs on different subpopulations, including comparing AIP effects across gender, culture/geographic region, and clinical populations. Our results from gender comparisons were in concurrence with previous findings (Westermann et al., 1996), with women exhibiting more sensitivity to AIPs, particularly negative AIPs, compared with men. As such, our review provides some evidence supporting the superior effectiveness of AIPs in women (or stated differently, the greater emotional reactivity of women to negative stimuli). Importantly, these gender differences may be attributable to the role of stereotypes in generating these self-reports. Given that women are often stereotyped as more emotional than men, their responses may follow these gender stereotypes by affecting their perceptions of self-report response anchors (e.g., men may feel less comfortable than women self-reporting emotional change in response to an AIP; Robinson & Clore, 2002). Future research would benefit from an examination of AIP effectiveness across gender using measures other than self-report such as physiological assessments or facial expression analysis (e.g., iMotions' Affectiva software or Noldus' FaceReader software).

We also compared the effectiveness of AIPs across different geographic regions. Although we recommend that researchers remain mindful of the emotional norms of a particular culture when designing their studies and selecting appropriate AIPs, our

results suggest that AIPs appear to be useful tools to manipulate emotions to a similar degree across geographical regions. One interesting exception to the lack of cross-cultural differences in our meta-analysis involved the notable absence of a negativity bias in Australian/New Zealand samples. Perhaps Australians' lack of negativity bias helps explain why some have found Australian samples to be higher in happiness than samples from other countries (Diener, Diener, & Diener, 1995)—an area that should be addressed in future research.

Finally, although AIPs are most frequently used to study non-clinical participants, some studies in our data utilized clinical samples (i.e., samples of individuals with clinically diagnosed disorders). The results suggest that AIPs are equally effective in clinical and non-clinical samples, although the higher levels of mean negative affect in clinical samples pre- and post-AIP warrant caution regarding the use of negative AIPs in these populations. We note that the effectiveness of positive AIPs in clinical samples may point to the importance of AIPs as a therapeutic tool, given that these individuals are often seeking treatment to improve their mood (e.g., Demyttenaere et al., 2015) and the current evidence suggests their mood can be improved via AIPs.

Do AIP Results Support the Coupling of Positive and Negative Emotions?

Our broad meta-analytic examination of emotional responsiveness to AIPs afforded us a unique approach to examine the bipolarity versus independence of emotions by assessing the patterns of coupled versus decoupled response of positive and negative emotions to positive and negative AIPs. We found that in addition to increasing positive emotions, positive AIPs also decreased negative emotions. Similarly, while increasing negative emotions, negative AIPs also decreased positive emotions. This pattern of findings highlights the coupled nature of emotional responsiveness to AIPs, providing evidence in support of the bipolarity hypothesis in which positive and negative emotions respond to emotional stimuli in tandem with one another. Interestingly, this coupled movement emerged using the PANAS measure (Watson et al., 1988), which is conceptually orthogonal and designed around a bivariate theory of emotions in which positive and negative emotions are decoupled. Our analyses suggest that a bipolar view of emotions is tenable—at least within experimental contexts and for short-term emotions (as shown by Diener and Emmons [1984], the case may be different for long-term moods). It is important to note, however, that this does not speak to whether mixed emotions—high (or low) levels of both positive and negative emotions simultaneously—can possibly be elicited through experimental protocols, as past research suggests that positive and negative emotions can be elicited independently from one another (Larsen & Green, 2013; Schimmack & Colcombe, 2007). However, there remain questions about whether these can be strongly elicited so that positivity and negativity are both strongly felt, which has not been shown in past studies to date (Tay & Kuykendall, 2017). It is also worth noting that although our results suggest coupled activation of positive and negative emotions, our results also suggest the AIPs did not induce true negative emotional states, suggesting that the coupled activation of positive and negative emotions should be revisited with an examination of AIPs that produce high mean-levels of negative emotions.

In summary, existing AIPs shift positive (or negative) emotions, but also move negative (or positive) emotions. Not only do these findings support a version of mood bipolarity for momentary emotions, but they also raise caution regarding how mood manipulation findings are interpreted. When AIPs affect dependent variables (e.g., cognitive or behavioral outcomes), we must be careful in interpreting whether these changes are attributable to increases in the level of the targeted mood, or perhaps decreases in the mood of the opposite valence. Therefore, we strongly encourage researchers to measure both types of emotions and to consider this dual movement when interpreting their results.

Do Negative AIPs Overcome the Positivity Offset?

