Psychological Bulletin
2018, Vol. 144, No. 4, 343–393

0033-2909/18/$12.00 http://dx.doi.org/10.1037/bul0000128

Emotion Fingerprints or Emotion Populations? A Meta-Analytic Investigation of Autonomic Features of Emotion Categories

Erika H. Siegel
University of California, San Francisco

Molly K. Sands
Northeastern University

Wim Van den Noortgate
University of Leuven

Paul Condon, Yale Chang, and Jennifer Dy
Northeastern University

Karen S. Quigley
Edith Nourse Rogers Memorial VA Hospital, Bedford, Massachusetts, and Northeastern University

Lisa Feldman Barrett
Northeastern University and Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts

The classical view of emotion hypothesizes that certain emotion categories have a specific autonomic nervous system (ANS) "fingerprint" that is distinct from other categories. Substantial ANS variation within a category is presumed to be epiphenomenal. The theory of constructed emotion hypothesizes that an emotion category is a population of context-specific, highly variable instances that need not share an ANS fingerprint. Instead, ANS variation within a category is a meaningful part of the nature of emotion. We present a meta-analysis of 202 studies measuring ANS reactivity during lab-based inductions of emotion in nonclinical samples of adults, using a random effects, multilevel meta-analysis and multivariate pattern classification analysis to test our hypotheses. We found increases in mean effect size for 59.4% of ANS variables across emotion categories, but the pattern of effect sizes did not clearly distinguish 1 emotion category from another. We also observed significant variation within emotion categories; heterogeneity accounted for a moderate to substantial percentage (i.e., $I^2 \geq 30\%$) of variability in 54% of these effect sizes. Experimental moderators epiphenomenal to emotion, such as induction type (e.g., films vs. imagery), did not explain a large portion of the variability. Correction for publication bias reduced estimated effect sizes even further, increasing heterogeneity of effect sizes for certain emotion categories. These findings, when considered in the broader empirical literature, are more consistent with population thinking and other principles from evolutionary biology found within the theory of constructed emotion, and offer insights for developing new hypotheses to understand the nature of emotion.

This article was published Online First February 1, 2018.

Erika H. Siegel, Department of Psychiatry, University of California, San Francisco; Molly K. Sands, Department of Psychology, Northeastern University; Wim Van den Noortgate, Department of Psychology and Educational Sciences, University of Leuven; Paul Condon, Department of Psychology, Northeastern University; Yale Chang and Jennifer Dy, Department of Electrical and Computer Engineering, Northeastern University; Karen S. Quigley, Department of Veterans Affairs, Edith Nourse Rogers Memorial VA Hospital, Bedford, Massachusetts, and Department of Psychology, Northeastern University; Lisa Feldman Barrett, Department of Psychology, Northeastern University, and Department of Psychiatry, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts.

Paul Condon is now at Department of Psychology, Northeastern University.

The Karen S. Quigley and Lisa Feldman Barrett share senior authorship. Special thanks to Justin Kopec, Lauren Sears, Sam Lyons, and Victoria Spring for their help with the literature search and for database(s) organization. Thanks also to Maria Gendron for her helpful comments on a draft of the article. Preparation of this article was supported by a U.S. National Institutes of Health (NIH) Director's Pioneer Award (DP1OD003312), a U.S. National Institute on Aging grant (R01AG030311), a U.S. National Science Foundation grant (BCS-1052790), and contracts from the U.S. Army Research Institute for the Behavioral and Social Sciences (Contract W5J9CQ-12-C-0049, W5J9CQ-11-C-0046) to Lisa Feldman Barrett. Army Research Institute (W911NF-16-1-0191), Individual differences in emotional experience and cognitive performance ($1,113,183). Principal Investigator: to Karen S. Quigley. National Cancer Institute (U01 CA193632), Fundamental subcortical mechanisms in affective processing ($3,180,229) to Lisa Feldman Barrett. National Institute for Mental Health (T32 MH019391) to Erika H. Siegel. The views, opinions, and findings contained in this article are those of the authors and should not be construed as an official position, policy, or decision of the U.S. National Institutes of Health or Department of the Army unless so designated by other documents.

Correspondence concerning this article should be addressed to Erika H. Siegel, Department of Psychiatry, University of California, San Francisco, San Francisco, CA 94118. E-mail: erika.siegel@ucsf.edu

Public Significance Statement
This meta-analytic investigation demonstrates that there is no 1-to-1 mapping between an emotion category and a specific autonomic nervous system response pattern. In addition, we observed substantial variability in autonomic nervous system changes during instances of the same emotion category that was not accounted for by experimental moderators (such as the way the emotion was induced). These findings suggest that autonomic nervous system changes during emotion are less like a bodily fingerprint and more like a population of variable, context sensitive instances.

