Emotion
2017, Vol. 17, No. 2, 267–295

1528-3542/17/$12.00 http://dx.doi.org/10.1037/emo0000226

The Jingle and Jangle of Emotion Assessment: Imprecise Measurement, Casual Scale Usage, and Conceptual Fuzziness in Emotion Research

Aaron C. Weidman, Conor M. Steckler, and Jessica L. Tracy
University of British Columbia

Although affective science has seen an explosion of interest in measuring subjectively experienced distinct emotional states, most existing self-report measures tap broad affect dimensions and dispositional emotional tendencies, rather than momentary distinct emotions. This raises the question of how emotion researchers *are* measuring momentary distinct emotions in their studies. To address this question, we reviewed the self-report measurement practices regularly used for the purpose of assessing momentary distinct emotions, by coding these practices as observed in a representative sample of articles published in *Emotion* from 2001–2011 (n = 467 articles; 751 studies; 356 measurement instances). This quantitative review produced several noteworthy findings. First, researchers assess many purportedly distinct emotions (n = 65), a number that differs substantially from previously developed emotion taxonomies. Second, researchers frequently use scales that were not systematically developed, and that include items also used to measure at least 1 other emotion on a separate scale in a separate study. Third, the majority of scales used include only a single item, and had unknown reliability. Together, these tactics may create ambiguity regarding which emotions are being measured in empirical studies, and conceptual inconsistency among measures of purportedly identical emotions across studies. We discuss the implications of these problematic practices, and conclude with recommendations for how the field might improve the way it measures emotions.

Keywords: affect, emotion, measurement, mood, self-report

Supplemental materials: http://dx.doi.org/10.1037/emo0000226.supp

Contemporary affective science has seen a surge of interest in distinct, momentary emotional states, typically defined as including specific patterns of subjective feelings, physiological changes, neural activity, cognitive appraisals, and motivated action tendencies (Ekman, 1992; Kragel & LaBar, 2014; Roseman, 2011; Tracy & Randles, 2011). As a result, emotion researchers in recent years have demonstrated an increase in their use of self-report measures to assess and draw conclusions about the momentary experience of distinct emotions. An informal survey of articles published over the past decade in *Psychological Science*, our field's flagship journal for new empirical findings, suggests that emotion researchers regularly use self-report methods to measure a range of emotional states, including, but not limited to, anger, anxiety, awe, disgust, elevation, fear, gratitude, guilt, jealousy, nostalgia, pride, and sadness (e.g., Amodio, Devine, & Harmon-Jones, 2007; Bartlett & DeSteno, 2006; Cryder, Lerner, Gross, & Dahl, 2008; Ford et al., 2010; Hodson & Costello, 2007; Lerner, Gonzalez, Small, & Fischhoff, 2003; Lerner, Small, & Loewenstein, 2004; Levy & Kelly, 2010; Rudd, Vohs, & Aaker, 2012; Schnall, Roper, & Fessler, 2010; Sherman, Haidt, & Clore, 2012; Williams & DeSteno, 2009; Zhou, Sedikides, Wildschut, & Gao, 2008).

The wide range of emotions studied is mirrored by the breadth of research topics examined via the assessment of momentary distinct emotional states. Researchers with interests as diverse as achievement, aging, altruism, attention, close relationships, economic decision-making, moral judgment, perception, physical health, prejudice, psychopathology, and social status have employed self-report measures of momentary distinct emotions in their research (e.g., Burns, 2006; Cottrell & Neuberg, 2005; Crusius & Mussweiler, 2012; Gonzaga, Turner, Keltner, Campos, & Altemus, 2006; Horberg, Oveis, Keltner, & Cohen, 2009; Hutcherson & Gross, 2011; Inbar, Pizarro, & Bloom, 2012; Isaacowitz, & Choi, 2011; Johnson, 2009; Ketelaar & Au, 2003; Labouvie-Vief, Lumley, Jain, & Heinze, 2003; Lelieveld, van Dijk, van Beest, & van Kleef, 2012; Lishner, Batson, & Huss, 2011; McGregor & Elliot, 2005; Most, Laurenceau, Graber, Belcher, & Smith, 2010; Nelissen, Leliveld, van Dijk, & Zeelenberg, 2011; Parker Tapias, Glaser, Keltner, Vasquez, & Wickens, 2007; Quartana & Burns, 2007; Rottenberg, Kasch, Gross, & Gotlib, 2002; Sbarra & Ferrer, 2006; Schnall, Haidt, Clore, & Jordan, 2008;

This article was published Online First September 19, 2016.

Aaron C. Weidman, Conor M. Steckler, and Jessica L. Tracy, Department of Psychology, University of British Columbia.

We thank Will Dunlop, Elizabeth Dunn, David Klonsky, Victoria Savalei, Toni Schmader, and the members of the Emotion and Self Lab, for their insightful comments on earlier versions of this article. We also thank Lauren Fratamico for her help in preparing the online coding file. We also acknowledge the generous support of the Social Sciences Research Council of Canada, Standard Research Grant 410-2009-2458, a Michael Smith Foundation for Health Research Scholar Award [CI-SCH-01862(07-1)], and a Canadian Institute for Health Research New Investigator Award.

Correspondence concerning this article should be addressed to Aaron C. Weidman, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC V6T 1Z4, Canada. E-mail: acweidman@psych.ubc.ca

Sweetman, Spears, Livingstone, & Manstead, 2013; Valdesolo & DeSteno, 2011; Williams & DeSteno, 2008). Indeed, a recent review of articles published in the first two sections of *Journal of Personality and Social Psychology* in 2011 observed that social-personality researchers quite frequently measure distinct emotions via self-report, typically treating them as mediators in their theoretical models (Inzlicht & Tritt, 2014). Taken together, the prevalence of self-report research into distinct emotional states suggests that substantial resources are being dedicated to understanding the unique antecedents, phenomenology, and functional consequences of these experiences across the field of affective science, and psychology more broadly.[1]

Many Studies, Many Emotions, Few Measures

Given the field's pervasive interest in the study of momentary distinct emotional states, it is essential that researchers have access to the right tools to precisely measure these states. However, an emphasis on measuring momentary distinct emotions, as opposed to dispositional emotional tendencies and broader affect dimensions, is a fairly novel development for which the field may not be prepared. Although theories of distinct emotions have a long history in psychology (e.g., Darwin, 1872; James, 1890), and during the past half-century researchers have identified a set of nonverbal expressions that are consistently and cross-culturally associated with distinct emotions (e.g., Ekman et al., 1987; Izard, 1971; Keltner, 1995; Tracy & Robins, 2008), there has been no concerted effort to develop a comprehensive means of measuring subjectively experienced, momentary distinct emotional states via self-report.

Instead, following interest in dimensionalist models of emotion in the 1980s (e.g., Russell, 1980; Watson & Tellegen, 1985), emotion researchers developed several measures of broad affect dimensions, including the Positive and Negative Affect Schedule, or PANAS (Watson, Clark, & Tellegen, 1988), and the Current Mood Adjectives (Barrett & Russell, 1998). These scales have become widely and frequently used; according to Google Scholar, the PANAS has been cited more than 22,000 times as of this writing, and the Current Mood Adjectives have been cited approximately 900 times, providing an indication of their prevalence in contemporary psychological science. In addition, numerous measures have been developed to assess the trait-like dispositional tendency to experience a number of distinct emotions, such as proneness to anger, awe, compassion, disgust, embarrassment, envy, gratitude, guilt and shame, happiness, and pride, to name a few (e.g., Buss & Durkee, 1957; Cohen, Wolf, Panter, & Insko, 2011; Haidt, McCauley, & Rozin, 1994; Lyubomirsky & Lepper, 1999; McCullough, Emmons, & Tsang, 2002; Modiglianai, 1968; Shiota, Keltner, & John, 2006; Smith, Parrott, Diener, Hoyle, & Kim, 1999; Tangney, Dearing, Wagner, & Gramzow, 2000; Tracy & Robins, 2007). However, these scales are generally not amenable to the measurement of momentary emotions, and, as far as we are aware, self-report scales of momentary distinct emotional states exist for only four emotions: anxiety (Spielberger, Gorssuch, Lushene, Vagg, & Jacobs, 1983), pride (Tracy & Robins, 2007), and shame and guilt (Marschall, Sanftner, & Tangney, 1994). In addition, the Profile of Mood States (POMS; McNair, Lorr, & Droppleman, 1971) is a measure of emotion blends (i.e., anger-hostility, tension-anxiety, depression-dejection) and other affective states (e.g., fatigue-inertia; vigor-activity; confusion-bewilderment).

Our Review

The apparent lack of existing scales to measure momentary distinct emotional states, coupled with the increasing frequency with which researchers seem to be empirically examining these states, raises the question of how researchers *are* assessing distinct emotional experiences in their studies. Indeed, some have called for a critical examination of the ways in which social-personality psychologists assess emotions, especially through self-report (Inzlicht & Tritt, 2014). The purpose of the present manuscript is to provide such an examination, by systematically reviewing a broad and representative sample of empirical articles published in the flagship journal for emotion research, *Emotion*.

In the sections that follow, we address three primary research questions which guided our review; the answer to each of these questions has implications for theoretical and empirical progress in distinct emotion research, and affective science more broadly. First, do currently measured distinct emotions reflect existing emotion taxonomies? Assuming that affective scientists have a sound theoretical understanding of which emotions should be considered distinct, there are important implications for understanding the correspondence (or lack thereof) between this theory and the empirical reality of which emotions are measured as distinct in individual studies. If existing taxonomies of distinct emotions adequately reflect the full range of human affective experience, and existing measurement practices match those theory-based taxonomies, it would suggest that the full range of human emotion is being measured and classified, and that each new empirical finding can be understood in terms of prior theory; this process would promote a cumulative advancement of our understanding of distinct emotions. In contrast, if theoretical accounts of distinct emotions are comprehensive and complete but empirical practice does not match these accounts, then researchers might be measuring more or fewer distinct emotions than exist. This would make it difficult to understand each new empirical finding in terms of existing theory about distinct emotions, leading to a scenario in which individual empirical findings accumulate in a piecemeal fashion, without building off one another to advance distinct emotion theory.

Second, are distinct emotions measured consistently and distinctly across studies? If each distinct emotion is operationalized with convergent sets of self-report items across studies—that is, sets of items that have been shown to capture the same construct through prior scale validation efforts—and these sets of items are largely unique to one emotion, then the emotions measured will have a consistent and distinct meaning across studies. This will facilitate direct comparison and integration of empirical effects across different studies and laboratories, which will allow the field to build a cumulative base of knowledge about the unique causes,

[1] There are, of course, many ways to assess emotion beyond self-report (e.g., facial expressions; physiology, neural activation; Mauss & Robinson, 2009), and each of these assessment channels corresponds to a central component in experts' definition of emotions (e.g., Izard, 2010). While acknowledging these other methods through which researchers can assess distinct emotions, we restrict the scope of this review to studies assessing emotions via self-report.

correlates, and consequences of each distinct emotion. In contrast, if each distinct emotion is operationalized across studies with sets of items that are not known to be convergent (i.e., sets of items that have not been empirically shown to capture the same emotion construct), or if the items used to measure one emotion overlap with those used to measure other emotions, then any given distinct emotion will not have a consensual or unique meaning across studies. As a result, affective scientists will be unable to compare and contrast empirical effects concerning the same emotion, given that differences across studies may result from inconsistent measurement; this state of affairs would hinder the cumulative development of distinct emotion theory.

Third, are currently used self-report scales of distinct emotions of adequate length? There are important implications for understanding whether self-report scales used to measure distinct emotions are long enough to capture the range of components associated with each emotion, and whether these scales are long enough to increase the likelihood that they will show high reliability. If researchers use scales with multiple items, each of these items can capture a different subjective component of emotions (e.g., thoughts, feelings, action tendencies), which together comprise many researchers' overarching definition of emotions (Izard, 2010). In addition, employing longer scales will increase the chances that those scales show high internal consistency reliability, which in turn helps prevent both false negatives (i.e., failing to detect true empirical effects) and false positives (i.e., detecting spurious empirical effects), and enhances researchers' ability to estimate effect sizes. Each of these outcomes will help promote a literature replete with reliable, replicable empirical findings, thereby improving theory development and cumulative science. In contrast, if very short or single-item scales are frequently used, they will often fail to capture the range of subjective components that comprise an emotion, and will often have low reliability, thereby hampering researchers ability to detect empirical effects, and potentially introducing false positives into the literature (Pashler & Harris, 2012).

To answer each of these three overarching questions, we coded 58% of the articles published in *Emotion* between the journal's inception in 2001 through the end of 2011 ($n = 30$ issues, 467 articles; 751 studies), by selecting 2 to 4 issues per year. Specifically, during the years 2001 through 2007, *Emotion* published four issues per year, and we selected two issues to code from each of those years (Issues 1 and 2). During the years 2008 through 2011, the journal published six issues per year, and we selected four issues to code from each of those years (Issues 1, 2, 5, and 6). The specific issues we coded were determined a priori, before any coding had been conducted, and without any knowledge of how many articles in each issue concerned distinct emotions. We treated individual studies as our unit of analysis; for example, if an article reported three studies, we coded each study individually (even if two of the studies were described as 1a and 1b). The first two authors (Aaron C. Weidman & Conor M. Steckler) coded all articles independently, and then met to verify each other's decisions and resolve discrepancies. After these discussions, agreement was reached on all coded items for all articles.

We focused on *Emotion* because it has the highest impact factor of journals primarily dedicated to publishing research on emotion. As a result, it is likely to contain a sample of the most widely read and influential articles in the field, that report research utilizing field-approved measurement practices. Additionally, *Emotion* attracts submissions from researchers interested in the nuanced similarities and differences among distinct emotions, making it an ideal outlet from which to draw a sample of research on momentary distinct emotions.

Measurement Approach of Coded Studies

For each of the 751 studies examined, we coded whether researchers measured momentary distinct emotions with self-report, and each separate instance in which a distinct emotional state was measured with self-report. We refer to this as the *observed measurement approach*, and based this coding entirely on the Method section of each article; a study was coded as measuring momentary distinct emotions if the authors employed a self-report scale to measure a momentary state and labeled this measure with a distinct emotion term. We identified 147 such studies (20% of studies examined), with 356 separate measurement instances (i.e., any time a distinct emotional state was measured within a study).[2] To provide a comparison standard, we also identified each study in which broader dimensions of positive and/or negative affect—independent of, though not mutually exclusive to, distinct emotions—were measured with self-report. We identified 138 such studies (18% of studies examined), suggesting that distinct emotions and positive/negative affect dimensions are measured with similar frequency by emotion researchers. Reflecting the broad interest in momentary distinct emotions across psychological disciplines, the 147 studies that assessed momentary distinct emotions using self-report were lead-authored by researchers with primary areas of affiliation including clinical, cognitive, developmental, health, personality, quantitative, and social psychology, as well as behavioral neuroscience, communications, economics, kinesiology, management, marketing, neuroscience, organizational behavior, and psychiatry.[3]

Theoretical Approach of Coded Studies

It is important to note that the observed measurement approach taken within each study need not reflect the theoretical perspective of the researchers who conducted the study. For example, certain researchers might use a distinct emotion term (e.g., *sadness*) to label their self-report scale out of convenience or simplicity, but not because their goal is to examine or draw conclusions about a distinct emotion. We were interested in both the overall assessment of distinct emotions through self-report scales, and the extent to which such scales are used by researchers who are actually seeking to examine a particular distinct emotion, as opposed to researchers who might use such a scale or label but are in fact more interested in studying a broader affect dimension.

[2] We excluded studies using several practices which could be viewed as instances of distinct-emotion measurement via self-report, but did not fit with our goal of capturing measurement practices for currently experienced distinct emotional states. These included: (a) measures of retrospective or forecasted emotions, (b) non-English scales, and (c) filler items that were not included in the scoring of emotion scales.

[3] Our raw coding data is available at http://ubc-emotionlab.ca/wp-content/uploads/2016/07/Weidman_Steckler_Tracy_Emotion_Study_Codes.xlsx. Syntax used to aggregate these data for the analyses reported in the paper is available on request from the first author.

To determine the extent to which the measurement practices we observed generalize across researchers adopting different theoretical approaches, we also quantified, for each of the 147 studies in which researchers measured one or more distinct emotions with self-report, whether it was conducted from a distinct-emotions perspective, a dimensionalist perspective (e.g., examining emotions at the level of positive and negative affect or valence and arousal; Russell, 1980; Watson & Tellegen, 1985), or a social or psychological constructivist perspective (i.e., treating distinct emotions as cognitively constructed concepts with little or no underlying or cross-culturally shared reality; Barrett, 2006; Lindquist, 2013). We coded articles for evidence of researchers having adopted one of these three approaches, in particular, because they represent the three most prevalent current theoretical orientations toward emotion research (see Barrett, 2014; Russell, 2014; Tracy, 2014).