In general, most people, while being more influenced by negative stimuli in the short term, tend to experience more positive than negative emotions, in general (Diener & Diener, 1996). We found evidence of this positivity bias in the present analysis; participants reported substantially higher positive emotion scores than negative emotion scores at baseline, prior to experiencing an AIP. Importantly, we found that this positivity bias was robust to negative AIPs. Negative AIPs only reduced the ratio of positive-to-negative emotions but did not reverse it; even after negative AIPs, participants still reported higher positive emotions than negative emotions. Thus, although negative AIPs are effective in changing emotion, they are not necessarily effective in inducing a predominately negative emotional state. Stated differently, our results suggest that negative AIPs are not strong enough to overcome the positivity bias.

This has implications for the interpretation of many findings resulting from the use of AIP analyses. Practically, our results suggest researchers can assess the effects of decreases in positive emotions or increases in negative emotions in response to a negative AIP, but, in general, it appears to be difficult to create a laboratory manipulation that creates a truly negative emotional state. This would suggest that previous research on the relationships between negative emotions and outcomes that has relied on AIPs may have underestimated these relationships, given that only mild negative emotions may have been induced. Of course, developing stronger negative AIPs may be difficult with respect to ethical research practices. However, researchers utilizing existing negative AIPs should be cautious in interpreting findings as the result of negative emotions, given that ratings of emotion after a negative AIP appear to be more positive than negative on average.

Limitations and Future Directions

Although the present work offers a broad and detailed analysis of the effects of AIPs on emotional responses, there are still limitations of this work that we must address. First, it is important to note that our meta-analytic effect sizes represent the primary studies that comprise them and as such, the strength of our meta-analytic effect sizes is influenced by the extent to which the more effective AIPs were chosen by the primary study authors. For example, although our overall meta-analytic effect size represents 874 samples and, as such, is a fairly stable estimate of AIP effectiveness, we note that more of these samples used a negative AIP than a positive AIP, which may have inflated the overall effectiveness estimate, given that negative AIPs exhibit larger

effect sizes. We also note that although many of our meta-analytic effect sizes indicated AIPs were significantly effective (i.e., the 95% confidence interval did not include zero), there was great variability in many of these effects (often, the correct standard deviation was as large as the effect size). Thus, researchers should be careful to examine both the mean effect size and the standard deviation in choosing an AIP that is appropriate for their study and avoid the assumption that AIPs with strong mean effect sizes are foolproof.

Second, we note that because of the low availability of discrete emotion AIPs, we could only present meta-analyses of happiness, sadness, anger, anxiety, disgust, and calm AIPs. Notably absent from this list is an examination of AIPs designed to induce guilt, shame, surprise, relief, pride, contempt, and so forth. We use the lack of available data as a signpost for future researchers to investigate these less popular AIPs. Third, we acknowledge that our test of the bipolarity hypothesis and positivity offset were based on a single measure of emotion, the PANAS (Watson et al., 1988). We chose the PANAS because it was the most widely used measure of both positive and negative emotion in our meta-analytic database, which allowed us to directly compare pre- and post-AIP positive and negative emotion on the same scale. Nevertheless, our results should be confirmed using a different measure. Fourth, we note that our comparison of AIPs across geographic region was not nearly as fine-grained as it ideally should have been, given that we lumped countries with fairly extensive cultural differences into single geographic regions. Future work may benefit from an examination of whether AIP strength varies directly with cultural values (e.g., individualism/collectivism).

Finally, in the process of meta-analyzing the AIP literature, we had hoped to comment on how long the effects of AIPs last. Unfortunately, of the 874 independent samples in our meta-analysis, only 12 reported the length of time between the induction and the post-induction assessment (we hoped to analyze this time lag as a predictor of effectiveness), and only 15 reported multiple post-AIP measures of affect that could be used to analyze how these effects degrade over time. Given the dearth of available data to examine how long AIPs last, we identify this as a critical issue for future research on AIPs, and we urge future work to report the time lag between induction and assessment and report multiple posttest effect sizes whenever possible. Similarly, we were also unable to comment on the effectiveness of reset inductions, or AIPs administered after a negative AIP to eliminate any negative effects, because although 192 samples reported the use of a reset induction, only 10 samples reported enough information to calculate the effectiveness of their reset inductions. These 10 samples are shown in Table 7 along with the effect size from each sample that indicates the standardized change in negative affect from pre-induction to post-reset. Positive *d* values indicate negative mood has not returned to baseline (a *d* value of zero would indicate a complete return to baseline), whereas negative *d* values indicate the participants left the experiment in a more positive state than when they began. From an examination of these effect sizes, the reset inductions appear to be successful (the majority were negative or very close to zero). The most effective reset induction involved thinking about a happy memory while positive music played for four minutes (Ridout, Noreen, & Johal, 2009). However, given the general lack of available information regarding the effectiveness of various reset inductions, we call on AIP researchers to report data following their reset inductions, allowing for the examination of whether negative AIPs exhibit residual effects beyond the experiment, and which reset inductions are most effective.