Keywords: emotion, peripheral psychophysiology, meta-analysis, autonomic nervous system specificity, theory of constructed emotion

Supplemental materials: http://dx.doi.org/10.1037/bul0000128.supp

Consider the following situations during which you might feel angry: You are cut off in traffic, your romantic partner has been unfaithful, you are served an overcooked piece of fish at a restaurant, you have difficulty finding parking at the airport and miss your flight, your roof leaks and damages some beloved photographs, and your coworker receives a promotion that you feel you deserved. Each emotional episode will involve some coordinated changes in your facial muscle movements, your specific actions, the activity of your autonomic nervous system (ANS; e.g., heart rate, respiration, activity of the sweat glands, etc.), and in the affective quality of your experience. A century-long debate exists over whether there is variation in these coordinated changes across different episodes of anger (or of any set of instances within the same emotion category), and if so, whether this variation is an intrinsic and functional part of emotion or whether variation across instances of emotion is caused by something epiphenomenal to emotion (such as the method used to induce emotion) and therefore masking a single coordinated pattern for all (or most) instances of an emotion category.

The classical view of emotion (including basic emotion theories and causal appraisal theories; Barrett, Mesquita Ochsner, & Gross, 2007; Gross & Barrett, 2011) proposes that the instances of each emotion category share a distinctive pattern of ANS activity (a metaphorical "fingerprint") and that different emotion categories have distinct, diagnostic fingerprints. Real fingerprints are used to identify a person because the pattern of ridges and valleys on each finger is unique. The observed pattern (i.e., the print of the ridges) varies somewhat from one instance to the next depending on the degree of pressure used, the surfaces touched, the amount of sweat present, and so on, but in principle, the unique fingerprint can still be identified as belonging to one individual and one individual only. In the same way, the autonomic pattern for "anger" need not be identical for every instance for the classical view of emotion to be correct, so some variation from instance to instance is permitted, but the pattern should be *sufficiently similar* to identify those instances as anger and to distinguish them from instances of other emotion categories (named by words such as "sadness," "fear," "disgust," or "happiness") that each have their own unique ANS fingerprints. An emotion category, therefore, is presumed to be a typological category whose instances are *physically* similar to one another. Significant variation in ANS activity is thought to be due to emotion regulation or other processes that are epiphenomenal to emotion itself. We call this the *fingerprint hypothesis*.

Alternatively, a constructionist view of emotion (Barrett, 2013, 2017a, 2017c; Barrett & Russell, 2015) hypothesizes that the ANS pattern occurring during an instance of an emotion category such as anger will be tailored to the specific demands of that situation. As such, ANS patterns are expected to be highly variable within an emotion category and to overlap with other categories. Within a constructionist framework, an emotion category is a *conceptual* category, where the similarity across instances of the same category and the differences across categories is in the mind of the person who is grouping perceptually different instances together for some purpose. An emotion category is thus suggested to be a grouping or *population* of context-dependent, variable instances. We call this the *population hypothesis* (see Table 1 for an articulation of the core assumptions for both the classical and constructionist views of emotion).

In this article, we use meta-analysis to empirically test whether emotion categories are more like physical categories with ANS "fingerprints" or like conceptual categories constructed as "populations" of highly variable, situated instances. The published scientific literature contains hundreds of individual studies, 10 qualitative reviews, four meta-analyses, and a handful of multivariate pattern classification analyses that have focused simply on whether or not emotion categories have ANS fingerprints (Barrett, 2006a; Cacioppo et al., 2000; Duffy, 1957; Friedman, 2010; Hunt, 1941; Kreibig, 2010; Lench, Flores, & Bench, 2011; Levenson, 1992, 2011, 2014; Lindquist, Siegel, Quigley, & Barrett, 2013; Mandler, 1984; Mauss & Robinson, 2009; Quigley & Barrett, 2014; Schachter & Singer, 1962; Stemmler, 2004). Repeated failures to observe widespread support for the classical view's fingerprint hypothesis have been largely viewed as evidence that ANS variation across instances of a given emotion category is due to random error or results from a process that epiphenomenal to emotion (e.g., the method by which the emotion was induced). In effect, this interpretation has made it difficult for empirical evidence to disconfirm the classical view of emotion. The population hypothesis, in contrast, proposes that ANS changes are tied to the metabolic demands associated with actual behaviors (e.g., cardiosomatic coupling; Obrist, Webb, Sutterer, & Howard, 1970) and expected behaviors (e.g., suprametabolic activity; Obrist, 1981; Sterling, 2012; Turner & Carroll, 1985) that vary with situational demands. From this theoretical perspective, significant ANS variation is a substantive finding about the nature of emotion.