Specifically, for each study included in our review, we coded the authors' (a) *theoretical approach*, or the extent to which the authors discussed their research rationale in terms of distinct emotions, emotion dimensions, or emotion concepts, in the article's Introductory section, and (b) *intended measurement approach*, or the extent to which the authors described their intended studies and predictions in terms of distinct emotions, emotion dimensions, or emotion concepts; the intended measurement approach was coded from text at the end of the Introductory section, typically in a section entitled The Present Research.

When coding the theoretical approach adopted for each study, we searched for evidence that the authors viewed emotions as distinct states with meaningful differences among them. For example, if authors of a given article noted an interest in distinguishing between the causes or consequences of two or more distinct emotions in their studies (e.g., pride and hope; Cavanaugh et al., 2011), or consistently used one distinct emotion term when discussing prior work or the current studies (e.g., gratitude; DeSteno et al., 2010), we coded these studies' authors as taking a *distinct emotions* theoretical approach. Furthermore, if authors of a given article formulated hypotheses for their studies that incorporated both distinct emotions and affect dimensions (e.g., anger and positive affect; Harmon-Jones et al., 2009), we coded these authors as also taking a *distinct emotions theoretical approach*, because such an approach typically incorporates the understanding that dimensions can be a useful way of categorizing or understanding differences and similarities among distinct emotion states. Although this conceptualization might have resulted in an overestimate of the number of studies authored by researchers who take a distinct-emotions theoretical approach, our review is likely to be relevant to all researchers whose work falls into this category broadly construed, including those who see emotions in terms of both distinct categories and dimensions; as these individuals likely assume that measures purporting to assess distinct emotions in fact capture meaningfully distinct states.

In contrast, we conceptualized the dimensionalist approach more narrowly, as the view that emotions are meaningfully distinct only in terms of continuous dimensions (e.g., pleasantness and arousal). Specifically, for studies with hypotheses that solely involved affect dimensions (e.g., Goldin & Gross, 2010; Storbeck & Clore, 2008), we coded the authors as taking a *dimensionalist theoretical approach*. Finally, if authors described emotions as concepts, or noted the importance of context and/or language in shaping how an emotion is experienced (e.g., Barger et al., 2010; Jakobs et al., 2001), we coded them as taking a *constructivist theoretical approach*.

Importantly, this coding of theoretical approach and intended measurement approach was conducted separately from our primary coding of the *observed measurement approach*. Although all the studies we reviewed measured momentary distinct emotions with self-report (i.e., based on our coding of their Method section), some of these studies were nonetheless authored by researchers who took a *theoretical approach* or *intended measurement approach* in which the stated goal was not to study distinct emotions. (i.e., based on our coding of the Introductory section).

The first author (Aaron C. Weidman) coded the theoretical and intended measurement approach of all studies; to verify these codes, the second author (Conor M. Steckler) coded 42 (29%) of studies. The first and second author showed 95% agreement on theoretical approach ($\kappa = .91$), and 98% agreement on intended measurement approach ($\kappa = .91$), suggesting that our coding results were reliable. We therefore used the first author's codes for all articles. However, to further verify this coding, the third author (Jessica L. Tracy) coded a sample of 21 (14%) studies that the first two authors viewed as particularly ambiguous. Based on the third author's decisions, and the criteria used to make those decisions, we adjusted several of the first author's coding decisions before calculating our final totals in each category.

Our coding of each study's theoretical approach suggested that the majority of researchers who measure momentary distinct emotions with self-report are indeed interested in the causes and consequences of these distinct states. Of the 147 studies included in our review, 108 (73%) studies were authored by researchers who took an explicitly distinct-emotions theoretical approach, and 119 (81%) studies were authored by researchers who took a distinct-emotions intended measurement approach. However, we also found that a minority of researchers who measured momentary distinct emotions with self-report were doing so with the goal of drawing conclusions about broader affect dimensions; 24 (16%) studies were authored by researchers who took a dimensionalist theoretical approach, and 14 (10%) studies were authored by researchers whose intended measurement approach included the assessment of affect dimensions. Additionally, 14 (10%) studies were authored by researchers whose intended measurement approach included a combination of distinct-emotion and dimensionalist assessment.

Only two (1%) of the studies we coded (i.e., studies in which distinct emotions were assessed) were authored by researchers who explicitly took a social constructivist theoretical approach. In addition, none of the studies we coded were authored by researchers whose intended measurement approach included the assessment of emotions from a constructivist approach. This may be attributable to constructivist researchers using a method that we would label as a distinct-emotions measurement approach, thus resulting in potential mis-categorizations. However, given that only 1% of the studies we coded were authored by individuals who appeared to hold a constructivist theoretical approach, based on their description of their research goals, theory, and hypotheses in the Introduction, it is unlikely that more than a small portion of studies we coded were conducted with the goal of measuring emotion concepts as opposed to distinct emotions.

Finally, we were unable to characterize the theoretical approach of 13 (9%) studies' authors; these were studies appearing in articles that, in most cases, used dimensionalist language in the Introductory section, but also referred to emotions like *happiness* and *sadness*, leaving open the question of whether the authors conceptualized happiness and sadness as distinct emotions or as convenient labels through which to study positive and negative affect.

Given that, for the majority of studies coded, researchers explicitly adopted a distinct-emotions theoretical approach and intended measurement approach to their research, in the following sections we focus on the implications of our findings for those researchers whose goal was to learn about the causes and consequences of distinct emotions. However, because a minority of researchers explicitly took a dimensionalist approach, in the final section of this article we discuss implications of the measurement practices observed in our review for those researchers who do not adopt a distinct-emotions perspective.

Question 1: Do Currently Measured Distinct Emotions Reflect Existing Theoretical Emotion Taxonomies?

We first sought to determine whether the set of emotions currently measured through self-report maps on to the set of emotions that have been conceptualized in prior theoretical taxonomies. Researchers have developed emotion taxonomies using a number of different approaches; although debate remains about which approach is likely to be best, we can seek convergence across these various methods to estimate the number of distinct emotions that would be predicted to emerge in the literature. First, based on sources of primarily nonself report evidence, researchers have identified between 6 and 10 distinct emotional states. For example, studies of nonverbal emotion expressions have identified 10 emotions that are reliably associated with distinct, cross-culturally recognized expressions (anger, contempt, disgust, embarrassment, fear, happiness, pride, sadness, shame, and surprise; Ekman & Friesen, 1971; Keltner, 1995; Tracy & Robins, 2008). Additionally, cross-species neuroscientific research has identified seven distinct emotional systems in the brain (CARE, FEAR, LUST, PANIC, PLAY, RAGE, SEEKING; Panksepp, 2007), and, in humans, neuroimaging research has pointed to four distinct patterns of brain activity corresponding to anger, fear, happiness, and sadness (Damasio et al., 2000).

However, researchers using self-report methods also have identified a number of distinct emotions that serve important social functions, or are likely to have served important functions in humans' evolutionary history, but are not known to be associated with distinct nonverbal expressions (e.g., gratitude, jealousy; Bartlett & DeSteno, 2006; Levy & Kelly, 2010) or brain activity (e.g., compassion, pride; Goetz, Keltner, & Simon-Thomas, 2010; Tracy & Robins, 2007). As a result, it is likely that the list of distinct emotions that are measurable through self-report outnumbers that identified through the non-self-report evidence that has accumulated thus far. Turning then to taxonomies developed through self-report methods, studies have identified 25 lower-order clusters of distinct states (including, e.g., jealousy and pride), which fall within 6 higher-order categories (anger, fear, joy, love, sadness, and surprise; Shaver, Schwartz, Kirson, & O'Connor, 1987).

In the face of this prior evidence—from both non-self-report and self-report sources—if theory and measurement practices align, we might expect the number and range of momentary distinct emotions regularly measured with self-report to approximately reflect these findings; as a result, we expected to see somewhere between 6 and 31 (6 higher-order clusters + 25 lower-order clusters) distinct states. However, in the absence of existing scales designed to reflect distinct emotions identified through prior taxonomies, researchers may regularly assess more emotions than prior taxonomies have included, for two reasons. First, researchers may create new scales impromptu for each study, without first examining whether the state they wish to measure is in fact a distinct emotion, based on an existing taxonomy. Second, researchers may apply new or inconsistent labels to self-report scales across studies, even if these scales contain similar items and likely measure the same emotion. The result of these two practices could be that researchers assess purportedly distinct emotions which are in fact slight variants of previously identified emotions, while treating them as distinct entities. This in turn could lead to a mismatch between theoretical accounts of distinct emotions, and the number of emotions that appear in the empirical literature.

To determine the number and range of distinct emotions currently measured in the empirical literature, we coded the distinct emotion that was measured at each observed measurement instance in the articles we reviewed. Importantly, emotions were coded exactly as they were conceptualized by the study's authors, regardless of the specific items used to measure them. For example, if three different studies used the item *happy* to measure *happiness*, *joy*, and *contentment*, respectively, we would code these three studies as measuring *happiness*, *joy*, and *contentment*, respectively, despite the fact that they used the same item. We identified 65 different emotions that were measured, each on average 5.48 times (SD = 10.04; Median: 2; Range: 1–49). Five emotions were measured on more than 10 occasions; these included anxiety, which was the most frequently measured emotion (n = 49 instances; 14% of total), and four emotions typically considered to fall within the class of basic emotions: anger, fear, happiness, and sadness (n = 142; 40%). Fifteen additional emotions were measured on 4 to 10 occasions (n = 89; 25%); these included disgust, guilt, joy, love, pride, *schadenfreude*, and awe, among others. Finally, 46 additional emotions were measured on three or fewer occasions (n = 76, 21%); these included anticipatory enthusiasm, astonishment, jealousy, nurturant love, symhedonia, and tension, among others (see Table 1 for full list).

Implication: Mismatch Between Theory and Measurement

The results of our quantitative review indicate that there is a mismatch between existing theoretical taxonomies of distinct emotions, and the distinct emotions currently measured in the empirical literature. Whereas so-called basic emotions continue to drive the bulk of distinct emotion research using self-report methods, a wide range of other emotions have entered the fray in recent years, creating a literature in which many emotions that have not been included in a prior theoretical taxonomies routinely appear. In light of the convergence across studies of nonverbal expressions and affective neuroscience pointing to the existence of 6 to 10 distinct basic-level emotions, and taxonomies developed through self-

Table 1
Frequency With Which Each Distinct Emotion Was Measured in Our Review

Emotion	Number of measurement occasions	Percentage of total measurement occasions	Number of distinct scales
Anxiety	49	13.76	10
Sadness	45	12.64	12
Anger	37	10.39	19
Happiness	36	10.11	10
Fear	24	6.74	9
Amusement	10	2.81	1
Disgust	10	2.81	4
Guilt	8	2.25	3
Joy	6	1.69	5
Love	6	1.69	4
Pride	6	1.69	1
Schadenfreude	6	1.69	4
Sympathy	6	1.69	5
Elation	5	1.40	2
Shame	5	1.40	3
Surprise	5	1.40	2
Awe	4	1.12	1
Calmness	4	1.12	2
Gratitude	4	1.12	2
Hope	4	1.12	1
Anticipatory enthusiasm	3	.84	2
Contentment	3	.84	1
Dejection	3	.84	2
Empathy	3	.84	3
Interest	3	.84	3
Jealousy	3	.84	1
Nurturant love	3	.84	2
Regret	3	.84	2
Tenderness	3	.84	1
Attachment love	2	.56	2
Compassion	2	.56	2
Contempt	2	.56	1
Depression	2	.56	2
Embarrassment	2	.56	1
Entertainment	2	.56	1
Excitement	2	.56	1
Frustration	2	.56	2
Hostility	2	.56	2
Irritation	2	.56	1
Symhedonia	2	.56	2
Tension	2	.56	2
Uneasiness	2	.56	1
Antagonism	1	.28	1
Astonishment	1	.28	1
Aversion	1	.28	1
Boredom	1	.28	1
Concern	1	.28	1
Confusion	1	.28	1
Desire	1	.28	1
Disappointment	1	.28	1
Discomfort	1	.28	1
Distress	1	.28	1
Enjoyment	1	.28	1
Envy	1	.28	1
Fatigue	1	.28	1
Inspiration	1	.28	1
Longing	1	.28	1
Melancholy	1	.28	1
Nervousness	1	.28	1
Nostalgia	1	.28	1
Pleasant relaxation	1	.28	1
Relief	1	.28	1
Shyness	1	.28	1
Touched	1	.28	1
Vigor	1	.28	1
Total 65	356	100	160

report methods pointing to approximately 31 higher-order and lower-order distinct emotion clusters, it may seem surprising that the list of distinct emotions routinely assessed through self-report stretches to at least as long as 65.[4] However, there are several possible explanations for this apparent discrepancy.

First, the mismatch may stem in part from limitations of existing taxonomies themselves, which may not capture the full range of distinct emotional experience. Each of the taxonomies reviewed above was developed on the basis of a specific source of evidence (e.g., distinct facial expressions, distinct neural signals; Ekman & Friesen, 1971; Panksepp, 2007), yet emotions are typically defined by a number of separate experiential criteria (Izard, 2010). As a result, taxonomies based on additional sources of evidence (e.g., taxonomies that account for emotions with distinct facial expressions *or* distinct vocal expressions *or* distinct physiology) may point to the existence of a greater number of distinct emotions than previously developed taxonomies which tend to focus on one source of evidence only.

Second, there may not be a one-to-one correspondence between terms used to measure distinct emotions in the current literature, and distinct emotions as theoretically conceptualized by researchers. For example, several terms used to measure distinct emotions may in fact refer to context-specific variants of other distinct emotions, which fall within prior taxonomies. Consider the emotion *schadenfreude*—observed on six measurement occasions in our review—which is typically defined as pleasure arising from the misfortune of others (Smith, Powell, Combs, & Schurtz, 2009). On one hand, perhaps *schadenfreude* is a distinct emotion in the same manner as happiness or anger, and should therefore be included in any taxonomy of distinct emotions. On the other hand, perhaps *schadenfreude* is best conceptualized as a form of *happiness*, which arises in a specific context (i.e., when observing someone else's misfortune). In that case, we would not expect to see *schadenfreude* emerge as a distinct emotion in a theoretical taxonomy. Similarly, several terms used to measure distinct emotions may in fact refer to specific components of distinct emotions that were identified in prior theoretical taxonomies. For example, consider *tension*, observed on two measurement occasions in our review. Tension may be a distinct emotion, but it may also be a physiological response associated with fear or anxiety—emotions that have been included in prior taxonomies (e.g., Ekman & Friesen, 1971; Shaver et al., 1987).

Third, several terms used to measure distinct emotions in the current literature may in fact refer to emotional states that are best conceptualized as falling within an *emotion family*, or broader category that encompasses several states that share a common set of emotional components (e.g., facial expression, physiology, subjective feeling; Ekman, 1992). For example, we observed that different researchers used each of the words *sadness*, *dejection*, *depression*, *disappointment*, and *melancholy* to refer to distinct emotions, across empirical studies. It is possible that the latter four of these states are all slight variants within the emotion family sadness, rather than each constituting a distinct emotion, particularly in light of the fact that lay persons have been shown to view the higher-order category of sadness as encompassing all of the latter four states (Shaver et al., 1987). In that case, prior theoretical taxonomies that include sadness (e.g., Ekman & Friesen, 1971; Shaver et al., 1987) will have accounted for the emergence of more narrow feeling states such as dejection, depression, disappointment, and melancholy.[5]

Fourth, it is possible that researchers have simply not arrived at consistent terminology with which to describe distinct emotional states that have been identified in prior theoretical taxonomies. For example, although *happiness* was one of the most frequently observed distinct emotions measured in our review ($n = 36$; 10%), we also identified several emotion terms that could reasonably be viewed as synonyms for happiness, such as joy ($n = 6$) or contentment ($n = 3$). Researchers using the words *happiness*, *joy* and *contentment* in their studies may not conceptualize each of these states as distinct, but rather may be using different words to refer to the same emotional experience. Indeed, even within theoretical taxonomies of distinct emotions, the terms *happiness* and *joy* have been used to refer to what appears to be the same experience (e.g., Ekman & Friesen, 1971; Shaver et al., 1987). Additionally, some affective scientists may prefer to use the terms joy and contentment when discussing distinct emotions—rather than happiness—to distinguish their line of inquiry from the considerable empirical literature on the causes and consequences of "happiness," conceptualized more broadly as a blend of life satisfaction and pleasant (vs. unpleasant) affect (Busseri & Sadava, 2011; Diener, 1984, 2000; Dunn & Norton, 2013; Kahneman, 1999; Larsen, 2000).

Regardless of the reasons for the gap between existing theoretical taxonomies and empirical measurement practices, if the emotional landscape is in fact best characterized as containing 65 or more distinct states, then researchers might benefit from constructing theoretical frameworks to account for how each of these distinct states may arise, and how each differs from each other, despite the fact that some may seem similar at face value. How could theory and measurement become more in line? If researchers wish to study a distinct emotion, they could first examine whether it is captured by a prior taxonomy, and if not, present a theoretical argument for how and why that emotion might be considered distinct from related emotions or broader emotion families that have already been examined. In addition to articulating a theoretical account for why an emotion of interest might be distinct, researchers could also articulate how they are measuring that emotion in a way that differentiates it from measures of closely related constructs. For example, returning to the example of *sadness*, *dejection*, *depression*, *distress*, *disappointment*, and *melancholy*, researchers who wish to study one of these states could explain why their state of interest differs from the broader emotion family (i.e., sadness) and perhaps from related subcategories as well, and also ensure that their measure is distinct from existing measures of the related states.