Table 7
Effect Sizes From Samples Reporting Reset Inductions With Postreset Affect Measurement

Study	$d_{baseline\text{-}postreset}$
Vanderlind, Stanton, Weinbrecht, Velkoff, & Joormann, 2017	0.27
Vara et al., 2016	0.16
Vara et al., 2016	0.04
Heene, De Raedt, Buysse, & Van Oost, 2007	0.01
Bernstein & McNally, 2017	−0.15
Macatee, Albanese, Schmidt, & Cougle, 2017	−0.24
Joormann, Talbot, & Gotlib, 2007	−0.52
Joormann et al., 2007	−0.71
Vara et al., 2016	−0.97
Ridout, Noreen, & Johal, 2009	−1.16

Note. $d_{baseline\text{-}postreset}$ is the standardized mean change from baseline measurement of negative affect to post-reset-induction measurement of negative affect. Positive values indicate negative affect has not returned to baseline, whereas negative values indicate the participants ended the experiment in a more positive state than when they began.

Conclusions

The present study took a broad approach in reviewing how existing AIPs induce a range of emotional responses. A number of important conclusions can be taken from this analysis. First, we found evidence that AIPs work; they are effective at inducing their intended emotions. We provide further information regarding which strategies work the best for eliciting particular emotions, as well as providing information regarding methodological features and sample moderators that influence the effectiveness of AIPs in eliciting target emotions. We recommend that researchers utilize these results in guiding their selection of AIPs for use in future studies and we provide a list of recommendations for researchers interested in using AIPs in Table 8. In addition, we argue that our findings are not only relevant for researchers who choose to use AIPs in their work; AIPs can also be used in applied settings by counselors/psychologists to train patients on how to use AIPs in their own lives, by leaders who acknowledge the importance of displaying positive emotions (Joseph, Dhanani, Shen, McHugh, & McCord, 2015) and may be interested in enhancing follower affect via AIP, by coaches/trainers/mentors/teachers who desire to enhance their mentee's affective state, and/or by employees who wish to enhance their own affective state at work.

In addition to providing an updated and nuanced quantitative review of the effects of AIPs on emotions, we leveraged the meta-analytic data to examine theoretical questions about emotions. Specifically, we found support for three theoretical propositions regarding emotion. First, we found evidence for the negativity bias, demonstrating that people respond to negative stimuli more strongly than they respond to positive stimuli. Second, our analyses revealed a bipolar coupling of positive and negative emotions in response to experimental manipulations, supporting the bipolarity hypothesis. Third, we found evidence supporting a

Table 8
Advice for Researchers Using Affect Induction Procedures

1. *Choose your AIP carefully.*
Although the average AIP is effective, the most effective AIP is eight times more effective than the least effective AIP, and the least effective AIPs can change mood in the opposite direction of what was intended.
2. *Choose your AIP type based on the emotion you are intending to induce.*
We recommend pictures of facial expressions for inducing broad negative affect, sadness, or happiness; film for inducing broad positive affect; and autobiographical recall for anxiety or anger.
3. *Don't assume that the use of multiple AIPs is more effective than a single AIP.*
Multiple AIPs were not more effective than single AIPs, on average.
4. *Be truthful about the purpose of the AIP to the participants whenever possible.*
Results demonstrate that truthful AIPs are more effective than deceptive AIPs. However, if demand characteristics are a concern, deceptive AIPs are effective, on average, and can be used.
5. *Assess whether your negative affect induction produces a negative emotional state.*
Results suggest that negative AIPs produce only mild negative emotions, on average. If an AIP is being used to test hypotheses that involve negative emotions, one should interpret mean levels of negative emotion after the AIP to determine whether the AIP is inducing negative affect to the desired degree.
6. *When interpreting results that involve an AIP, acknowledge the coupled activation of positive and negative emotion.*
Our findings suggest that the induction of positive emotion reduces negative emotion, and the induction of negative emotion reduces positive emotion. This should be acknowledged when interpreting the results of an AIP (e.g., change due to an increase in negative mood may also be explained as change due to a decrease in positive mood).
7. *Report the timing of the induction and post-AIP assessment data.*
Given the need for additional data about how long AIPs last, all AIP research should report the time lag between the induction and the post-induction assessment(s). All effect sizes involving post-induction assessments should be reported.
8. *Use a reset induction after a negative AIP and report effect sizes from this reset to assess effectiveness.*
Reset inductions should be given to all negative AIP participants (rather than offered to participants). This is especially important in clinical samples, who have higher mean post-AIP negative affect than non-clinical samples.
9. *Measure the induced emotion.*
Rather than inducing a discrete emotion (e.g., anxiety) and assessing the effectiveness of the induction with a broad affect measure (e.g., negative affect), match the measure with the induced emotion (e.g., an anxiety induction should be assessed with an anxiety measure).