We begin this article with a brief review of the classical and constructionist approaches that have guided hypothesis testing on

Table 1
Core Assumptions of Classical and Constructionist Views of Emotion

Assumption	Classical		Constructionist
	Strong version	Weak version	
1. How is an emotion category defined?	A physical category. A collection of instances sharing a set of core physical features (a fingerprint) that reflect a central emotional state (i.e., an essence).	Same as strong version.	A conceptual category. A population of context-dependent, variable instances that are treated as similar for some purpose.
2. Are there natural discontinuities in the variation?	Yes. Nature defines what is, and what is not, in a given emotion category. Emotion categories are perceiver-independent.	Same as strong version.	No. Collective intentionality defines what is, and what is not, in a given emotion category. Some variation is attended to (forming a category boundary) and some variation is ignored so that instances become similar in the service of some function or goal. The similarity between instances is in the mind of the perceiver, so that emotion categories are perceiver-dependent.
3. Do emotions have unique patterns of activity (i.e., fingerprints) in the autonomic nervous system (ANS)?	Yes. Instances of the same emotion category (barring error) involve a consistent and specific pattern of ANS activity.	Same as the strong version.	No. ANS patterns occurring during instances of emotion will be tailored to the specific action demands of the situation. Autonomic patterns are expected to be highly variable within an emotion category and to have some similarity across categories.
4. Is there any variation in the ANS pattern within an emotion category?	Yes, but it is limited. More substantial variation is considered epiphenomenal to emotion.	Same as strong version, but more variation might be observed.	There is substantial ANS variation within each emotion category, and ANS similarities across categories. Additional ingredients are necessary to transform ANS changes into an instance of an emotion.
5. Evolutionary inspiration.	*The Expression of Emotion in Man and Animals* (1872, 2005).	Same as strong version.	*On the Origin of Species* (1859, 2001).

the nature of ANS changes during emotion, introducing their competing hypotheses in strong inferential terms. Next, we review past meta-analyses and how they relate to our present hypotheses. We then present a comprehensive, multilevel meta-analytic strategy, which examines empirical studies that have measured ANS changes during emotional episodes. Our analysis is unique in two respects. First, we directly compared the predictions that derive from the fingerprint hypothesis (and the classical view of emotion) versus the population hypothesis (and the constructionist view of emotion). Second, we report a multivariate pattern classification analysis that searched for evidence of higher-dimensional ANS patterns. We end the article by discussing the implications of the results for the future of emotion research.[1]

The Classical View of Emotion: The Fingerprint Hypothesis

The classical view of emotion includes a variety of ideas about the existence of basic emotion and causal appraisal processes (e.g., Anderson & Adolphs, 2014; Ekman, 1984; Ekman, 2007; Ekman & Cordaro, 2011; Ekman & Davidson, 1994; Ekman, Levenson, & Friesen, 1983; Frijda, 1986; Izard, 1977; Lange, 1922; Levenson, 2011; Levenson, 2014; Panksepp, 1998; Roseman, 1996; Roseman, 2011; Tooby & Cosmides, 1990; Tracy & Randles, 2011). These ideas were deeply inspired by Darwin's *The Expression of the Emotions in Man and Animals* (Darwin, 1872/2005) that infers the existence of emotion essences (see Barrett, 2011a)[2] that evolved as adaptive responses to the demands of recurring situations. The classical view hypothesizes that the pattern of ANS changes is highly similar across instances of the same emotion category, reflecting some central emotional state. Regardless of whether you cry in anger, shout in anger, smile in anger, freeze in anger, or laugh in the face of anger, the classical view hypothesizes that each time you will be in the same central anger state characterized by a diagnostic ANS pattern. This pattern is hypothesized

[1] Here we do not deal with the brain basis of emotion, which has been meta-analytically summarized elsewhere by our group and others (Lindquist et al., 2016; Lindquist et al., 2012; Vytal & Hamann, 2010; Wager et al., 2015). We also excluded facial electromyographic measurements of emotion, which we will address in a separate report. Finally, we do not address broader meta-analytic questions about the nature of affect (pleasure/displeasure, arousal), which we will also address in a separate report.

[2] Descriptions of Darwin's writings suffer from a number of errors that continue to be perpetuated in the published literature on emotion. For example, in *The Expression of the Emotions in Man and Animals*, Darwin did not write that emotions or their expressions are functional adaptations. He wrote that they were functionless vestiges (Barrett, 2011a, 2017a; Fridlund, 1992); instead, it was Floyd Allport (1924) who reinterpreted Darwin (and it is this reinterpretation that modern authors mistakenly describe).

to be *consistent* across contexts and individuals regardless of age and culture. An ANS pattern should be *specific* to one (and only one) emotion category. This typological approach to defining an emotion category corresponds to a classical view of categories (Murphy, 2002), where each emotion category is presumed to have necessary and sufficient features that function as a metaphorical "fingerprint." There is a general consensus within the classical view that "anger," "disgust," "fear," "happiness," and "sadness" qualify as "basic" categories meaning they have a biological fingerprint and are universally expressed and recognized (Ekman & Cordaro, 2011; Tracy & Randles, 2011), although there continue to be disagreements about whether other categories meet these criteria, (e.g., Ortony & Turner, 1990).