[4] Importantly, given that we coded only a subset of articles published in *Emotion* from 2001 through 2011, it is likely that we did not code all distinct emotions that have been studied over the course of that time, suggesting that the list of distinct emotions studied in the current literature may be even longer than 65.

[5] We acknowledge that the term *depression* is often used to refer more broadly to a clinical syndrome—and not merely a distinct emotion. However, in the studies we coded, it was treated as a distinct emotion.

What Are the Advantages and Disadvantages of Having a Consensus Taxonomy in a Research Area?

The development of a consensual taxonomy of distinct emotions, and of sets of measures that validly capture the emotions included in this taxonomy, could benefit affective scientists. History suggests that a field in which theoretical frameworks and empirical measurement strategies do not align can produce a scattered body of data that impedes empirical progress. For example, in the 1950s and 1960s, the field of personality psychology experienced just such a crisis, as researchers regularly measured and studied a wide array of personality constructs in isolation, without constructing a unified theoretical framework of the person (Adelson, 1969; McAdams, 1997); this in turn drew critical appraisals (e.g., Mischel, 1968). Our review suggests that the field of distinct emotion research may currently face the same problem of numerous, scattered constructs, given the present misalignment between theory and measurement.

The mismatch between theoretical taxonomies and empirical measurement practices in distinct emotion research is likely exacerbated by the lack of existing scales for measuring these many purportedly distinct states. One driving force behind the renaissance of personality psychology, following its mid-20th century crisis, was the adoption of the Big Five framework as a unifying model, and the subsequent development of self-report scales through which to assess these five core personality traits (e.g., Costa & McCrae, 1992; Goldberg, 1992; John & Srivastava, 1999). Similar efforts have proven fruitful in other areas of psychology too; for example, cognitive psychologists have begun to develop The Cognitive Atlas, a database that catalogues a taxonomy of mental functions (e.g., working memory), interrelations among these functions, and specific measurement tools and tasks that are known to index each (Poldrack, 2010). In contrast, emotion researchers have neither constructed a comprehensive taxonomy of distinct emotions, nor developed scales to measure the states contained in such a taxonomy. Emotion researchers might therefore benefit from following the model set forth by personality psychologists and cognitive psychologists, and seeking to develop a comprehensive taxonomy of distinct emotions and systematically construct self-report scales aimed to measure them.[6]

Importantly, developing a taxonomy of distinct emotions could spark novel theoretical advancements, including the discovery of new distinct emotions, as researchers could begin to explicitly conceptualize additional emotional states in juxtaposition to those contained in existing taxonomies. Similar scenarios have played out in personality psychology since the adoption of the Big Five. For example, personality researchers have identified Big Five Facets as more narrow personality dimensions that fall below Big Five traits within the hierarchical structure of that taxonomy (e.g., Costa & McCrae, 1992; DeYoung, Quilty, & Peterson, 2007), and have argued that these narrow facets (vs. broad traits) have predictive value when examining narrow (vs. broad) outcomes (e.g., Paunonen, 1998; Paunonen & Ashton, 2013). Similarly, personality researchers have pinpointed gaps in the Big Five's coverage of human personality and developed theories about traits that lie outside the Big Five, such as the Dark Triad (e.g., Paulhus & Williams, 2002), Honest-Humility (e.g., Ashton, Lee, & DeVries, 2014), and motives such as Need for Achievement (e.g., McClelland, 1987; Winter, John, Stewart, Klohnen, & Duncan, 1998). To be sure, only time will tell if a comprehensive distinct emotions taxonomy would be beneficial; nonetheless, historical and contemporary parallels between personality psychology and affective science point to the likelihood of major advances for the field if a taxonomy is developed.

Question 2: Are Distinct Emotions Measured Consistently and Distinctly Across Studies?

We next sought to examine whether distinct emotions are measured with convergent sets of items across studies, and, in each case, a set that is largely exclusive to the emotion in question. Distinct emotions are constructs that generally involve consistent and specific patterns of subjective feelings, physiological changes, neural activity, cognitive appraisals, and motivated action tendencies (Ekman, 1992; Kragel & LaBar, 2014; Roseman, 2011; see Tracy & Randles, 2011). As a result, across studies, a given distinct emotion would be expected to be measured with sets of self-report items which are largely exclusive to that distinct emotion, unless empirical or theoretical work suggests overlap in the subjective components of multiple emotions. However, given the apparent absence of a systematically developed set of self-report scales to measure emotions, it is possible that researchers may not regularly measure emotions in a convergent and distinct manner across studies.

There are two reasons for this expectation. First, the absence of systematically developed scales increases the chance that a single emotion may be measured with a different set of items across studies, yet labeled as the same emotion, creating the misleading impression that the two measures capture identical constructs (i.e., the *Jingle Fallacy*; Thorndike, 1904). This potential problem is likely to occur in the absence of systematically developed scales because researchers are more likely to create scales impromptu each time they wish to measure a given emotion, and, across multiple studies and laboratories, these many impromptu scales will not necessarily contain converging sets of items. As a result, these different scales may in fact measure different psychological states. The result would be a set of scales thought to measure the same emotion, but which show an inconsistent pattern of relations with other variables across studies, creating the spurious impression that a single distinct emotion is inconsistently related to other variables, when in fact the inconsistency arises purely as a result of inconsistent measurement.

Second, the absence of systematically developed scales increases the chance that multiple distinct emotions are measured with a similar set of items across studies, creating the misleading impression that these scales measure distinct emotional states, when in fact they capture the same or closely related states (i.e., *the Jangle Fallacy*; Kelley, 1927). This potential problem is likely to occur in absence of developed scales because, if researchers create scales impromptu to measure distinct emotions, then across studies they are likely to use slightly different labels to refer to these scales, despite their similar content (e.g., dejection vs. depression); this would likely lead to the use of the same scale items to measure purportedly distinct emotions. The result would be findings sug-

[6] It should be noted that the development of self-report measures of momentary distinct emotions will involve somewhat different procedures than the development of measures of stable personality dispositions (e.g., discriminant validity may be of primary importance when distinguishing closely related distinct emotions; test–retest reliability is only useful for personality measures [McCrae et al., 2011]).

gesting that multiple distinct emotions have similar empirical correlates, creating the spurious impression that they are not distinct, even though this similarity arises as a result of unintended overlap in how the two emotions are measured.

To examine the extent to which these problematic trends are likely to occur, we first examined whether distinct emotions were measured consistently across studies. To do so, we categorized the scales we coded into one of four categories based on how the scale was developed. Three categories constitute scales that are likely to lead to inconsistent measurement across studies: (a) *impromptu*, or scales that were developed for a given measurement instance with no reference to prior research and without a systematic scale development process[7]; (b) *cited impromptu*, or scales explicitly taken from a previous study in which they had been developed impromptu; and (c) *cited existing altered*, or scales that had been systematically developed in previous research but had been altered for the present measurement instance. In contrast, one category constituted scales that were likely to lead to consistent measurement across studies: (d) *cited existing unaltered*, or scales that had been systematically developed in previous research and had not been altered for the present measurement instance. To distinguish between *cited existing altered* and *cited existing unaltered*, for each measurement instance in which a scale was accompanied by a citation to a previous article, we examined the cited article to determine whether the previously developed scale was altered for the present measurement. However, in cases where authors explicitly reported altering a previously developed scale, we simply coded that usage as *cited existing altered* without checking the prior cited article.

We found that the majority of scales we coded were used in ways that are likely to promote inconsistent measurement across studies. As is shown in Figure 1, among the 356 measurement instances coded in our review, scales for a total of 246 (69%) measurement instances were developed *impromptu*, scales for 27 (7.6%) measurement instances were *cited impromptu*, and scales for 30 (8.4%) measurement instances were *cited existing altered*. The latter included cases where, to use a hypothetical example, a researcher used only three of the five items included in the guilt subscale of the State Shame and Guilt Scale (SSGS; Marschall et al., 1994) to measure state guilt, and cited the SSGS. In contrast, scales for 43 (12%) measurement instances were *cited existing unaltered*. Scales were not reported for 10 (3%) of measurement instances. In sum, these results indicate that scales for a full 85% of measurement instances were likely to contribute to inconsistent measurement of distinct emotions across studies.[8]

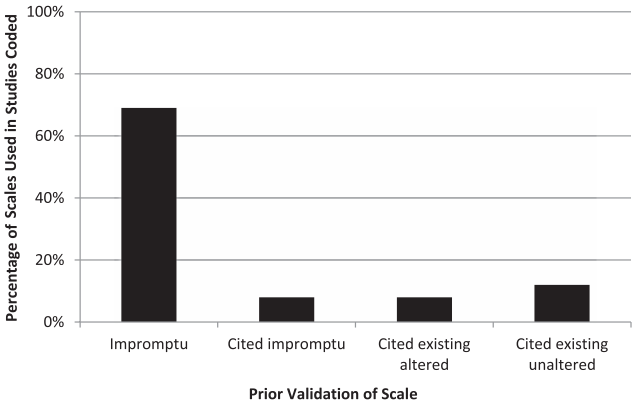

Figure 1. Prior development of scales used to measure distinct emotions. Impromptu = scale was developed in an impromptu fashion, with no reference to prior research. Cited impromptu = scale was explicitly taken from a previous study in which it had been developed in an impromptu fashion. Cited existing altered = scale was systematically developed in previous research but altered for the present measurement instance. Cited existing unaltered = scale was systematically developed in previous research and was not altered for the present measurement instance.

Next, to determine whether emotions were measured in ways that captured their distinctness, we coded the specific items comprising each self-report scale. We coded both single words (e.g., *anxious*, *happy*) and short phrases (e.g., *I would like to be in the shoes of* [someone]; *I violated a norm*). Each word or phrase was coded separately for each distinct emotion it was used to measure (e.g., *anxious* might be coded as used to measure both anxiety and fear). We ignored part of speech when coding words (e.g., *amused* and *amusement* would be coded as one item). The number of words or phrases that were used to measure each emotion varied widely ($M = 4.23$; $SD = 6.76$; Median: 2; Range: 0–45).[9] Table 2 presents each word or phrase that was used as a scale item to measure a distinct emotion, and, for each scale item, other distinct emotions that were also measured with that item (henceforth "overlapping emotions"). For example, the first entry in Table 2 indicates that the emotion *amusement* was measured with the word *amused/amusement*, and that this item was also used, in at least one other study, to measure the emotion *happiness*. In total, 178 distinct words and phrases were used in 160 distinct scales (i.e., unique combinations of items).[10]

[7] Drawing on prior research, we defined systematic scale development as involving the following steps: (a) drawing on prior theoretical accounts of a distinct emotion of interest, (b) assessing lay knowledge of a distinct emotion, and/or (c) using factor analytic methods to arrive at a set of items that comprehensively captures the content of the emotion (see Clark & Watson, 1995 and Reise, Waller, & Comrey, 2000, for further discussion).

[8] Notably, among the 73 instances in which a systematically developed scale was used (either altered or unaltered), relatively few scales were used repeatedly. Only four systematically developed measures were used in more than one article: the STAI-S ($n = 28$ measurement instances; Spielberger et al., 1983), the Profile of Mood States ($n = 8$; POMS; McNair et al., 1971), the Worry-Emotionality Questionnaire (Morris, Davis, & Hutchings, 1981; $n = 3$), and the Positive and Negative Affect Schedule ($n = 19$; PANAS; Watson et al., 1988). The STAI-S was never reported to be altered, the POMS was altered in 2 instances, and the PANAS was altered in 16 instances. Although the PANAS was typically used to measure a distinct emotion with one of its subscales (e.g., fear, hostility), we also identified multiple examples in which the PANAS was used in ways other than originally intended (i.e., to measure positive and negative affect dimensions), such as administering the entire PANAS but using scores from only a few selected items to measure a distinct emotion.

[9] We only recorded scale items that were either reported in the text, or were available in the literature (i.e., not part of a proprietary scale). If an emotion is recorded as being measured with 0 scale items, it therefore means that none of the items used to measure that emotion were available to readers of the paper.

[10] In calculating the number of distinct scales, all measurement instances for a given emotion in which items were not reported were counted as one scale, given that each unreported set of items could conceivably have been identical. Given the likelihood that at least some of these scales in fact used different sets of items, the total number of scales reported represents a conservative estimate.

Table 2
Words and Phrases Used to Measure Distinct Emotions

Emotion measured	Words used to measure emotion	Overlapping emotions
amusement	amused	happiness
anger	I feel like swearing	[none]
	aggressive	[none]
	agitated	[none]
	angry	antagonism
		disgust
		hostility
	annoyed	frustration
	disgusted	antagonism
		disgust
		hostility
	feel like hitting someone	[none]
	frustrated	frustration
	furious	[none]
	hostile	anxiety
		disgust
		hostility
	irritated	antagonism
		anxiety
		disgust
		hostility
		irritation
	mad	[none]
	peeved	[none]
	rage	[none]
	scornful	disgust
		hostility
	sore	[none]
	want to get back at someone	[none]
	want to lash out	[none]
	want to overcome some obstacle	[none]
	want to strike out at someone	[none]
antagonism	*angry*	anger
		disgust
		hostility
	disgusted	anger
		disgust
		hostility
	irritated	anger
		anxiety
		disgust
		hostility
		irritation
anticipatory enthusiasm	*enthusiastic*	elation
		happiness
		joy
	excited	elation
		excitement
		interest
anxiety	I am so tense that my stomach is upset	[none]
	I do not feel very confident	[none]
	I feel I may not do as well as I could	[none]
	I feel my heart beating fast	[none]
	I feel that others will be disappointed in me	[none]
	afraid	fear
	anxious	fear
	arms and legs feel stiff	[none]
	ashamed	guilt
		sadness
		shame
	avoid uncomfortable thoughts	[none]
	breathing is fast and shallow	[none]
	butterflies in the stomach	[none]
	can't get thoughts out of mind	[none]
anxiety (continued)	can't make up my mind	[none]

Table 2 (*continued*)

Emotion measured	Words used to measure emotion	Overlapping emotions
	cannot control thoughts	[none]
	distressed	distress
	dizzy	[none]
	face feels hot	[none]
	feel agonized over problem	[none]
	guilty	guilt
		sadness
		shame
	heart beats fast	[none]
	hostile	anger
		disgust
		hostility
	irrelevant thoughts intruding	[none]
	irritated	anger
		antagonism
		disgust
		hostility
		irritation
	jittery	fear
	muscles are tense	[none]
	muscles feel weak	[none]
	nervous	distress
		fear
		nervousness
	palms feel clammy	[none]
	panicky	fear
	picture future misfortunes	[none]
	regretful	dejection
		regret
		shame
	relaxed[a]	pleasant relaxation
	scared	fear
	self-conscious	[none]
	shaky	fear
	tense	tension
	think others won't approve	[none]
	think worst will happen	[none]
	throat feels dry	[none]
	trembly	[none]
	trouble remembering things	[none]
	uneasy	uneasiness
	upset	distress
	worried	fear
astonishment	surprise	surprise
attachment love	attachment	[none]
	love	love
aversion	aversion	[none]
	repugnance	[none]
awe	awe	[none]
boredom	boring	interest
calmness	calm	happiness
		pleasant relaxation
compassion	*compassionate*	empathy
		nurturant love
		sadness
		sympathy
	pity	[none]
	sympathetic	empathy
		love
		sympathy
concerned	concern	empathy
confusion	[none]	—
contempt	contempt	hostility
contentment	content	happiness
contentment (continued)		joy
dejection	*blue*	sadness

(*table continues*)

Table 2 (*continued*)

Emotion measured	Words used to measure emotion	Overlapping emotions
	disappointed	disappointment
	discouraged	frustration
	low	[none]
	regretful	anxiety
		regret
		shame
	sad	happiness
		sadness
		sympathy
	self-pity	[none]
depression	*depressed*	sadness
	dull	[none]
	Tired	[none]
desire	desire	[none]
disappointment	*disappointed*	dejection
discomfort	discomfort	[none]
disgust	*angry*	anger
		antagonism
		hostility
	disgusted	anger
		antagonism
		hostility
	feel like throwing up	[none]
	grossed-out	[none]
	hostile	anger
		anxiety
		hostility
	irritated	anger
		antagonism
		anxiety
		hostility
		irritation
	loathing	hostility
	repulsed	[none]
	scornful	anger
		hostility
	sickened	[none]
	turn away from something or someone	[none]
	want to avoid something	[none]
	want to get rid of something	[none]
	want to move away from something	[none]
distress	*distressed*	anxiety
	nervous	anxiety
		fear
		nervousness
	stressed	[none]
	upset	anxiety
elation	elation	[none]
	enthusiastic	anticipatory enthusiasm
		happiness
		joy
	excited	anticipatory enthusiasm
		excitement
		interest
	happy	happiness
		joy
		sadness
		schadenfreude
		symhedonia
embarrassment	*embarrassed*	shame
empathy	*compassionate*	compassion
		nurturant love