Note. AIP = affect induction procedures.

positivity offset, in which individuals tend to feel positive emotions, in general, and these positive emotions appear to be stronger than negative emotions following negative AIPs, again supporting individuals' tendencies to maintain positive emotions. AIPs are very important tools that have enabled researchers to study the relationships between emotions and many psychological processes, and we hope these results will aid researchers in leveraging AIPs effectively to continue answering important questions about the influence of emotions across domains.

References

Albersnagel, F. A. (1988). Velten and musical mood induction procedures: A comparison with accessibility of thought associations. *Behaviour Research and Therapy, 26,* 79–95. http://dx.doi.org/10.1016/0005-7967(88)90035-6

Avramova, Y. R., & Stapel, D. A. (2008). Moods as spotlights: The influence of mood on accessibility effects. *Journal of Personality and Social Psychology, 95,* 542–554. http://dx.doi.org/10.1037/a0012560

Baas, M., De Dreu, C. K., & Nijstad, B. A. (2008). A meta-analysis of 25 years of mood-creativity research: Hedonic tone, activation, or regulatory focus? *Psychological Bulletin, 134,* 779–806. http://dx.doi.org/10.1037/a0012815

Barrett, L. F., & Bliss-Moreau, E. (2009). She's emotional. He's having a bad day: Attributional explanations for emotion stereotypes. *Emotion, 9,* 649–658. http://dx.doi.org/10.1037/a0016821

Barrett, L. F., Robin, L., Pietromonaco, P. R., & Eyssell, K. M. (1998). Are women the "more emotional" sex? Evidence from emotional experiences in social context. *Cognition and Emotion, 12,* 555–578. http://dx.doi.org/10.1080/026999398379565

Barrett, L. F., & Russell, J. A. (2003, February). Introduction. In L. F. Barrett & J. A. Russell (Chairs), *Core affect: Theory and research.* Symposium conducted at the annual meeting of the Society for Personality and Social Psychology, Los Angeles, CA.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5,* 323–370. http://dx.doi.org/10.1037/1089-2680.5.4.323

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50,* 1088–1101. http://dx.doi.org/10.2307/2533446

Bernstein, E. E., & McNally, R. J. (2017). Acute aerobic exercise helps overcome emotion regulation deficits. *Cognition and Emotion, 31,* 834–843. http://dx.doi.org/10.1080/02699931.2016.1168284

Bless, H., Bohner, G., Schwarz, N., & Strack, F. (1990). Mood and persuasion: A cognitive response analysis. *Personality and Social Psychology Bulletin, 16,* 331–345. http://dx.doi.org/10.1177/0146167290162013

Bless, H., Schwarz, N., & Kemmelmeier, M. (1996). Mood and stereotyping: Affective states and the use of general knowledge structures. *European Review of Social Psychology, 7,* 63–93. http://dx.doi.org/10.1080/14792779443000102

Bodenhausen, G. V., Kramer, G. P., & Süsser, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of Personality and Social Psychology, 66,* 621–632. http://dx.doi.org/10.1037/0022-3514.66.4.621

Bond, M. H. (1993). Emotions and their expression in Chinese culture. *Journal of Nonverbal Behavior, 17,* 245–262. http://dx.doi.org/10.1007/BF00987240

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36,* 129–148. http://dx.doi.org/10.1037/0003-066X.36.2.129

Brannick, M. T., Yang, L.-Q., & Cafri, G. (2011). Comparison of weights for meta-analysis of *r* and *d* under realistic conditions. *Organizational Research Methods, 14,* 587–607. http://dx.doi.org/10.1177/1094428110368725

Brenner, E. (2000). Mood induction in children: Methodological issues and clinical implications. *Review of General Psychology, 4,* 264–283. http://dx.doi.org/10.1037/1089-2680.4.3.264

Brody, L. R., & Hall, J. A. (2008). Gender and emotion in context. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (pp. 338–349). New York, NY: Guilford.