Table 2 (*continued*)

Emotion measured	Words used to measure emotion	Overlapping emotions
		sadness
		sympathy
	concerned	empathy
	empathetic	[none]
	moved	sadness
		sympathy
	soft-hearted	sympathy
	sympathetic	compassion
		love
		sympathy
	tender	nurturant love
		sympathy
		tenderness
enjoyment	enjoyment	[none]
entertainment	entertained	[none]
envy	I feel less good when I compare my own results with those of . . .	[none]
	I would like to be in the position of . . .	[none]
	I would like to be in the shoes of . . .	[none]
	jealous	jealousy
excitement	excited	anticipatory enthusiasm
		elation
		interest
fatigue	[none]	—
fear	*afraid*	anxiety
	anxious	anxiety
	frightened	nervousness
	jittery	anxiety
	nervous	anxiety
		distress
		nervousness
	panicky	anxiety
	scared	anxiety
	shaky	anxiety
	timid	[none]
	worried	anxiety
frustration	annoyed	anger
	discouraged	dejection
	frustrated	anger
gratitude	appreciative	[none]
	grateful	[none]
	positive	symhedonia
guilt	a bad person	[none]
	apologize	[none]
	ashamed	anxiety
		sadness
		shame
	be forgiven	[none]
	guilty	anxiety
		sadness
		shame
	violated a norm	[none]
happiness	*amused*	amusement
	bad[a]	[none]
	calm	calmness
		pleasant relaxation
	cheerful	[none]
	content	contentment
		joy
	delighted	joy
	determined	[none]
	enthusiastic	anticipatory enthusiasm
		elation
		joy
happiness (continued)	gay	[none]
	glad	joy
	good	[none]
	happy	elation

(*table continues*)

Table 2 (*continued*)

Emotion measured	Words used to measure emotion	Overlapping emotions
		joy
		sadness
		schadenfreude
		symhedonia
	inspired	inspiration
	joyful	joy
	pleased	[none]
	proud	proud
	sad[a]	dejection
		sadness
		sympathy
	satisfied	joy
		schadenfreude
	tranquil	[none]
	well	[none]
hope	hope	[none]
hostility	I dislike	[none]
	angry	anger
		antagonism
		disgust
	contempt	contempt
	disgusted	anger
		antagonism
		disgust
	hate	[none]
	hostile	anger
		anxiety
		disgust
	irritated	anger
		antagonism
		disgust
		irritation
	loathing	disgust
	scornful	anger
		disgust
inspiration	*inspired*	happiness
interest	*boring*[a]	boredom
	curious	[none]
	excited	anticipatory enthusiasm
		elation
		excitement
	interested	[none]
irritation	*irritated*	anger
		antagonism
		anxiety
		disgust
		hostility
jealousy	*jealous*	envy
joy	*content*	contentment
		happiness
	delighted	happiness
	enthusiastic	anticipatory enthusiasm
		elation
		happiness
	glad	happiness
	happy	elation
		happiness
		sadness
		schadenfreude
		symhedonia
	joyful	happiness
joy (continued)	lively	[none]
	satisfied	happiness
		schadenfreude
longing	longing	[none]
love	I feel I can confide in [my dating partner] about anything	[none]

Table 2 (*continued*)

Emotion measured	Words used to measure emotion	Overlapping emotions
	I would do almost anything for [my dating partner]	[none]
	If I were lonely my first thought would be to seek [my dating partner] out	[none]
	affection	[none]
	caring	[none]
	fondness	[none]
	love	attachment love
	sympathetic	compassion
		empathy
		sympathy
melancholy	melancholy	[none]
nervousness	*frightened*	fear
	nervous	anxiety
		distress
		fear
nostalgia	nostalgia	[none]
nurturant love	*compassionate*	compassion
		empathy
		sadness
		sympathy
	nurturance	[none]
	tender	empathy
		sympathy
		tenderness
pleasant relaxation	*calm*	calmness
		happiness
	relaxed	anxiety
pride	*proud*	happiness
regret	kicking self	[none]
	missed an opportunity	[none]
	regretful	anxiety
		dejection
		shame
	should have known better	[none]
	undo what had happened	[none]
relief	*relieved*	schadenfreude
sadness	*ashamed*	anxiety
		guilt
		shame
	blue	dejection
	compassionate	compassion
		empathy
		nurturant love
		sympathy
	dejected	[none]
	depressed	depression
	down	[none]
	gloomy	[none]
	guilty	anxiety
		guilt
		shame
	happy[a]	elation
		happiness
		joy
		schadenfreude
		symhedonia
	hopeless	[none]
	lonely	[none]
	miserable	[none]
	moved	empathy
		sympathy
	sad	dejection
sadness (continued)		happiness
		sympathy
	sorrow	[none]
	unhappy	[none]
schadenfreude	actually I had to laugh a little	[none]
	I couldn't resist to smile a little	[none]

(*table continues*)

Table 2 (*continued*)

Emotion measured	Words used to measure emotion	Overlapping emotions
	I like what happened to . . .	[none]
	happy	elation
		happiness
		joy
		sadness
		symhedonia
	relieved	relief
	satisfied	happiness
		joy
	schadenfreude	[none]
shame	*ashamed*	anxiety
		guilt
		sadness
	embarrassed	embarrassment
	foolish	[none]
	guilty	anxiety
		guilt
		sadness
	regretful	anxiety
		dejection
		regret
	ridiculed	[none]
shyness	shyness	[none]
surprise	astonishment	[none]
	surprise	astonishment
symhedonia	*happy*	elation
		happiness
		joy
		sadness
		schadenfreude
	positive	gratitude
sympathy	I commiserate with [target] about what happened	[none]
	compassionate	compassion
		empathy
		nurturant love
		sadness
	moved	empathy
		sadness
	negative	[none]
	sad	dejection
		happiness
		sadness
	soft-hearted	empathy
	sympathetic	compassion
		empathy
		love
	tender	empathy
		nurturant love
		tenderness
tenderness	*tender*	empathy
		nurturant love
		sympathy
tension	*tense*	anxiety
touched	touched	[none]
uneasiness	*uneasy*	anxiety
vigor	[none]	—

Note. Overlapping emotions: Additional emotions for which the word or phrase was used in a self-report scale. [none] in "Words used to measure emotion" indicates that none of the scale items used to measure the emotion in question were available to readers (i.e., they were part of a proprietary scale, and not reported in the manuscript). *Italicized* words or phrases are those used to measure at least one overlapping emotion.
[a] Word or phrase was used to measure the absence of a given emotion.

At the level of individual emotions, 51 of the 65 (78%) emotions measured were assessed with at least one word or phrase that was also used to measure an overlapping emotion; these 51 emotions were each measured with a set of words that were used to measure an average of 4.96 distinct emotions, across studies (Median = 4, *SD* = 3.55, Range: 2–17). At the level of individual items, 55 of the 125 (44%) words used to measure emotions were used to measure more than one emotion; these 55 words were each used to

measure an average of 2.67 emotions, across studies (Median = 2, *SD* = 1.06, Range: 2–6). In contrast, of the 53 short phrases used, none were used to measure an overlapping emotion.

Of note, inconsistent scale use—which presumably resulted from a lack of existing scales to measure emotions—exacerbated the problem of overlap among items used to measure distinct emotions. Emotions that were measured with a greater number of unique scales also tended to be measured with a larger number of different words and phrases (i.e., the correlation between the number of scales used to measure a given emotion, and the number of distinct words/phrases used to measure that emotion, was $r = .78$, $p < .001$). Emotions that were measured with a greater number of words or phrases, in turn, tended to be measured with items that were also used to measure a greater number of other emotions (i.e., the correlation between the number of words/phrases used to measure an emotion, and the number of other emotions that were measured with the same set of words or phrases, was $r = .80$, $p < .001$). These results suggest that the use of impromptu scales is directly associated with the substantial item-level overlap among emotion scales observed in our review.

Whereas the lack of existing scales appears to have exacerbated overlap among items used to measure distinct emotions, our review suggests that the existence of systematically developed scales could help reduce the overlap in items used to measure distinct emotions. For example, the emotion *anxiety* appeared more times than *anger* in the articles we reviewed (*n*s = 49 and 37, respectively), yet was assessed with approximately half as many distinct scales as anger (*n*s = 10 and 19, respectively). This is likely because of the availability of a systematically developed self-report scale for measuring state anxiety (i.e., the STAI-S [Spielberger et al., 1983], which was used in 28 measurement instances in our review), but not anger. This suggests that the existence of a systematically developed self-report scale may reduce the frequency with which researchers rely on different scales to measure the same emotion across studies—without knowing whether those scales contain convergent sets of items—which is in turn likely to improve the rate at which researchers measure a given emotion with items are unique to that emotion.

Implications: Widespread Occurrence of the Jingle and Jangle Fallacies

Our review of the scale items used to measure distinct emotions suggests that the majority of self-report scales contain items that are also used to measure other purportedly distinct emotions. Using Venn diagrams, Figures 2 and 3 provide a visual illustration of the extent of overlap among single words used to measure distinct emotions. As anticipated, overlap among words used in distinct emotion scales led to frequent instances in which researchers succumb to the *Jingle* and *Jangle* Fallacies (Thorndike, 1904; Kelley, 1927), of which we present examples below. That said, the observed overlap might also reflect the actual structure of emotions, if distinct emotions are characterized in part by overlapping constellations of words, rather than unique sets of words. For example, the presence of several larger circles corresponding to broader emotions (e.g., anger, anxiety, sadness), each containing several words that appear to describe slight variants or more narrow components of the broader state, is consistent with the emotion family account (Ekman, 1992). One interpretation of these Venn diagrams is therefore that a large proportion of affec-

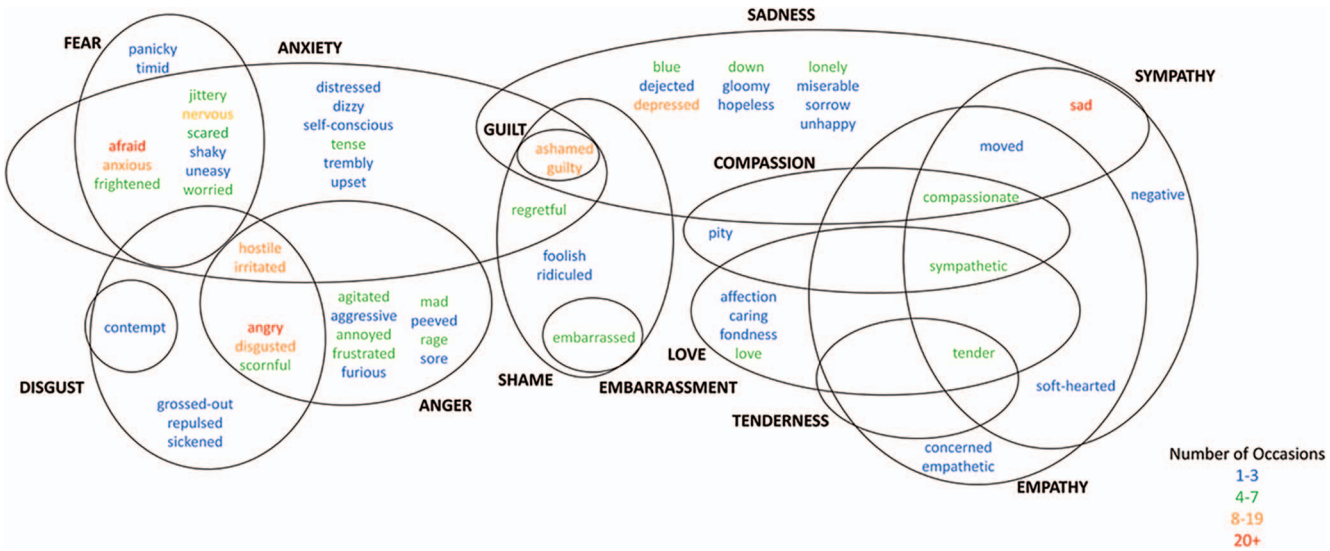

Figure 2. Overlap among words used to measure frequently studied negative emotions and related states. Each circle represents one purportedly distinct emotion measured in the studies we coded, and the words falling within that circle represent the single words used to measure those emotions across all coded studies. For example, the word *sad* falls within the circles for sadness and sympathy because it was used to measure these two emotions, in different studies, whereas the word *compassionate* falls within the circles for compassion, empathy, sadness, and sympathy because it was used to measure these four emotions, in different studies. Color is used to indicate the number of measurement occasions in which a given word was used as a scale item; for example, the word *disgusted* was used 14 times, whereas the word *foolish* was used twice.

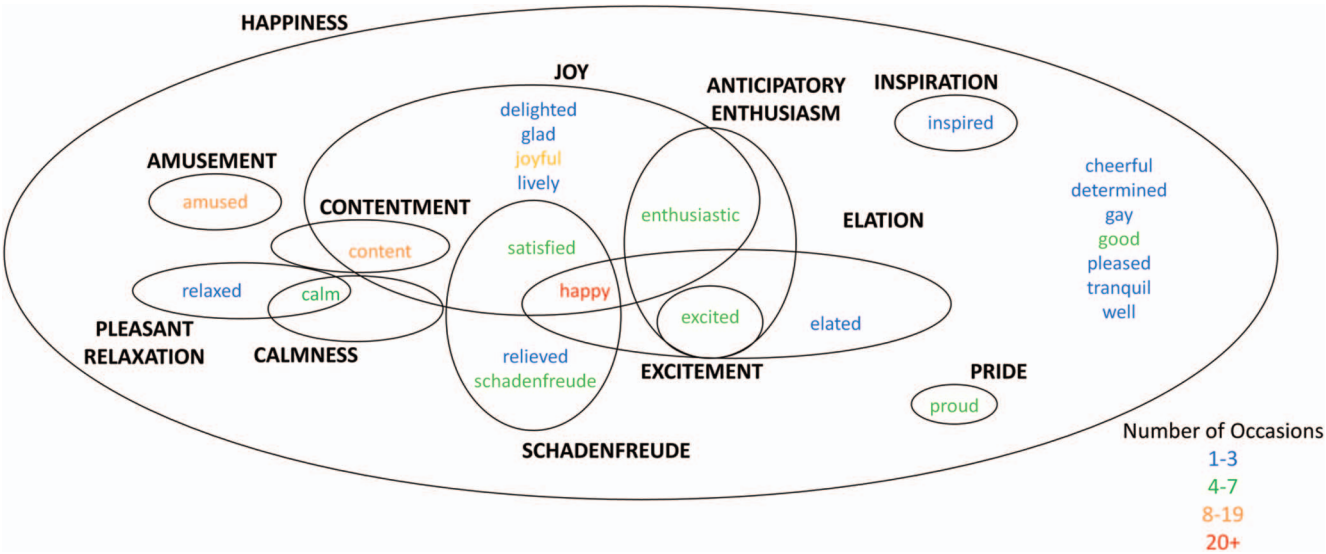

Figure 3. Overlap among words used to measure happiness and related states. Each circle represents one purportedly distinct emotion measured in the studies we coded, and the words falling within that circle represent the single words used to measure those emotions across all coded studies. For example, the word *amused* falls within the circles for amusement and happiness because it was used to measure these two emotions, in different studies, whereas the word *enthusiastic* falls within the circles for anticipatory enthusiasm, elation, joy, and happiness because it was used to measure these four emotions, in different studies. Color is used to indicate the number of measurement occasions in which a given word was used as a scale item; for example, the word *amused* was used 12 times, whereas the word *cheerful* was used 3 times.

tive scientists currently view distinct emotions as best conceptualized as members of broader emotion families, rather than as fully distinct phenomenological entities themselves.

As tempting as it is to draw this type of conclusion based on Figures 1 and 2, it is crucial to keep in mind that our review speaks only to how distinct emotions are currently measured (i.e., epistemology), and not necessarily to their actual nature (i.e., ontology), though both questions are of course important for future research. In the recommendations section of our article, we discuss how researchers might use the observations made in this review to empirically examine the ontological nature of distinct emotions.

Jingle Fallacy. We observed instances in which a single emotion was measured with different sets of words across studies, without having established that those words capture the same emotional experience; for example, researchers used many different sets of items to measure anger (e.g., (a) *angry, infuriated, outraged*, (b) *angry*, *agitated*, *frustrated*, *hostile*, *irritated*; (c) *angry*, *aggressive*, *annoyed*). Given that different anger-related words show varying degrees of similarity to one another (Shaver et al., 1987), varying degrees of centrality to the anger concept (Russell & Fehr, 1994), and varying relations with other distinct emotions (Nabi, 2002), these different sets of words are likely to capture slightly different psychological states, despite the fact that they purportedly measure the single emotion of anger. As a result, these different scales may show variable relations to other constructs, leading to the spurious conclusion that the same emotion (anger) lacks a consistent profile of external correlates. However, it is also possible that several scales comprised of different sets of anger-related words would correlate quite highly with one another, and therefore produce similar empirical effects across studies. This question could be answered empirically by comparing the convergent and predictive validity of several such measures in a single study. Until such work is performed, however, we cannot assume that different sets of words in fact capture identical constructs.