Buchwald, A. M., Strack, S., & Coyne, J. C. (1981). Demand characteristics and the Velten mood induction procedure. *Journal of Consulting and Clinical Psychology, 49,* 478–479. http://dx.doi.org/10.1037/0022-006X.49.3.478

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115,* 401–423. http://dx.doi.org/10.1037/0033-2909.115.3.401

Cacioppo, J. T., & Berntson, G. G. (1999). The affect system: Architecture and operating characteristics. *Current Directions in Psychological Science, 8,* 133–137. http://dx.doi.org/10.1111/1467-8721.00031

Chacón-Moscoso, S., Sanduvete-Chaves, S., & Sánchez-Martín, M. (2016). The development of a checklist to enhance methodological quality in intervention programs. *Frontiers in Psychology, 7,* 1811. http://dx.doi.org/10.3389/fpsyg.2016.01811

Chakrabarti, B., Bullmore, E., & Baron-Cohen, S. (2006). Empathizing with basic emotions: Common and discrete neural substrates. *Social Neuroscience, 1,* 364–384. http://dx.doi.org/10.1080/17470910601041317

Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In bad taste: Evidence for the oral origins of moral disgust. *Science, 323,* 1222–1226. http://dx.doi.org/10.1126/science.1165565

Clark, D. M. (1983). On the induction of depressed mood in the laboratory: Evaluation and comparison of the Velten and musical procedures. *Advances in Behaviour Research and Therapy, 5,* 27–49. http://dx.doi.org/10.1016/0146-6402(83)90014-0

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671–684. http://dx.doi.org/10.1037/h0043943

Davey, G. C. L. (1994). Self-reported fears to common indigenous animals in an adult U. K. population: The role of disgust sensitivity. *British Journal of Psychology, 85,* 541–554. http://dx.doi.org/10.1111/j.2044-8295.1994.tb02540.x

Demyttenaere, K., Donneau, A. F., Albert, A., Ansseau, M., Constant, E., & van Heeringen, K. (2015). What is important in being cured from depression? Discordance between physicians and patients. *Journal of Affective Disorders, 174,* 390–396. http://dx.doi.org/10.1016/j.jad.2014.12.004

Diener, E., & Diener, C. (1996). Most people are happy. *Psychological Science, 7,* 181–185. http://dx.doi.org/10.1111/j.1467-9280.1996.tb00354.x

Diener, E., Diener, M., & Diener, C. (1995). Factors predicting the subjective well-being of nations. *Journal of Personality and Social Psychology, 69,* 851–864. http://dx.doi.org/10.1037/0022-3514.69.5.851

Diener, E., & Emmons, R. A. (1984). The independence of positive and negative affect. *Journal of Personality and Social Psychology, 47,* 1105–1117. http://dx.doi.org/10.1037/0022-3514.47.5.1105

Diener, E., Kanazawa, S., Suh, E. M., & Oishi, S. (2015). Why people are in a generally good mood. *Personality and Social Psychology Review, 19,* 235–256. http://dx.doi.org/10.1177/1088868314544467

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455–463. http://dx.doi.org/10.1111/j.0006-341X.2000.00455.x

Ekman, P. (1994). Moods, emotions, and traits. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 56–58). New York, NY: Oxford University Press.

Else-Quest, N. M., Higgins, A., Allison, C., & Morton, L. C. (2012). Gender differences in self-conscious emotional experience: A meta-analysis. *Psychological Bulletin, 138,* 947–981. http://dx.doi.org/10.1037/a0027930

Feldman-Barrett, L., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science, 8,* 10–14. http://dx.doi.org/10.1111/1467-8721.00003

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York, NY: McGraw-Hill.

Fredrickson, B. L. (1998). What good are positive emotions? *Review of General Psychology, 2,* 300–319. http://dx.doi.org/10.1037/1089-2680.2.3.300

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology. The broaden-and-build theory of positive emotions. *American Psychologist, 56,* 218–226. http://dx.doi.org/10.1037/0003-066X.56.3.218

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition and Emotion, 19,* 313–332. http://dx.doi.org/10.1080/02699930441000238

Gasper, K. (2003). When necessity is the mother of invention: Mood and problem solving. *Journal of Experimental Social Psychology, 39,* 248–262. http://dx.doi.org/10.1016/S0022-1031(03)00023-4

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. http://dx.doi.org/10.1111/j.2044-8295.1994.tb02508.x

Goodwin, A. M., & Williams, J. M. G. (1982). Mood-induction research—Its implications for clinical depression. *Behaviour Research and Therapy, 20,* 373–382. http://dx.doi.org/10.1016/0005-7967(82)90097-3