Jangle Fallacy. We observed instances in which multiple purportedly distinct emotions were measured with the same words. For example, in the studies we coded, the words *anxious*, *afraid*, *jittery*, *scared*, and *worried* (among others) were all used to measure the momentary experience of *anxiety* and the momentary experience of *fear*. Yet, evidence from non-self-report research has elucidated ways in which these states may be distinct (see Öhman, 2008, for an overview); for example, studies have suggested that these emotions are associated with activity in distinct brain regions (Walker, Toufexis, & Davis, 2003), distinct facial expressions (Perkins, Inchley-Mort, Pickering, Corr, & Burgess, 2012), distinct heritability components (Hettema, Prescott, Myers, Neale, & Kendler, 2005), distinct patterns of arousal and startle responses (Cuthbert et al., 2003), and distinct classes of psychopathology (Watson, 2005). These findings suggest that researchers measuring anxiety and fear with scales that contain overlapping items may not be able to capture the unique properties of these states that have been identified in prior research. As a result, researchers may reach the incorrect conclusion that fear and anxiety are the same emotion, or that their subjective experiences predict the same outcomes, due to similarities in their empirical correlates, when in fact this similarity is attributable to a failure to develop self-report scales that measure the two states in distinct ways.

What Leads to the Jingle and Jangle Fallacies?

What leads researchers to measure the same emotion with different items, or to measure different emotions with the same items, across studies? Several fairly benign possibilities come to mind; perhaps researchers have sound theoretical reasons for emphasizing one component of an emotion in their scale rather than another, or hold different overarching conceptualizations of an emotion, based on their past research experiences. Or perhaps researchers believe that a close synonym of a word used in a prior scale better captures the emotion they wish to assess than the full scale does. Regardless of the cause of these practices, however, they are likely to amount to an increase in "researcher degrees of freedom," in the form of post-hoc scale construction. In other words, although researchers may not be aware that they are doing so, their decisions to assess emotions using impromptu scales may reflect the fact that most researchers regularly and often unconsciously take advantage of flexibility in their designs to maximize their chances of observing statistically significant effects (e.g., Simmons, Nelson, & Simonsohn, 2011). For example, a researcher might administer a long list of items when conducting a study, and then select the item or set of items that allow for statistically significant results to emerge (i.e., the items that "worked"). He or she may do so explicitly because the chosen item or items seem most appropriate for the given question, but researchers' ability to make such choices in a post-hoc fashion is a key factor leading to the inflation of false positives (Kerr, 1998). To be clear, the present research includes no direct evidence of such "questionable research practices" (John, Loewenstein, & Prelec, 2012). However, in light of the current replicability crisis in psychology (e.g., John et al., 2012; Pashler & Harris, 2012; Simmons et al., 2011), it is noteworthy that the pattern of inconsistent scale usage we observed is consistent with the kinds of practices that have been labeled as "*p* hacking" (Simonsohn, Nelson, & Simmons, 2014), and could therefore be a sign that nonreplicable studies have infiltrated the affective science literature.

Question 3: Are Currently Used Self-Report Scales of Distinct Emotions of Adequate Length?

Finally, we sought to examine the length of scales used to measure distinct emotions. Measuring emotions with scales of adequate length is important for two reasons. First, the experience of distinct emotions—including both basic emotions such as anger, fear, and sadness (Russell & Fehr, 1994; Shaver et al., 1987), and arguably more cognitively complex emotions such as gratitude (Lambert, Graham, & Fincham, 2009), love (Fehr & Russell, 1991; Fehr & Sprecher, 2009), jealousy (Sharpsteen, 1993), and nostalgia (Hepper, Ritchie, Sedikides, & Wildschut, 2012)—is known to be characterized by a broad range of thoughts, feelings, and behaviors (see Russell, 1991b, for an overview). As a result, comprehensively capturing the experiential components of distinct emotions is likely to require the use of multiple self-report items. For example, a researcher wishing to measure joy may need self-report items capturing the extent to which someone is feeling friendly toward others, displaying physical animation, and feeling a sense of belonging; similarly, a researcher wishing to study sadness may need self-report items capturing the extent to which someone wants to withdraw from social contact, feels tired and run down, and has a negative outlook on life (Shaver et al., 1987). These types of nuanced, multi-item scales will in turn demonstrate content validity, typically established when a scale is shown to capture a representative sample of the entire range of content known to be associated with a construct under investigation (Cronbach & Meehl, 1955; Loevinger, 1957).

Second, employing scales of adequate length will increase the chances that those scales demonstrate good reliability. Short measures tend to contain greater error variance (Gulliksen, 1950), in part because they benefit less from aggregation across multiple items (e.g., Epstein, 1983). For example, a single item may capture an idiosyncratic representation of one's feelings due to factors unrelated to one's actual subjective emotional state (e.g., variable interpretation of synonymous single words from occasion to occasion; Goldberg & Kilkowski, 1985), thereby rendering the item score not indicative of the distinct emotion of interest. Indeed, a recent review of personality inventories ranging from two to eight items in length found a strong positive correlation between scale length and reliability ($r = .77$; Credé, Harms, Niehorster, & Gaye-Valentine, 2012), and Classical Test Theory suggests that this principle should apply to self-report emotion scales as well (Gulliksen, 1950).

Achieving adequate reliability is, in turn, important for three reasons. First, low reliability could hamper empirical discoveries, thereby leading to Type-II errors—that is, the failure to observe effects that are, in fact, real. Given that the relation between two constructs is limited by the reliability of either individual construct, if self-report scales show low reliability, observed empirical effects will be attenuated, impeding researchers' ability to detect real relations between distinct emotions and other variables. A second, related, issue is that low reliability could indirectly lead to a greater incidence of Type-I errors, or false positives. The rate of false positives in any set of studies increases if the average statistical power of those studies is low (Pashler & Harris, 2012). To the extent that low reliability curtails statistical power by hampering affective scientists' ability to detect empirical effects of interest, it will lead to an increased rate of false positives in the empirical literature on distinct emotions (i.e., those effects that emerge as statistically significant despite being based on unreliable scales are more likely to be false positives). Third, given that unreliable measures contain a preponderance of measurement error, it is difficult to interpret the size of observed effects between those unreliable measures and measures of other constructs (Kashy, Donnellan, Ackerman, & Russell, 2009; Schmidt & Hunter, 1996), which is particularly problematic in light of the field's current emphasis on more precise estimation of effect sizes (e.g., Cumming, 2014; Eich, 2014; Funder et al., 2014).

To examine scale length, we coded the number of items in each scale used to measure emotions in an observed measurement instance. We found that researchers tended to use relatively short scales ($M = 3.72$ items; $SD = 5.37$; Median: 1 Range: 1–20; see Figure 4). A total of 199 measurements (58%) used a single item, suggesting that this is the modal tendency in measuring distinct emotions through self-report.

To examine whether the preponderance of short scales was associated with observed reliabilities, we coded the internal consistency (i.e., coefficient alpha) of the 147 scales observed in our review that were comprised of more than a single item. We focused on coefficient alpha for two reasons: (a) it is most fre-

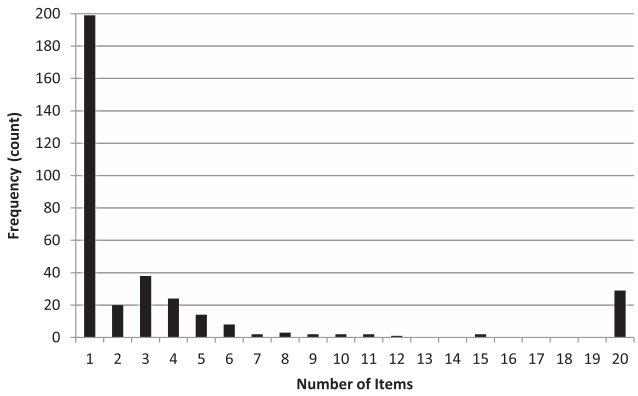

Figure 4. Length of scales used to measure distinct emotions. All but one of the 20-item scales were instances in which researchers used the State–Trait Anxiety Inventory (Spielberger et al., 1983).

quently reported, facilitating comparisons across studies; and (b) it is more appropriate than other commonly used indices (e.g., test–retest reliability) for calculating the reliability of a construct that tends to exhibit true score change over time (as would be expected of distinct emotions; McCrae, Kurtz, Yamagata, & Terracciano, 2011). Coefficient alpha was reported in 66 (45%) of the studies that used a multi-item scale. In these studies, alphas tended to be high (M = .84; SD = .09; Median = .86; Range: .51–.95), though this finding should be interpreted with caution due to severe underreporting.[11]

Implication: Short Measures May Capture Narrow Emotions and Hamper Effect Estimation

Our review suggests that researchers tend to measure momentary distinct emotions with short or single-item scales. This practice is likely to be problematic; given that distinct emotions are characterized by a range of thoughts, feelings, and behaviors, short scales will often fail to comprehensively capture a target emotion. Importantly, even emotions that are assumed to possess relatively simple semantic structures—and therefore likely to be amenable to measurement with a single, face-valid item—may yield variable interpretations when measured with brief scales, especially if samples are comprised of individuals from varying cultural backgrounds (Heider, 1991; Hupka, Lenton, & Hutchison, 1999; Romney, Moore, & Rusch, 1997; Russell, 1991a). Happiness provides a good example; though some would argue that this word possesses a straightforward meaning, researchers have debated whether it captures a preponderance of positive (vs. negative) affect (e.g., Larsen, 2000), a global judgment of life satisfaction (e.g., Diener, Emmons, Larsen, & Griffin, 1985), or an appraisal of well-being across multiple specific life domains (e.g., Ryff, 1989; see Busseri & Sadava, 2011, for a review). Prior empirical work has suggested that the manner in which researchers operationalize happiness influences observed relations between happiness and other variables (e.g., Diener, Ng, Harter, & Arora, 2010; Kahneman & Deaton, 2010; Schimmack, Schupp, & Wagner, 2008; Weidman & Dunn, 2016). For example, Kahneman and Deaton (2010) showed that income correlates positively with global life satisfaction well into the highest income brackets, whereas income correlates positively with daily mood only among individuals making less than $75,000 per year, and has no relation with daily mood among extremely wealthy individuals. These findings highlight the importance of capturing distinctions within the broad emotion construct of happiness by using nuanced, multi-item scales.

Additionally, our review suggests that the internal consistency reliability of self-report scales used to measure emotions is often unknown, because of both the frequent use of single-item scales—for which internal consistency generally cannot be estimated[12]—and the frequency with which internal consistency of scales was not reported. Scales that lack high internal consistency can demonstrate high reliability through another metric (e.g., test–retest reliability), but this was not the case for any of the short or single-item measures we observed in our coding. Unknown reliability will negatively affect the empirical literature on distinct emotions by hampering effect detection and estimation in empirical studies (Kashy et al., 2009; Schmidt & Hunter, 1996). Additionally, to the extent that the frequent use of short and single-item measures leads self-report scales to exhibit low reliability, this will lead to an increased rate of both Type-II and Type-I errors in the distinct emotion literature (e.g., Gulliksen, 1950; Pashler & Harris, 2012).

When might a short measure be useful? They are particularly useful in survey studies with large samples, and experience-sampling studies involving many assessments over a short period of time, in which cases a priority is placed on maximizing economy and efficiency, and reducing participant boredom and fatigue (Burisch, 1984). Single words may also be adequate if the emotion of interest appears to have a relatively straightforward meaning to lay individuals. Although, as noted above, emotion words with seemingly obvious meanings (e.g., *happy*) can be interpreted in different ways, it is reasonable to assume that single, face-valid words may do an adequate job of capturing the subjective experience of certain basic emotions, such as fear or anger, especially when pragmatic considerations increase the utility of short measures. Indeed, short and single-item measures have been successfully adopted to measure relatively straightforward constructs in other domains of psychology (e.g., personality traits, self-esteem; Gosling et al., 2003; Robins, Hendin, & Trzesniewski, 2001), and some proponents of short measures have argued that the optimal trade-off between reliability and validity occurs around just 2 to 4 items (e.g., Burisch, 1997).

[11] The scales for which reliability was reported tended to be shorter (M = 4.83; SD = 3.67) than those for which reliability was not reported (M = 9.49; SD = 5.61; $t(145)$ = 5.81, $p < .001$). However, when excluding the 20-item STAI-S from the list of scales for which reliability was not reported (in 26 of 28 measurement instances, reliability of the STAI was not reported), the mean length of these scales was much shorter, and not significantly different from the mean length of scales for which reliability was reported (M = 4.53, SD = 3.30, $t(119)$ = .47, p = .64).

[12] Researchers relying on single-item measures can use various procedures other than computing coefficient alpha to estimate the reliability of their scales (e.g., correction for attenuation, factor analysis, structural equation modeling; Heise, 1969; Wanous & Hudy, 2001). However, these procedures are typically not amenable to the single time-point measurement laboratory paradigms that frequently characterize the emotion literature.

Relying on a single, face-valid word to measure a straightforward construct is also far preferable to lengthening a scale by adding synonymous words for the sake of boosting internal consistency (e.g., adding the words *fearful* and *frightened* to a scale comprising the word *afraid*). Lengthening a scale by adding redundant items can reduce the scale's predictive or convergent validity, as it can cause the scale to contain many items that assess a single facet of a construct while neglecting to assess the construct's full breadth (Loevinger, 1954). Lengthening a scale may have an additional consequence specific to emotion research; momentary emotions are by definition somewhat transient phenomena, such that if a participant is asked to complete an extremely long scale, her emotional feeling may decay while she is still being asked to report it. Length is therefore only one factor to consider when developing self-report scales of distinct emotions, and there are contexts where the costs of a long scale are not worth the drawbacks.

In scenarios that call for short measures, there are several steps researchers can take to improve the comprehensiveness and interpretability of brief scales (see Gosling, Rentfrow, & Swann, 2003; Rammstedt & John, 2007), including cluster analysis (Wood, Nye, & Saucier, 2010) and algorithm-based item selection (Yarkoni, 2010). Personality researchers have also developed methods for constructing comprehensive single-item measures, by including brief definitions of each pole of a bipolar item, so as to enhance content coverage and retain validity (e.g., Konstabel, Lönnqvist, Walkowitz, Konstabel, & Verkasalo, 2012; Woods & Hampson, 2005). Emotion researchers might similarly develop scales that include brief descriptions of the antecedents, subjective feelings, and functional consequences associated with a given emotion, based on authoritative reviews of prior research and theory regarding the construct. For example, a researcher wishing to measure *compassion* with a single item might provide participants with a definition such as, "the feeling that arises in witnessing another's suffering and that motivates a subsequent desire to help," taken from a review of the research literature on compassion (e.g., Goetz, Keltner, & Simon-Thomas, 2010). Researchers could then assess the extent to which participants felt "compassion." Comprehensive single items such as these have been shown to retain modest reliability and validity when used to measure Big Five personality traits (Konstabel et al., 2012; Woods & Hampson, 2005), suggesting that they may represent an adequate middle ground for affective scientists who wish to measure a distinct emotion with a single item, while using a scale that retains good psychometric properties. Comprehensive single items have significant drawbacks, however, primarily the fact that the descriptions of the construct are akin to a double-barreled (or triple or quadruple barreled) scale item; in the example provided here, the description of compassion is comprehensive but consequently contains several different emotional components, which must be weighed in tandem by participants responding to the item. A multi-item scale, which separately assesses each of these components, would therefore still be optimal in most research contexts.

Implications for Researchers Who do Not Take a Distinct Emotions Perspective

As evidenced by our coding of researchers' theoretical approaches to their emotion research, approximately 16% of the studies that measured momentary distinct emotions with self-report scales appeared to be conducted primarily in the interest of testing hypotheses regarding broader affect dimensions, rather than distinct emotions per se. This list includes 16 (11%) studies that were conducted by researchers who took a dimensionalist theoretical approach but nonetheless measured distinct emotions. For example, several authors framed their broad research goals in terms of examining the effect of positive or negative affect on various outcomes, then described their more specific study goals as examining the effect of specific states such as *happiness*, *amusement*, *sadness*, or *fear* (e.g., Rottenberg et al., 2002; Storbeck & Clore, 2008). This list also includes 8 (5%) studies conducted by authors who adopted a dimensionalist theoretical approach *and* described their studies as involving the measurement of emotion dimensions, but used measures that were labeled with distinct-emotion terms (e.g., Klimstra et al., 2011; Pronin et al., 2008). When researchers use different terms to describe their studies, predictions, and measures, it again renders effects difficult to compare across studies.