Gray, J. A. (1981). A critique of Eysenck's theory of personality. In H. J. Eysenck (Ed.), *A model for personality* (pp. 246–276). New York, NY: Springer. http://dx.doi.org/10.1007/978-3-642-67783-0_8

Gross, J. J. (2010). The future's so bright, I gotta wear shades. *Emotion Review, 2,* 212–216. http://dx.doi.org/10.1177/1754073910361982

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion, 9,* 87–108. http://dx.doi.org/10.1080/02699939508408966

Grossman, M., & Wood, W. (1993). Sex differences in intensity of emotional experience: A social role interpretation. *Journal of Personality and Social Psychology, 65,* 1010–1022. http://dx.doi.org/10.1037/0022-3514.65.5.1010

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Heene, E., De Raedt, R., Buysse, A., & Van Oost, P. (2007). Does negative mood influence self-report assessment of individual and relational measures? An experimental analysis. *Assessment, 14,* 86–93. http://dx.doi.org/10.1177/1073191106293980

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology, 82,* 903–918. http://dx.doi.org/10.1037/0022-3514.82.6.903

Henderson, J. R., & Lohr, J. M. (1982). The effect of statement valence and referent content upon mood and anagram performance. *Cognitive Therapy and Research, 6,* 461–464. http://dx.doi.org/10.1007/BF01184013

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781412985031

Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology, 52,* 1122–1131. http://dx.doi.org/10.1037/0022-3514.52.6.1122

Isen, A. M., Shalker, T. E., Clark, M., & Karp, L. (1978). Affect, accessibility of material in memory, and behavior: A cognitive loop? *Journal of Personality and Social Psychology, 36,* 1–12. http://dx.doi.org/10.1037/0022-3514.36.1.1

Ito, T. A., Cacioppo, J. T., & Lang, P. J. (1998). Eliciting affect using the International Affective Picture System: Trajectories through evaluative space. *Personality and Social Psychology Bulletin, 24,* 855–879. http://dx.doi.org/10.1177/0146167298248006

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology, 75,* 887–900. http://dx.doi.org/10.1037/0022-3514.75.4.887

Joormann, J., & Gotlib, I. H. (2007). Selective attention to emotional faces following recovery from depression. *Journal of Abnormal Psychology, 116,* 80–85. http://dx.doi.org/10.1037/0021-843X.116.1.80

Joormann, J., Talbot, L., & Gotlib, I. H. (2007). Biased processing of emotional information in girls at risk for depression. *Journal of Abnormal Psychology, 116,* 135–143. http://dx.doi.org/10.1037/0021-843X.116.1.135

Joseph, D. L., Dhanani, L. Y., Shen, W., McHugh, B. C., & McCord, M. A. (2015). Is a happy leader a good leader? A meta-analytic investigation of leader trait affect and leadership. *The Leadership Quarterly, 26,* 557–577. http://dx.doi.org/10.1016/j.leaqua.2015.04.001

Kenealy, P. M. (1986). The Velten mood induction procedure: A methodological review. *Motivation and Emotion, 10,* 315–335. http://dx.doi.org/10.1007/BF00992107

Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods, 15,* 624–662. http://dx.doi.org/10.1177/1094428112452760

Kitayama, S., Markus, H. R., & Kurokawa, M. (2000). Culture, emotion, and well-being: Good feelings in Japan and the United States. *Cognition and Emotion, 14,* 93–124. http://dx.doi.org/10.1080/026999300379003

Landis, C. (1924). Studies of emotional reactions, II. General behavior and facial expression. *Journal of Comparative Psychology, 4,* 447–510. http://dx.doi.org/10.1037/h0073039

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*. Gainesville: University of Florida.

Larsen, J. T., & Green, J. D. (2013). Evidence for mixed feelings of happiness and sadness from brief moments in time. *Cognition and Emotion, 27,* 1469–1477. http://dx.doi.org/10.1080/02699931.2013.790782

Larsen, J. T., & McGraw, A. P. (2011). Further evidence for mixed emotions. *Journal of Personality and Social Psychology, 100,* 1095–1110. http://dx.doi.org/10.1037/a0021846

Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology, 81,* 684–696. http://dx.doi.org/10.1037/0022-3514.81.4.684

Larsen, R. J., & Ketelaar, T. (1989). Extraversion, neuroticism and susceptibility to positive and negative mood induction procedures. *Personality and Individual Differences, 10,* 1221–1228. http://dx.doi.org/10.1016/0191-8869(89)90233-X

Larsen, R. J., & Sinnett, L. M. (1991). Meta-analysis of experimental manipulations: Some factors affecting the Velten mood induction procedure. *Personality and Social Psychology Bulletin, 17,* 323–334. http://dx.doi.org/10.1177/0146167291173013

Lench, H. C., Flores, S. A., & Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: A meta-analysis of experimental emotion elicitations. *Psychological Bulletin, 137,* 834–855. http://dx.doi.org/10.1037/a0024244

Lenton, S. R., & Martin, P. R. (1991). The contribution of music vs instructions in the Musical Mood Induction Procedure. *Behaviour Research and Therapy, 29,* 623–625. http://dx.doi.org/10.1016/0005-7967(91)90011-Q

Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition and Emotion, 14,* 473–493. http://dx.doi.org/10.1080/026999300402763

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology, 66,* 799–823. http://dx.doi.org/10.1146/annurev-psych-010213-115043

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage

Macatee, R. J., Albanese, B. J., Schmidt, N. B., & Cougle, J. R. (2017). The moderating influence of heart rate variability on stressor-elicited change in pupillary and attentional indices of emotional processing: An eye-Tracking study. *Biological Psychology, 123,* 83–93. http://dx.doi.org/10.1016/j.biopsycho.2016.11.013

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95,* 15–20. http://dx.doi.org/10.1037/0021-843X.95.1.15

Marín-Martínez, F., & Sánchez-Meca, J. (2010). Weighting by inverse variance or by sample size in random-effects meta-analysis. *Educational and Psychological Measurement, 70,* 56–73. http://dx.doi.org/10.1177/0013164409344534

Marston, A., Hart, J., Hileman, C., & Faunce, W. (1984). Toward the laboratory study of sadness and crying. *The American Journal of Psychology, 97,* 127–131. http://dx.doi.org/10.2307/1422552

Martin, M. (1990). On the induction of mood. *Clinical Psychology Review, 10,* 669–697. http://dx.doi.org/10.1016/0272-7358(90)90075-L

Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology, 1,* 167–195. http://dx.doi.org/10.1146/annurev.clinpsy.1.102803.143916

Miller, N. E. (1961). Some recent studies on conflict behavior and drugs. *American Psychologist, 16,* 12–24. http://dx.doi.org/10.1037/h0048720

Morelli, S. A., & Lieberman, M. D. (2013). The role of automaticity and attention in neural processes underlying empathy for happiness, sadness, and anxiety. *Frontiers in Human Neuroscience, 7,* 160. http://dx.doi.org/10.3389/fnhum.2013.00160

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11,* 364–386. http://dx.doi.org/10.1177/1094428106291059

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105–125. http://dx.doi.org/10.1037/1082-989X.7.1.105

Nummenmaa, L., & Niemi, P. (2004). Inducing affective states with success-failure manipulations: A meta-analysis. *Emotion, 4,* 207–214. http://dx.doi.org/10.1037/1528-3542.4.2.207

Olatunji, B. O., & Wolitzky-Taylor, K. B. (2009). Anxiety sensitivity and the anxiety disorders: A meta-analytic review and synthesis. *Psychological Bulletin, 135,* 974–999. http://dx.doi.org/10.1037/a0017428

Papousek, I., Schulter, G., & Lang, B. (2009). Effects of emotionally contagious films on changes in hemisphere-specific cognitive performance. *Emotion, 9,* 510–519. http://dx.doi.org/10.1037/a0016299

Park, J., & Banaji, M. R. (2000). Mood and heuristics: The influence of happy and sad states on sensitivity and bias in stereotyping. *Journal of Personality and Social Psychology, 78,* 1005–1023. http://dx.doi.org/10.1037/0022-3514.78.6.1005

Pennebaker, J. W., & Roberts, T.-A. (1992). Toward a his and hers theory of emotion: Gender differences in visceral perception. *Journal of Social and Clinical Psychology, 11,* 199–212. http://dx.doi.org/10.1521/jscp.1992.11.3.199

Petty, R. E., & Briñol, P. (2015). Emotion and persuasion: Cognitive and meta-cognitive processes impact attitudes. *Cognition and Emotion, 29,* 1–26. http://dx.doi.org/10.1080/02699931.2014.967183

Philippot, P. (1993). Inducing and assessing differentiated emotion-feeling states in the laboratory. *Cognition and Emotion, 7,* 171–193. http://dx.doi.org/10.1080/02699939308409183