Studies that take an explicitly dimensionalist theoretical approach but then use measures that are labeled as assessing distinct emotions can introduce confusion into the empirical literature. Consider two hypothetical studies purporting to examine negative affect, but doing so with two different items or sets of items (e.g., Study 1 uses "angry"; Study 2 uses "afraid"). Emotions such as anger and fear are each thought, by many researchers, to have distinct causes and functional consequences (e.g., Carver & Harmon-Jones, 2009; Öhman, 2008). As a result, these two hypothetical studies of "negative affect" may produce divergent results; for example, based on prior theory, we might expect that Study 1 would show that negative affect involves an approach orientation (Carver & Harmon-Jones, 2009), whereas Study 2 would show that negative affect involves an avoidance orientation (Öhman, 2008). Yet these divergent results would be primarily due to the varied measurement tactics used to operationalize negative affect, and not necessarily to any true variability in the broad construct of negative affect. These divergent results would, in turn, create spurious inconsistency in the literature; making it difficult for readers to compare and integrate findings across studies in which the same theoretical construct (i.e., negative affect) was purportedly measured.

The simplest remedy for closing this sort of gap between theory and measurement is for researchers to label their scale on the basis of the terms or concepts that comprise it, and discuss any empirical findings accordingly. For example, a study seeking to test a hypothesis related to negative affect would ideally include a scale designed to measure negative affect (e.g., the PANAS NA scale, which combines several negative affect adjectives), and would label that scale as such, instead of using a scale that in fact measures state anxiety, anger, depressed affect, or some other more specific construct (see Harmon-Jones, Bastian, & Harmon-Jones, 2016, for a similar discussion). If researchers discuss their studies and measures in the same way as they discuss their theory and hypotheses, theory and measurement will become more consistent and coherent, allowing the field to integrate many studies into a cumulative knowledge base.

Recommendations: Systematically Develop Self-Report Scales

Thus far, we have discussed several ways in which current self-report methods for assessing momentary distinct emotions are problematic, in that they do not allow researchers to gain insight into the unique subjective properties of a theoretically driven set of distinct emotions, nor to uncover the full range of causes and correlates that accompany these emotions.[13] It is important to reiterate, however, that our review pertains only to the *epistemology* of distinct emotions (i.e., what current measurement practices allow researchers to know about the subjective experience of distinct emotions), and not to the *ontology* of distinct emotions (i.e., the true nature of the subjective experience of distinct emotions). Thus, although the measurement practices we have documented may limit researchers' ability to identify the subjective properties and nomological network of distinct emotions, these emotions may nonetheless be characterized by largely unique sets of thoughts, feelings, behaviors, causes, and correlates, especially in light of existing evidence for distinct emotions from other domains (e.g., Ekman, 1992; Panksepp, 2007; Shaver et al., 1987; see Tracy & Randles, 2011, for a review). Stated differently, the limitations inherent in our current capacity for comprehensive self-report measurement do not necessarily speak to the reality of distinct emotional experiences; to know whether this is the case, much more systematic research is needed.

However, given that the ultimate purpose of this review is to encourage further understanding of the ontology of distinct emotions—rather than merely documenting their epistemology—we will conclude with several recommendations for how researchers might develop self-report scales with which to measure momentary distinct emotions. Theory development and scale development have historically been seen as advancing in tandem (Strauss & Smith, 2009); one must have an initial theory about a construct in order to operationally define, and subsequently measure, that construct (Loevinger, 1957), and each subsequent attempt to measure a construct and thereby examine its nomological network provides an opportunity to further refine the initial theory of the construct (Cronbach & Meehl, 1955). In the same manner, we suggest that rigorous scale development efforts for distinct emotions, and subsequent employment of those scales in empirical studies, will help affective scientists refine existing theories about distinct emotions, thereby informing ontology of these states. Specifically, if a researcher has a theory about a distinct emotion, she can test this theory by employing a well-validated measure of this emotion—which itself was developed based on a theory of the emotion—and in turn use the results of this test to further build a theory about the emotion, its unique subjective properties, and how it is distinct from other emotions. Importantly, this bidirectional process between theory and scale development has historical precedent for advancing the field of affective science. In the 1980s, several researchers developed dimensionalist theories of emotion (e.g., Russell, 1980; Watson & Tellegen, 1985); these dimensionalist theories led directly to the development of several scales to measure emotion dimensions (e.g., Barrett & Russell, 1998; Watson et al., 1988), and these scales have been widely used in the interim decades to further refine dimensionalist theories of emotion.

One specific way in which distinct emotion scale development could help refine existing distinct emotion theory is by enabling the integration of findings across studies, by ensuring that the same emotion is assessed with convergent sets of items in different research contexts. Consider a series of 10 studies in which researchers seek to understand the causes, correlates, and consequences of anger, and in which each study involves a different sample of participants and a different social context. If each of the 10 researchers measures their emotion of interest differently, and not with previously validated measures of anger—but all 10 label this emotion *anger*, it becomes impossible to integrate the findings of the 10 studies to reach a broad and generalizable conclusion about *anger*, because that term no longer has a consensual, operational definition across studies. In contrast, if all 10 researchers use measures that are known to have convergent validity to assess anger, then conclusions can be drawn across studies, and contextual and sample differences can be most accurately documented. For example, if the experience of anger is different for someone of Asian cultural descent, compared with an individual of European cultural descent, the only way to demonstrate that difference is to use convergent measures within both populations.

Below we outline three guiding principles researchers might follow in future scale development efforts: draw on existing research and theory, capture lay knowledge, and use short phrases as items. As noted in our review, the gap between distinct emotion measurement and theory, the frequent use of impromptu scales with overlapping items, and the preponderance of single-item measures with low and unknown reliability, together have the potential to hinder the progress of cumulative science within the field of distinct emotions, while increasing the rate of both type-I and type-II errors. The following recommendations are therefore important both for understanding the ontology of distinct emotions, and for promoting more replicable research in the field.

Principle 1: Draw on Existing Research and Theory

By drawing on existing research and theory to guide scale construction, researchers can provide evidence that a distinct emotion of interest is in fact distinct from other previously identified emotions. Existing research and theory can also be used to identify subtle differences within broader, previously identified emotional experiences, which may indicate potentially novel, or previously unstudied, emotional states.

For example, theoretical accounts of shame and guilt have been helpful in identifying potential differences between these two emotional states (see Tangney & Dearing, 2002, for an overview), and the hypothesized differences between these two emotions have since informed scale constructions (e.g., Marschall et al., 1994; Tangney et al., 2000). By formulating and measuring shame and guilt in a manner that mirrors existing conceptualizations of these emotions, researchers increase the probability that they will tap into the distinct content specific to those emotions across studies. In contrast, other studies often measure momentary experiences of

[13] Of note, when conducting our review, we did not observe any systematic trends in measurement practices over time. Specifically, we examined whether any of the primary findings in our review differed across the 11 years of *Emotion* that we coded (i.e., frequency of impromptu scales, frequency of single-item measures, average scale length, frequency with which reliability was not reported); these values showed variability across years, but no interpretable trends emerged.

shame and guilt with single-item self-report scales. Given that lay individuals do not differentiate these constructs at the level of single words (Ellsworth & Smith, 1988; Watson et al., 1988), this practice can lead to considerable confusion. Across studies, despite theoretical distinctions between *shame* and *guilt*, the two emotions may be shown to relate to similar outcomes, in part due to the use of single-item scales that do not differentiate between them, but rather primarily capture their shared negative self-consciousness (Paulhus, Robins, Trzesniewski, & Tracy, 2004; Tignor & Colvin, 2016; though see also Leach & Cidam, 2015).

Theory can also be used to dissociate phenomenologically similar affective states that may in fact be distinguishable based on their other properties. For example, from an evolutionary view, different kinds of love serve different functions, suggesting that the subjective affective experiences that guide the behaviors associated with those functions may also differ. Although love experienced toward one's romantic partner and love toward kin or offspring are similar in terms of valence and arousal, their phenomenological experiences should differ in key respects related to evolutionary selection pressures that shaped the capacity to experience each. Based on inclusive fitness theory (Hamilton, 1964), the love one feels for relatives (e.g., offspring or a brother) motivates one to perform costly acts that benefit those relatives. Since relatives share genes, these behaviors also benefit one's own genes. However, if love was a generic and undifferentiated feeling, individuals might become romantically and sexually attracted to their relatives, which would increase the chances of resultant offspring expressing deleterious double recessive genes, which can lead to severe negative health and/or reproductive consequences (e.g., Fisher, Aron, & Brown, 2006; Joshi et al., 2015; Keller et al., 1994). This suggests that love should be differentiated into at least two distinct states: love toward kin (i.e., attachment love; Shiota et al., 2014) and love toward sex partners (i.e., romantic love; Gonzaga et al., 2006). Evolutionary theory therefore provides an *a priori* reason to expect different kinds of love to differ experientially; if researchers draw on this theoretical perspective to develop measures that capture the components that dissociate these otherwise similar emotions, the resulting measures will likely correspond to ontologically distinct emotional states.

A related advantage to drawing on theory when constructing scales is that doing so will likely help researchers measure the most relevant emotions in a given research context. Take the example above, in which we describe how people should experience two forms of love. If a researcher held a theory about attachment love, but had not used that theory to identify the components that comprise that particular form of love—or to construct a scale based on those components—she might instead employ a measure that primarily captures components of romantic love, or a blend of romantic and attachment love. This would preclude a strong test of her theory about attachment love, because participants would not have been given the opportunity to report on the unique components of attachment love that are most central to the research hypotheses.

Principle 2: Capture Lay Knowledge

Content validity is established by documenting that a set of items on a scale captures a representative sample of the entire universe of content known to characterize the construct under investigation (Cronbach & Meehl, 1955; Loevinger, 1957). To demonstrate content validity, researchers typically first define a universe of content for a construct, and then sample items from that universe when creating a scale (Clark & Watson, 1995). Similarly, if distinct emotion researchers create self-report scales in a bottom-up manner, by first drawing on lay knowledge of the target emotion to establish the content universe, and then developing the scale by selecting items from this universe, the resultant scale will be more likely to capture the emotion of interest as it is experienced by research participants. A bottom-up approach can also help determine whether a distinct emotion of interest is in fact experientially distinct from previously identified emotions—according to research participants. This would allow for a more informed determination of whether the emotion is worthy of study as a novel state, rather than as part of a previously identified, broader state.

Researchers interested in the prototype structure of emotions have provided a blueprint for how such an investigation might be conducted (e.g., Fehr & Russell, 1991; Fehr & Sprecher, 2009; Hepper et al., 2012; Lambert et al., 2009; Russell & Fehr, 1994; Sharpsteen, 1993): Ask participants to list synonyms and features associated with a broad emotion category, and then rate these components on their centrality or prototypicality to the construct. The result is a list of features that are closely associated with the target emotion concept in the minds of the individuals who comprise the research population of interest. A scale constructed from these words is thus likely to comprehensively capture the semantic content central to the target emotion. In addition, if researchers employ large samples of participants with varying cultural and socioeconomic backgrounds, and ask these participants to generate features of an emotion based on their own personal experiences, the resultant list of features will likely capture the core, consistent content of each distinct emotion across a range of different populations and situations. The result will be a set of scales that are useful to researchers across populations and contexts.

This approach has been used previously to facilitate the construction of distinct emotion scales; for example, studies using this method to explore the semantic structure of pride revealed that this emotion consists of two distinct facets: *authentic pride*, associated with words such as *confident*, *accomplished*, and *achieving*; and *hubristic pride*, associated with *haughty*, *boastful*, and *egotistic* (Tracy & Robins, 2007). This investigation eventually resulted in the development of scales to measure each separate pride facet, ensuring that researchers who wish to study pride can tap into each distinctive component of each facet as it is represented in the minds of lay individuals completing the measures. In contrast, a momentary measure of pride that relied on the single item *proud* would result in ambiguity as to which facet was driving any observed relations with other variables of interest. Using these scales, a recent cross-cultural analysis demonstrated that a very similar set of words and phrases are used to describe authentic and hubristic pride experiences in Mainland China and South Korea, suggesting that the two-facet structure of pride generalizes across diverse cultural contexts (Shi et al., 2015). This line of research provides a good example of how a bottom-up approach to scale development can facilitate the derivation of items that capture an emotion across social and cultural contexts, allowing researchers to subsequently measure the emotion in a consistent and meaningful way in different populations and research situations.

Principle 3: Use Short Phrases as Items

Based on our review, single words (as opposed to short phrases) appear to be entirely responsible for the problematic overlap among scale items used to measure distinct emotions. This is likely because many emotions share a common phenomenological core, but nevertheless involve complex appraisals, feelings, and action tendencies. Items comprised of short phrases may better capture these nuances while also reducing conceptual overlap among measures. For example, the emotion *schadenfreude* is defined as pleasure arising from the misfortune of others (Smith, Powell, Combs, & Schurtz, 2009), and, across the various articles reviewed that examined it, was measured with four different words (i.e., *happiness*, *relief*, *satisfied/satisfaction*, *schadenfreude*) and three short phrases (i.e., *I like what happened to* [target of misfortune], *I couldn't resist to smile a little* [at target's misfortune], *Actually, I had to laugh a little at target's misfortune*]). Notably, three of the four single words (all except for *schadenfreude*) overlapped with items included in scales used to measure other emotions (e.g., happiness, elation, relief), but none of the short phrases did. This suggests that the phenomenological core of *schadenfreude* (i.e., pleasure), as captured by single words, may be difficult to distinguish from other emotions sharing that phenomenological core, whereas short phrases capturing the emotion's antecedents and target can better pinpoint differences between *schadenfreude* and other emotions involving a pleasant hedonic core.[14]

The use of short phrases, rather than single words, may also be helpful for developing scales that effectively distinguish between similar yet distinct emotions. For example, although lay persons view *anger* and *contempt* as highly similar emotions (Shaver et al., 1987), prior work has shown that they involve distinct desires: Anger evokes a desire to take corrective action against another individual's perceived wrongdoing, whereas contempt evokes a desire to derogate and avoid another individual (Fischer & Roseman, 2007). These divergent motivations could be incorporated into measures that aim to distinguish the two emotions. Similarly, although lay persons view sadness and depression as similar emotion terms (Shaver et al., 1987), studies suggest that sadness is the more transient experience, whereas depression typically manifests as a longer-lasting syndrome with more debilitating physical and motivational effects (Beck, Steer, & Brown, 1996; Strauman, 2002). In all of these examples, short phrases may be more fruitful than single words in capturing the unique, dissociable properties of these distinct emotions.

Conclusion

In our review of typical practices in emotion research—encompassing studies conducted by researchers who were, for the most part, seeking to make theoretical claims about distinct emotions—we identified a number of trends in the measurement of self-reported momentary distinct emotions that are likely to be problematic. Researchers tend to employ short, impromptu scales that have not been systematically developed and have unknown reliability, and these scales introduce substantial overlap among measured emotions, as the same words and phrases are often used, by different researchers, in different studies, to measure purportedly distinct emotions. These practices have created a pool of distinct emotions measured in the current literature that does not match the distinct emotions that are included in prior taxonomies, a trend that has the potential to create a mismatch between theory and empirical practice that could preclude the advancement of cumulative distinct emotion science. These practices also hinder researchers from integrating findings about any single emotion across multiple studies. Furthermore, these practices may create the potentially misleading impression that emotions are not in fact experienced as distinct phenomenological entities, but rather are characterized by substantial overlap among, due entirely to imprecise measurement.[15]

By alerting the field to the scope and depth of these trends, we hope that our review will encourage researchers to take caution when presenting and interpreting empirical findings. When presenting and interpreting findings regarding a distinct emotion, if researchers and readers carefully examine the scale used to measure that emotion, and consider the items on that scale, and the scale's psychometric properties, they will be able to determine whether the scale in fact captures the construct under investigation in the present study. If researchers ensure that they in fact measure one specified distinct emotion, rather than another closely related emotion, a blend of multiple distinct emotions, a broader state such as positive or negative affect, or a higher-level emotion family, they will better be able to substantiate empirical claims made in their articles.

We also hope that our review will encourage researchers to draw on existing theories of distinct emotions to inform rigorous scale development efforts. In turn, we hope that these scale development efforts will help ensure that self-report scales used in empirical studies capture the thoughts, feelings, and behaviors specific to distinct emotions of interest, and do so consistently across studies. In the final section of this manuscript, we attempted to provide guiding principles for how we think the field can improve upon current measurement practices to ensure the continued advancement of knowledge of distinct emotions. Given the ubiquity with which researchers from a range of psychological subdisciplines assess distinct emotions with self-report, the immediate development and use of scales, based on existing theories of distinct emotions, should be a paramount objective. Researchers have, to date, amassed a considerable amount of evidence regarding how distinct emotions likely evolved, their cognitive antecedents, and how they are expressed in our nonverbal behaviors. The time is now ripe for researchers to pin down the nuanced ways in which these emotions are subjectively experienced.

[14] Researchers using single words to measure *schadenfreude* at times specified the context in which that feeling occurred (e.g., *happy* regarding another's misfortune). This additional specificity may be another useful way to address the problems associated with using single words.

[15] It is worth pointing out that the problematic measurement practices we identified in our review are not necessarily unique to emotion research; it is conceivable that other areas of social-personality psychology are characterized by similar issues. Although we do not have any evidence speaking to whether these practices infiltrate other subdisciplines, we hope that our review provides a blueprint for how researchers might assess the modal measurement practices of a given subfield, and determine whether these practices are facilitating theoretical discovery and cumulative science.

References

Adelson, J. (1969). Personality. *Annual Review of Psychology, 20,* 217–252. http://dx.doi.org/10.1146/annurev.ps.20.020169.001245

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2007). A dynamic model of guilt: Implications for motivation and self-regulation in the context of prejudice. *Psychological Science, 18,* 524–530. http://dx.doi.org/10.1111/j.1467-9280.2007.01933.x

Ashton, M. C., Lee, K., & de Vries, R. E. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality factors: A review of research and theory. *Personality and Social Psychology Review, 18,* 139–152. http://dx.doi.org/10.1177/1088868314523838

Barger, B., Nabi, R., & Hong, L. Y. (2010). Standard back-translation procedures may not capture proper emotion concepts: A case study of Chinese disgust terms. *Emotion, 10,* 703–711. http://dx.doi.org/10.1037/a0021453

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science, 1,* 28–58. http://dx.doi.org/10.1111/j.1745-6916.2006.00003.x

Barrett, L. F. (2014). The conceptual-act theory: A précis. *Emotion Review, 6,* 292–297. http://dx.doi.org/10.1177/1754073914534479

Barrett, L. F., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology, 74,* 967–984. http://dx.doi.org/10.1037/0022-3514.74.4.967

Bartlett, M. Y., & DeSteno, D. (2006). Gratitude and prosocial behavior: Helping when it costs you. *Psychological Science, 17,* 319–325. http://dx.doi.org/10.1111/j.1467-9280.2006.01705.x

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory* (2nd ed.). Boston, MA: Harcourt, Brace, & Co.

Burisch, M. (1984). Approaches to personality inventory construction: A comparison of merits. *American Psychologist, 39,* 214–227. http://dx.doi.org/10.1037/0003-066X.39.3.214

Burisch, M. (1997). Test length and validity revisited. *European Journal of Personality, 11,* 303–315. http://dx.doi.org/10.1002/(SICI)1099-0984(199711)11:4<303::AID-PER292>3.0.CO;2-#

Burns, J. W. (2006). Arousal of negative emotions and symptom-specific reactivity in chronic low back pain patients. *Emotion, 6,* 309–319. http://dx.doi.org/10.1037/1528-3542.6.2.309

Buss, A. H., & Durkee, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology, 21,* 343–349. http://dx.doi.org/10.1037/h0046900

Busseri, M. A., & Sadava, S. W. (2011). A review of the tripartite structure of subjective well-being: Implications for conceptualization, operationalization, analysis, and synthesis. *Personality and Social Psychology Review, 15,* 290–314. http://dx.doi.org/10.1177/1088868310391271

Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect: Evidence and implications. *Psychological Bulletin, 135,* 183–204. http://dx.doi.org/10.1037/a0013965

Cavanaugh, L. A., Cutright, K. M., Luce, M. F., & Bettman, J. R. (2011). Hope, pride, and processing during optimal and nonoptimal times of day. *Emotion, 11,* 38–46. http://dx.doi.org/10.1037/a0022016

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7,* 309–319. http://dx.doi.org/10.1037/1040-3590.7.3.309

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: A new measure of guilt and shame proneness. *Journal of Personality and Social Psychology, 100,* 947–966. http://dx.doi.org/10.1037/a0022641

Costa, P. T., & McCrae, R. R. (1992). *NEO-PI-R professional manual*. Odessa, FL: Psychological Assessment Resources.

Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to "prejudice". *Journal of Personality and Social Psychology, 88,* 770–789. http://dx.doi.org/10.1037/0022-3514.88.5.770

Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology, 102,* 874–888. http://dx.doi.org/10.1037/a0027403

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302. http://dx.doi.org/10.1037/h0040957

Crusius, J., & Mussweiler, T. (2012). When people want what others have: The impulsive side of envious desire. *Emotion, 12,* 142–153. http://dx.doi.org/10.1037/a0023523

Cryder, C. E., Lerner, J. S., Gross, J. J., & Dahl, R. E. (2008). Misery is not miserly: Sad and self-focused individuals spend more. *Psychological Science, 19,* 525–530. http://dx.doi.org/10.1111/j.1467-9280.2008.02118.x

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25,* 7–29. http://dx.doi.org/10.1177/0956797613504966

Cuthbert, B. N., Lang, P. J., Strauss, C., Drobes, D., Patrick, C. J., & Bradley, M. M. (2003). The psychophysiology of anxiety disorder: Fear memory imagery. *Psychophysiology, 40,* 407–422. http://dx.doi.org/10.1111/1469-8986.00043

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L. B., Parvizi, J., & Hichwa, R. D. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience, 3,* 1049–1056. http://dx.doi.org/10.1038/79871

Darwin, C. (1872). *The descent of man and selection in relation to sex*. New York, NY: Appleton & Co.

DeSteno, D., Bartlett, M. Y., Baumann, J., Williams, L. A., & Dickens, L. (2010). Gratitude as moral sentiment: Emotion-guided cooperation in economic exchange. *Emotion, 10,* 289–293. http://dx.doi.org/10.1037/a0017883

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93,* 880–896. http://dx.doi.org/10.1037/0022-3514.93.5.880

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95,* 542–575. http://dx.doi.org/10.1037/0033-2909.95.3.542

Diener, E. (2000). Subjective well-being. The science of happiness and a proposal for a national index. *American Psychologist, 55,* 34–43. http://dx.doi.org/10.1037/0003-066X.55.1.34

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49,* 71–75. http://dx.doi.org/10.1207/s15327752jpa4901_13

Diener, E., Ng, W., Harter, J., & Arora, R. (2010). Wealth and happiness across the world: Material prosperity predicts life evaluation, whereas psychosocial prosperity predicts positive feeling. *Journal of Personality and Social Psychology, 99,* 52–61. http://dx.doi.org/10.1037/a0018066

Dunn, E. W., & Norton, M. I. (2013). *Happy money: The science of smarter spending*. New York, NY: Simon & Schuster.

Eich, E. (2014). Business not as usual. *Psychological Science, 25,* 3–6. http://dx.doi.org/10.1177/0956797613512465

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6,* 169–200. http://dx.doi.org/10.1080/02699939208411068

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17,* 124–129. http://dx.doi.org/10.1037/h0030377

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., . . . Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53,* 712–717. http://dx.doi.org/10.1037/0022-3514.53.4.712

Ellsworth, P. C., & Smith, C. A. (1988). From appraisal to emotion: Differences among unpleasant feelings. *Motivation and Emotion, 12,* 271–302. http://dx.doi.org/10.1007/BF00993115

Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality, 51,* 360–392. http://dx.doi.org/10.1111/j.1467-6494.1983.tb00338.x

Fehr, B., & Russell, J. A. (1991). The concept of love viewed from a prototype perspective. *Journal of Personality and Social Psychology, 60,* 425–438. http://dx.doi.org/10.1037/0022-3514.60.3.425

Fehr, B., & Sprecher, S. (2009). Prototype analysis of the concept of compassionate love. *Personal Relationships, 16,* 343–364. http://dx.doi.org/10.1111/j.1475-6811.2009.01227.x

Fischer, A. H., & Roseman, I. J. (2007). Beat them or ban them: The characteristics and social functions of anger and contempt. *Journal of Personality and Social Psychology, 93,* 103–115. http://dx.doi.org/10.1037/0022-3514.93.1.103

Fisher, H. E., Aron, A., & Brown, L. L. (2006). Romantic love: A mammalian brain system for mate choice. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 361,* 2173–2186. http://dx.doi.org/10.1098/rstb.2006.1938

Ford, B. Q., Tamir, M., Brunyé, T. T., Shirer, W. R., Mahoney, C. R., & Taylor, H. A. (2010). Keeping your eyes on the prize: Anger and visual attention to threats and rewards. *Psychological Science, 21,* 1098–1105. http://dx.doi.org/10.1177/0956797610375450

Funder, D. C., Levine, J. M., Mackie, D. M., Morf, C. C., Sansone, C., Vazire, S., & West, S. G. (2014). Improving the dependability of research in personality and social psychology: Recommendations for research and educational practice. *Personality and Social Psychology Review, 18,* 3–12. http://dx.doi.org/10.1177/1088868313507536

Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: An evolutionary analysis and empirical review. *Psychological Bulletin, 136,* 351–374. http://dx.doi.org/10.1037/a0018807

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4,* 26–42. http://dx.doi.org/10.1037/1040-3590.4.1.26

Goldberg, L. R., & Kilkowski, J. M. (1985). The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms that affect the consistency of responses to synonym and antonym pairs. *Journal of Personality and Social Psychology, 48,* 82–98. http://dx.doi.org/10.1037/0022-3514.48.1.82

Goldin, P. R., & Gross, J. J. (2010). Effects of mindfulness-based stress reduction (MBSR) on emotion regulation in social anxiety disorder. *Emotion, 10,* 83–91. http://dx.doi.org/10.1037/a0018441

Gonzaga, G. C., Turner, R. A., Keltner, D., Campos, B., & Altemus, M. (2006). Romantic love and sexual desire in close relationships. *Emotion, 6,* 163–179. http://dx.doi.org/10.1037/1528-3542.6.2.163

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big Five personality domains. *Journal of Research in Personality, 37,* 504–528. http://dx.doi.org/10.1016/S0092-6566(03)00046-1

Gulliksen, H. (1950). *Theory of mental tests*. New York, NY: Wiley. http://dx.doi.org/10.1037/13240-000

Haidt, J., McCauley, C., & Rozin, P. (1994). Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Personality and Individual Differences, 16,* 701–713. http://dx.doi.org/10.1016/0191-8869(94)90212-7

Hamilton, W. D. (1964). The genetical evolution of social behaviour. I. *Journal of Theoretical Biology, 7,* 1–16. http://dx.doi.org/10.1016/0022-5193(64)90038-4

Harmon-Jones, C., Bastian, B., & Harmon-Jones, E. (2016). Detecting transient emotional responses with improved self-report measures and instructions. *Emotion*. Advance online publication. http://dx.doi.org/10.1037/emo0000216

Harmon-Jones, E., Harmon-Jones, C., Abramson, L., & Peterson, C. K. (2009). PANAS positive activation is associated with anger. *Emotion, 9,* 183–196. http://dx.doi.org/10.1037/a0014959

Heider, K. G. (1991). *Landscapes of emotion*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511527715

Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review, 34,* 93–101. http://dx.doi.org/10.2307/2092790

Hepper, E. G., Ritchie, T. D., Sedikides, C., & Wildschut, T. (2012). Odyssey's end: Lay conceptions of nostalgia reflect its original Homeric meaning. *Emotion, 12,* 102–119. http://dx.doi.org/10.1037/a0025167

Hettema, J. M., Prescott, C. A., Myers, J. M., Neale, M. C., & Kendler, K. S. (2005). The structure of genetic and environmental risk factors for anxiety disorders in men and women. *Archives of General Psychiatry, 62,* 182–189. http://dx.doi.org/10.1001/archpsyc.62.2.182

Hodson, G., & Costello, K. (2007). Interpersonal disgust, ideological orientations, and dehumanization as predictors of intergroup attitudes. *Psychological Science, 18,* 691–698. http://dx.doi.org/10.1111/j.1467-9280.2007.01962.x

Horberg, E. J., Oveis, C., Keltner, D., & Cohen, A. B. (2009). Disgust and the moralization of purity. *Journal of Personality and Social Psychology, 97,* 963–976. http://dx.doi.org/10.1037/a0017423

Hupka, R. B., Lenton, A. P., & Hutchison, K. A. (1999). Universal development of emotion categories in natural language. *Journal of Personality and Social Psychology, 77,* 247–278. http://dx.doi.org/10.1037/0022-3514.77.2.247

Hutcherson, C. A., & Gross, J. J. (2011). The moral emotions: A social-functionalist account of anger, disgust, and contempt. *Journal of Personality and Social Psychology, 100,* 719–737. http://dx.doi.org/10.1037/a0022408

Inbar, Y., Pizarro, D. A., & Bloom, P. (2012). Disgusting smells cause decreased liking of gay men. *Emotion, 12,* 23–27. http://dx.doi.org/10.1037/a0023984

Inzlicht, M., & Tritt, S. (2014, February). Where's the passion? The relative neglect of emotion process variables in social psychology. Paper presented at the annual meeting of the Society for Personality and Social Psychology. Austin, TX.

Isaacowitz, D. M., & Choi, Y. (2011). The malleability of age-related positive gaze preferences: Training to change gaze and mood. *Emotion, 11,* 90–100. http://dx.doi.org/10.1037/a0021551

Izard, C. E. (1971). *The face of emotion*. New York, NY: Appleton-Century-Crofts.

Izard, C. E. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. *Emotion Review, 2,* 363–370. http://dx.doi.org/10.1177/1754073910374661

Jakobs, E., Manstead, A. S. R., & Fischer, A. H. (2001). Social context effects on facial activity in a negative emotional setting. *Emotion, 1,* 51–69. http://dx.doi.org/10.1037/1528-3542.1.1.51

James, W. (1890). What is an emotion? *Mind, 9,* 188–205.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23,* 524–532. http://dx.doi.org/10.1177/0956797611430953

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol. 2, pp. 102–138). New York, NY: Guilford Press.

Johnson, D. R. (2009). Emotional attention set-shifting and its relationship to anxiety and emotion regulation. *Emotion, 9,* 681–690. http://dx.doi.org/10.1037/a0017095

Joshi, P. K., Esko, T., Mattsson, H., Eklund, N., Gandin, I., Nutile, T., . . . Wilson, J. F. (2015). Directional dominance on stature and cognition in diverse human populations. *Nature, 523,* 459–462.

Kahneman, D. (1999). Objective happiness. In D. Kahneman, E. Deiner, & N. Schwarz (Eds.), *Well-being: Foundations of hedonic psychology* (pp. 3–25). New York, NY: Russell Sage Foundation.

Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy*

of Sciences of the United States of America, 107, 16489–16493. http://dx.doi.org/10.1073/pnas.1011492107

Kashy, D. A., Donnellan, M. B., Ackerman, R. A., & Russell, D. W. (2009). Reporting and interpreting research in PSPB: Practices, principles, and pragmatics. *Personality and Social Psychology Bulletin, 35,* 1131–1142. http://dx.doi.org/10.1177/0146167208331253

Keller, L. F., Arcese, P., Smith, J. N., Hochachka, W. M., & Stearns, S. C. (1994). Selection against inbred song sparrows during a natural population bottleneck. *Nature, 372,* 356–357. http://dx.doi.org/10.1038/372356a0

Kelley, T. L. (1927). *Interpretation of educational measurements*. New York, NY: World Book Company.