Pineles, S. L., & Mineka, S. (2005). Attentional biases to internal and external sources of potential threat in social anxiety. *Journal of Abnor-*

mal Psychology, 114, 314–318. http://dx.doi.org/10.1037/0021-843X.114.2.314

Ridout, N., Noreen, A., & Johal, J. (2009). Memory for emotional faces in naturally occurring dysphoria and induced sadness. *Behaviour Research and Therapy, 47,* 851–860. http://dx.doi.org/10.1016/j.brat.2009.06.013

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin, 128,* 934–960. http://dx.doi.org/10.1037/0033-2909.128.6.934

Rottenberg, J., Ray, R. R., & Gross, J. J. (2007). Emotion elicitation using films. In J. A. Coan & J. J. B. Allen (Eds.), *The handbook of emotion elicitation and assessment* (pp. 9–28). New York, NY: Oxford University Press.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5,* 296–320. http://dx.doi.org/10.1207/S15327957PSPR0504_2

Russell, J. A. (1983). Pancultural aspects of the human conceptual organization of emotions. *Journal of Personality and Social Psychology, 45,* 1281–1288. http://dx.doi.org/10.1037/0022-3514.45.6.1281

Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological Bulletin, 110,* 426–450. http://dx.doi.org/10.1037/0033-2909.110.3.426

Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin, 125,* 3–30. http://dx.doi.org/10.1037/0033-2909.125.1.3

Scherer, K. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. E. Ekman (Eds.), *Approaches to emotion* (pp. 293–317). Hillsdale, NJ: Erlbaum.

Schimmack, U., & Colcombe, S. J. (2007). Eliciting mixed feelings with the paired-picture paradigm: A tribute to Kellog (1915). *Cognition and Emotion, 21,* 1546–1553. http://dx.doi.org/10.1080/02699930601057011

Schneider, F., Gur, R. C., Gur, R. E., & Muenz, L. R. (1994). Standardized mood induction with happy and sad facial expressions. *Psychiatry Research, 51,* 19–31. http://dx.doi.org/10.1016/0165-1781(94)90044-2

Shields, S. A. (2002). *Speaking from the heart: Gender and the social meaning of emotion*. New York, NY: Cambridge University Press.

Sutherland, G., Newman, B., & Rachman, S. (1982). Experimental investigations of the relations between mood and intrusive unwanted cognitions. *British Journal of Medical Psychology, 55,* 127–138. http://dx.doi.org/10.1111/j.2044-8341.1982.tb01491.x

Tay, L., & Kuykendall, L. (2017). Why self-reports of happiness and sadness may not necessarily contradict bipolarity: A psychometric review and proposal. *Emotion Review, 9,* 146–154. http://dx.doi.org/10.1177/1754073916637656

Terhaar, J., Boettger, M. K., Schwier, C., Wagner, G., Israel, A. K., & Bär, K. J. (2010). Increased sensitivity to heat pain after sad mood induction in female patients with major depression. *European Journal of Pain, 14,* 559–563. http://dx.doi.org/10.1016/j.ejpain.2009.09.004

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134,* 383–403. http://dx.doi.org/10.1037/0033-2909.134.3.383

Vanderlind, W. M., Stanton, C. H., Weinbrecht, A., Velkoff, E. A., & Joormann, J. (2017). Remembering the good ole days: Fear of positive emotion relates to affect repair using positive memories. *Cognitive Therapy and Research, 41,* 362–368. http://dx.doi.org/10.1007/s10608-016-9775-z

Vara, M. D., Baños, R. M., Rasal, P., Rodríguez, A., Rey, B., Wrzesien, M., & Alcañiz, M. (2016). A game for emotional regulation in adolescents: The (body) interface device matters. *Computers in Human Behavior, 57,* 267–273. http://dx.doi.org/10.1016/j.chb.2015.12.033

Velten, E., Jr. (1968). A laboratory task for induction of mood states. *Behaviour Research and Therapy, 6,* 473–482. http://dx.doi.org/10.1016/0005-7967(68)90028-4

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063–1070. http://dx.doi.org/10.1037/0022-3514.54.6.1063

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98,* 219–235. http://dx.doi.org/10.1037/0033-2909.98.2.219

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. http://dx.doi.org/10.1002/(SICI)1099-0992(199607)26:4<557::AID-EJSP769>3.0.CO;2-4

Wilson, D. B. (2019). Meta-analysis macros for SAS, SPSS, and Stata. Retrieved August, 1, 2019, from http://mason.gmu.edu/~dwilsonb/ma.html

Received January 26, 2018
Revision received October 28, 2019
Accepted December 7, 2019 ■