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology, 68,* 441–454. http://dx.doi.org/10.1037/0022-3514.68.3.441

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2,* 196–217. http://dx.doi.org/10.1207/s15327957pspr0203_4

Ketelaar, T., & Au, W. T. (2003). The effects of feelings of guilt on the behaviour of uncooperative individuals in repeated social bargaining games: An affect-as-information interpretation of the role of emotion in social interaction. *Cognition and Emotion, 17,* 429–453. http://dx.doi.org/10.1080/02699930143000662

Klimstra, T. A., Frijns, T., Keijsers, L., Denissen, J. J. A., Raaijmakers, Q. A. W., van Aken, M. A. W., . . . Meeus, W. H. J. (2011). Come rain or come shine: Individual differences in how weather affects mood. *Emotion, 11,* 1495–1499. http://dx.doi.org/10.1037/a0024649

Konstabel, K., Lönnqvist, J., Walkowitz, G., Konstabel, K., & Verkasalo, M. (2012). The 'Short Five' (S5): Measuring personality traits using comprehensive single-items. *European Journal of Personality, 26,* 13–29. http://dx.doi.org/10.1002/per.813

Kragel, P. A., & LaBar, K. S. (2014). Advancing emotion theory with multivariate pattern classification. *Emotion Review, 6,* 160–174. http://dx.doi.org/10.1177/1754073913512519

Labouvie-Vief, G., Lumley, M. A., Jain, E., & Heinze, H. (2003). Age and gender differences in cardiac reactivity and subjective emotion responses to emotional autobiographical memories. *Emotion, 3,* 115–126. http://dx.doi.org/10.1037/1528-3542.3.2.115

Lambert, N. M., Graham, S. M., & Fincham, F. D. (2009). A prototype analysis of gratitude: Varieties of gratitude experiences. *Personality and Social Psychology Bulletin, 35,* 1193–1207. http://dx.doi.org/10.1177/0146167209338071

Larsen, R. J. (2000). Toward a science of mood regulation. *Psychological Inquiry, 11,* 129–141. http://dx.doi.org/10.1207/S15327965PLI1103_01

Leach, C. W., & Cidam, A. (2015). When is shame linked to constructive approach orientation? A meta-analysis. *Journal of Personality and Social Psychology, 109,* 983–1002. http://dx.doi.org/10.1037/pspa0000037

Lelieveld, G. J., Van Dijk, E., Van Beest, I., & Van Kleef, G. A. (2012). Why anger and disappointment affect other's bargaining behavior differently: The moderating role of power and the mediating role of reciprocal and complementary emotions. *Personality and Social Psychology Bulletin, 38,* 1209–1221. http://dx.doi.org/10.1177/0146167212446938

Lerner, J. S., Gonzalez, R. M., Small, D. A., & Fischhoff, B. (2003). Effects of fear and anger on perceived risks of terrorism: A national field experiment. *Psychological Science, 14,* 144–150. http://dx.doi.org/10.1111/1467-9280.01433

Lerner, J. S., Small, D. A., & Loewenstein, G. (2004). Heart strings and purse strings: Carryover effects of emotions on economic decisions. *Psychological Science, 15,* 337–341. http://dx.doi.org/10.1111/j.0956-7976.2004.00679.x

Levy, K. N., & Kelly, K. M. (2010). Sex differences in jealousy: A contribution from attachment theory. *Psychological Science, 21,* 168–173. http://dx.doi.org/10.1177/0956797609357708

Lindquist, K. A. (2013). Emotions emerge from more basic psychological ingredients: A modern psychological constructionist model. *Emotion Review, 5,* 356–368. http://dx.doi.org/10.1177/1754073913489750

Lishner, D. A., Batson, C. D., & Huss, E. (2011). Tenderness and sympathy: Distinct empathic emotions elicited by different forms of need. *Personality and Social Psychology Bulletin, 37,* 614–625. http://dx.doi.org/10.1177/0146167211403157

Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin, 51,* 493–504. http://dx.doi.org/10.1037/h0058543

Loevinger, J. (1957). Objective tests as instruments for psychological theory. *Psychological Reports, 3,* 635–694. http://dx.doi.org/10.2466/pr0.1957.3.3.635

Lyubomirsky, S., & Lepper, H. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research, 46,* 137–155. http://dx.doi.org/10.1023/A:1006824100041

Marschall, D., Sanftner, J., & Tangney, J. P. (1994). *The state shame and guilt scale*. Fairfax, VA: George Mason University.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion, 23,* 209–237. http://dx.doi.org/10.1080/02699930802204677

McAdams, D. P. (1997). A conceptual history of personality psychology. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 3–39). San Diego, CA: Academic Press. http://dx.doi.org/10.1016/B978-012134645-4/50002-0

McClelland, D. C. (1987). *Human motivation*. Cambridge, UK: Cambridge University Press.

McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review, 15,* 28–50. http://dx.doi.org/10.1177/1088868310366253

McCullough, M. E., Emmons, R. A., & Tsang, J. A. (2002). The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology, 82,* 112–127. http://dx.doi.org/10.1037/0022-3514.82.1.112

McGregor, H. A., & Elliot, A. J. (2005). The shame of failure: Examining the link between fear of failure and shame. *Personality and Social Psychology Bulletin, 31,* 218–231. http://dx.doi.org/10.1177/0146167204271420

McNair, D., Lorr, M., & Droppleman, L. (1971). *Edits manual: Profile of mood states*. San Diego, CA: Educational and Industrial Testing Services.

Mischel, W. (1968). *Personality and assessment*. New York, NY: Wiley.

Modiglianai, A. (1968). Embarrassment and embarrassability. *Sociometry, 31,* 313–326. http://dx.doi.org/10.2307/2786616

Morris, L. W., Davis, M. A., & Hutchings, C. H. (1981). Cognitive and emotional components of anxiety: Literature review and a revised Worry-Emotionality Scale. *Journal of Educational Psychology, 73,* 541–555.

Most, S. B., Laurenceau, J. P., Graber, E., Belcher, A., & Smith, C. V. (2010). Blind jealousy? Romantic insecurity increases emotion-induced failures of visual perception. *Emotion, 10,* 250–256. http://dx.doi.org/10.1037/a0019007

Nabi, R. L. (2002). The theoretical versus the lay meaning of disgust: Implications for emotion research. *Cognition and Emotion, 16,* 695–703. http://dx.doi.org/10.1080/02699930143000437

Nelissen, R. M. A., Leliveld, M. C., van Dijk, E., & Zeelenberg, M. (2011). Fear and guilt in proposers: Using emotions to explain offers in ultimatum bargaining. *European Journal of Social Psychology, 41,* 78–85. http://dx.doi.org/10.1002/ejsp.735

Öhman, A. (2008). Fear and anxiety: Overlaps and dissociations. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 709–729). New York, NY: Guilford Press.

Panksepp, J. (2007). Neurologizing the psychology of affects: How appraisal-based constructivism and basic emotion theory can coexist. *Perspectives on Psychological Science, 2,* 281–296. http://dx.doi.org/10.1111/j.1745-6916.2007.00045.x

Parker Tapias, M., Glaser, J., Keltner, D., Vasquez, K., & Wickens, T. (2007). Emotions and prejudice: Specific emotions toward outgroups. *Group Processes & Intergroup Relations, 10,* 27–39. http://dx.doi.org/10.1177/1368430207071338

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7,* 531–536. http://dx.doi.org/10.1177/1745691612463401

Paulhus, D. L., Robins, R. W., Trzesniewski, K. H., & Tracy, J. L. (2004). Two replicable suppressor situations in personality research. *Multivariate Behavioral Research, 39,* 303–328. http://dx.doi.org/10.1207/s15327906mbr3902_7

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality, 36,* 556–563. http://dx.doi.org/10.1016/S0092-6566(02)00505-6

Paunonen, S. V. (1998). Hierarchical organization of personality and prediction of behavior. *Journal of Personality and Social Psychology, 74,* 538–556. http://dx.doi.org/10.1037/0022-3514.74.2.538

Paunonen, S. V., & Ashton, M. C. (2013). On the prediction of academic performance with personality traits: A replication study. *Journal of Research in Personality, 47,* 778–781. http://dx.doi.org/10.1016/j.jrp.2013.08.003

Perkins, A. M., Inchley-Mort, S. L., Pickering, A. D., Corr, P. J., & Burgess, A. P. (2012). A facial expression for anxiety. *Journal of Personality and Social Psychology, 102,* 910–924. http://dx.doi.org/10.1037/a0026825

Poldrack, R. A. (2010). Mapping mental function to brain structure: How can cognitive neuroimaging succeed? *Perspectives on Psychological Science, 5,* 753–761. http://dx.doi.org/10.1177/1745691610388777

Pronin, E., Jacobs, E., & Wegner, D. M. (2008). Psychological effects of thought acceleration. *Emotion, 8,* 597–612. http://dx.doi.org/10.1037/a0013268

Quartana, P. J., & Burns, J. W. (2007). Painful consequences of anger suppression. *Emotion, 7,* 400–414. http://dx.doi.org/10.1037/1528-3542.7.2.400

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41,* 203–212. http://dx.doi.org/10.1016/j.jrp.2006.02.001

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12,* 287–297. http://dx.doi.org/10.1037/1040-3590.12.3.287

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem scale. *Personality and Social Psychology Bulletin, 27,* 151–161. http://dx.doi.org/10.1177/0146167201272002

Romney, A. K., Moore, C. C., & Rusch, C. D. (1997). Cultural universals: Measuring the semantic structure of emotion terms in English and Japanese. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 94,* 5489–5494. http://dx.doi.org/10.1073/pnas.94.10.5489

Roseman, I. J. (2011). Emotional behaviors, emotivational goals, emotion strategies: Multiple levels of organization integrate variable and consistent responses. *Emotion Review, 3,* 434–443. http://dx.doi.org/10.1177/1754073911410744

Rottenberg, J., Kasch, K. L., Gross, J. J., & Gotlib, I. H. (2002). Sadness and amusement reactivity differentially predict concurrent and prospective functioning in major depressive disorder. *Emotion, 2,* 135–146. http://dx.doi.org/10.1037/1528-3542.2.2.135

Rudd, M., Vohs, K. D., & Aaker, J. (2012). Awe expands people's perception of time, alters decision making, and enhances well-being. *Psychological Science, 23,* 1130–1136. http://dx.doi.org/10.1177/0956797612438731

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178. http://dx.doi.org/10.1037/h0077714

Russell, J. A. (1991a). 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. (1991b). In defense of a prototype approach to emotion concepts. *Journal of Personality and Social Psychology, 60,* 37–47. http://dx.doi.org/10.1037/0022-3514.60.1.37

Russell, J. A. (2014). Four perspectives on the psychology of emotion: An introduction. *Emotion Review, 6,* 291. http://dx.doi.org/10.1177/1754073914534558

Russell, J. A., & Fehr, B. (1994). Fuzzy concepts in a fuzzy hierarchy: Varieties of anger. *Journal of Personality and Social Psychology, 67,* 186–205. http://dx.doi.org/10.1037/0022-3514.67.2.186

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57,* 1069–1081. http://dx.doi.org/10.1037/0022-3514.57.6.1069

Sbarra, D. A., & Ferrer, E. (2006). The structure and process of emotional experience following nonmarital relationship dissolution: Dynamic factor analyses of love, anger, and sadness. *Emotion, 6,* 224–238.

Schimmack, U., Schupp, J., & Wagner, G. G. (2008). The influence of environment and personality on the affective and cognitive component of subjective well-being. *Social Indicators Research, 89,* 41–60. http://dx.doi.org/10.1007/s11205-007-9230-3

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1,* 199–223. http://dx.doi.org/10.1037/1082-989X.1.2.199

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin, 34,* 1096–1109. http://dx.doi.org/10.1177/0146167208317771

Schnall, S., Roper, J., & Fessler, D. M. (2010). Elevation leads to altruistic behavior. *Psychological Science, 21,* 315–320. http://dx.doi.org/10.1177/0956797609359882

Sharpsteen, D. J. (1993). Romantic jealousy as an emotion concept: A prototype analysis. *Journal of Social and Personal Relationships, 10,* 69–82. http://dx.doi.org/10.1177/0265407593101005

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology, 52,* 1061–1086. http://dx.doi.org/10.1037/0022-3514.52.6.1061

Sherman, G. D., Haidt, J., & Clore, G. L. (2012). The faintest speck of dirt: Disgust enhances the detection of impurity. *Psychological Science, 23,* 1506–1514. http://dx.doi.org/10.1177/0956797612445318

Shi, Y., Chung, J. M., Cheng, J. T., Tracy, J. L., Robins, R. W., Chen, X., & Zheng, Y. (2015). Cross-cultural evidence for the two-facet structure of pride. *Journal of Research in Personality, 55,* 61–74. http://dx.doi.org/10.1016/j.jrp.2015.01.004

Shiota, M. N., Keltner, D., & John, O. P. (2006). Positive emotion dispositions differentially associated with big five personality and attachment style. *The Journal of Positive Psychology, 1,* 61–71. http://dx.doi.org/10.1080/17439760500510833

Shiota, M. N., Neufeld, S. L., Danvers, A. F., Osborne, E. A., Sng, O., & Yee, C. I. (2014). Positive emotion differentiation: A functional ap-

proach. *Social and Personality Psychology Compass, 8,* 104–117. http://dx.doi.org/10.1111/spc3.12092

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359–1366. http://dx.doi.org/10.1177/0956797611417632

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143,* 534–547. http://dx.doi.org/10.1037/a0033242

Smith, R. H., Parrott, W. G., Diener, E. F., Hoyle, R. H., & Kim, S. H. (1999). Dispositional envy. *Personality and Social Psychology Bulletin, 25,* 1007–1020. http://dx.doi.org/10.1177/01461672992511008

Smith, R. H., Powell, C. A. J., Combs, D. J. Y., & Schurtz, D. R. (2009). Exploring the when and why of schadenfreude. *Social and Personality Psychology Compass, 3,* 530–546. http://dx.doi.org/10.1111/j.1751-9004.2009.00181.x

Spielberger, C. D., Gorssuch, R. L., Lushene, P. R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory*. Mountain View, CA: Consulting Psychologists Press.

Storbeck, J., & Clore, G. L. (2008). The affective regulation of cognitive priming. *Emotion, 8,* 208–215. http://dx.doi.org/10.1037/1528-3542.8.2.208

Strauman, T. J. (2002). Self-regulation and depression. *Self and Identity, 1,* 151–157. http://dx.doi.org/10.1080/152988602317319339

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology, 5,* 1–25.

Sweetman, J., Spears, R., Livingstone, A. G., & Manstead, A. S. R. (2013). Admiration regulates social hierarchy: Antecedents, dispositions, and effects on intergroup behavior. *Journal of Experimental Social Psychology, 49,* 534–542. http://dx.doi.org/10.1016/j.jesp.2012.10.007

Tangney, J. P., & Dearing, R. L. (2002). *Shame and guilt*. New York, NY: Guilford Press. http://dx.doi.org/10.4135/9781412950664.n388

Tangney, J. P., Dearing, R. L., Wagner, P. E., & Gramzow, R. (2000). *The test of self-conscious affect–3 (TOSCA-3)*. Fairfax, VA: George Mason University.

Thorndike, E. L. (1904). *An introduction to the theory of mental and social measurements*. New York, NY: Columbia University Press. http://dx.doi.org/10.1037/13283-000

Tignor, S. M., & Colvin, C. R. (2016). The interpersonal adaptiveness of dispositional guilt and shame: A meta-analytic investigation. *Journal of Personality*. http://dx.doi.org/10.1111/jopy.12244

Tracy, J. L. (2014). Author reply: Incompatible conclusions or different levels of analysis. *Emotion Review, 6,* 330–331. http://dx.doi.org/10.1177/1754073914534504

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman & Cordaro, Izard, Levenson, and Panksepp & Watt. *Emotion Review, 3,* 397–405. http://dx.doi.org/10.1177/1754073911410747

Tracy, J. L., & Robins, R. W. (2007). The psychological structure of pride: A tale of two facets. *Journal of Personality and Social Psychology, 92,* 506–525. http://dx.doi.org/10.1037/0022-3514.92.3.506

Tracy, J. L., & Robins, R. W. (2008). The nonverbal expression of pride: Evidence for cross-cultural recognition. *Journal of Personality and Social Psychology, 94,* 516–530. http://dx.doi.org/10.1037/0022-3514.94.3.516

Valdesolo, P., & Desteno, D. (2011). Synchrony and the social tuning of compassion. *Emotion, 11,* 262–266. http://dx.doi.org/10.1037/a0021302

Walker, D. L., Toufexis, D. J., & Davis, M. (2003). Role of the bed nucleus of the stria terminalis versus the amygdala in fear, stress, and anxiety. *European Journal of Pharmacology, 463,* 199–216. http://dx.doi.org/10.1016/S0014-2999(03)01282-2

Wanous, J. P., & Hudy, M. J. (2001). Single-item reliability: A replication and extension. *Organizational Research Methods, 4,* 361–375. http://dx.doi.org/10.1177/109442810144003

Watson, D. (2005). Rethinking the mood and anxiety disorders: A quantitative hierarchical model for *DSM-V*. *Journal of Abnormal Psychology, 114,* 522–536. http://dx.doi.org/10.1037/0021-843X.114.4.522

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

Weidman, A. C., & Dunn, E. W. (2016). The unsung benefits of material things: Material purchases provide more frequent momentary happiness than experiential purchases. *Social Psychological and Personality Science, 7,* 390–399. http://dx.doi.org/10.1177/1948550615619761

Williams, L. A., & DeSteno, D. (2008). Pride and perseverance: The motivational role of pride. *Journal of Personality and Social Psychology, 94,* 1007–1017. http://dx.doi.org/10.1037/0022-3514.94.6.1007

Williams, L. A., & DeSteno, D. (2009). Pride: Adaptive social emotion or seventh sin? *Psychological Science, 20,* 284–288. http://dx.doi.org/10.1111/j.1467-9280.2009.02292.x

Winter, D. G., John, O. P., Stewart, A. J., Klohnen, E. C., & Duncan, L. E. (1998). Traits and motives: Toward an integration of two traditions in personality research. *Psychological Review, 105,* 230–250. http://dx.doi.org/10.1037/0033-295X.105.2.230

Wood, D., Nye, C. D., & Saucier, G. (2010). Identification and measurement of a more comprehensive set of person-descriptive trait markers from the English lexicon. *Journal of Research in Personality, 44,* 258–272. http://dx.doi.org/10.1016/j.jrp.2010.02.003

Woods, S. A., & Hampson, S. E. (2005). Measuring the Big Five with single items using a bipolar response scale. *European Journal of Personality, 19,* 373–390. http://dx.doi.org/10.1002/per.542

Yarkoni, T. (2010). The abbreviation of personality, or how to measure 200 personality scales with 200 items. *Journal of Research in Personality, 44,* 180–198. http://dx.doi.org/10.1016/j.jrp.2010.01.002

Zhou, X., Sedikides, C., Wildschut, T., & Gao, D. G. (2008). Counteracting loneliness: On the restorative function of nostalgia. *Psychological Science, 19,* 1023–1029. http://dx.doi.org/10.1111/j.1467-9280.2008.02194.x

Received May 28, 2015
Revision received July 28, 2016
Accepted August 2, 2016 ■