The classical view of emotion permits limited variation in ANS changes across instances of the same emotion category. For example, the word "anger" is said to refer to a family of responses that includes "annoyance," "frustration," and "aggravation." Even amid any such variation, however, each family member is hypothesized to share a characteristic pattern of ANS physiology that is *consistently present and recognizably different* from the patterns found in other emotion families, such as the fear and happiness families (for specific quotations, see Ekman, 1992, p. 550; Ekman & Cordaro, 2011, p. 364; Levenson, 2011, p. 379; Scarantino & Griffiths, 2011, pp. 448–449). Variation around the presumed fingerprint is usually thought to be epiphenomenal to the central emotion state itself, reflecting stochastic fluctuations, individual differences in the structure of the nervous system, display rules, emotion regulation strategies, differences in induction methods, or measurement error (Levenson, 2011, 2014; Matsumoto, 1990; Roseman, 2001; Rychlowska et al., 2015).

A weaker version of the fingerprint hypothesis Levenson (1992, 2011, 2014) proposes that ANS changes observed in the lab may be insufficient to distinguish emotion categories from one another because emotion inductions (such as viewing still images of emotionally significant objects) are insufficiently potent. By contrast, ANS fingerprints are presumed to emerge in the "real world." Therefore, in the lab, when the induction method is sufficiently strong, a unique (specific) pattern of ANS change (i.e., a fingerprint) should emerge for a given emotion category. Since the pattern would be conditioned on the method used in a given study, we might expect to observe the fingerprint less consistently across studies. Therefore, reduced consistency in autonomic nervous system changes (as the result of methodological variations in induction method) separates the "weak" version of the fingerprint hypothesis from the "strong" version (see Tables 1 and 2). A key assumption of all approaches of the classical view, however, is that an emotion category's fingerprint is specific enough to diagnose its instances in a perceiver-independent (i.e., objective) way (see Table 2).

The Constructionist View of Emotion: The Population Hypothesis

The constructionist view of emotion includes social construction theories (e.g., Averill, 1980; De Leersnyder, Boiger, & Mesquita, 2013), psychological construction theories (e.g., Barrett, 2006b; Barrett & Russell, 2015; Cunningham, Dunfield, & Stillman, 2013; James, 1894/1994; Russell, 2003), and descriptive appraisal theories (e.g., Ortony & Clore, 2015), as well as the theory of constructed emotion that integrates social construction and psychological construction, as well as neuroconstructive and rational constructionist perspectives (Barrett, 2017a, 2017c; Barrett & Satpute, 2013; Barrett & Simmons, 2015; Chanes & Barrett, 2016; Lindquist et al., 2012). The constructionist view of emotion is consistent with Darwin's discovery of *population thinking* as articulated in *On the Origin of Species* (Darwin, 1859/2001). In Darwin's view, a biological category, like a species, is a conceptual category populated with heterogeneous individuals. Even though population thinking is considered one of Darwin's great conceptual innovations (Mayr, 2004), it is missing from Darwin's more essentialist writings in *The Expression* (Darwin, 1872/2005); for a discussion of Darwin's writings in this regard, see Barrett, 2017a).[3]

Constructionist theories of emotion hypothesize that an emotion category, as a biological category, is a *conceptual category* (i.e., perceptually variable instances are treated as similar for some function or purpose; Barrett, 2013, 2017a, 2017c).[4] Instances of the same emotion category have variable ANS patterns. The ANS variation is not random, but is meaningful and functional because behaviors and actions vary from situation to situation during the same emotion (i.e., ANS activity during an emotion is situated in support of the specific action in a specific context, Barrett, 2006a). Peripheral physiological changes are tied to the metabolic demands associated with action (e.g., cardiosomatic coupling; Obrist, Webb, Sutterer, & Howard, 1970) or anticipated action (e.g., suprametabolic activity; Obrist, 1981; Sterling, 2012; Turner & Carroll, 1985), so that, for example, crying, shouting, smiling, freezing, and laughing in anger will each be supported by a distinct pattern of ANS change. Simply put, substantial variation in ANS patterns within an emotion category is a feature; it should be expected because it confers evolutionary advantage. It is inherent to the nature of emotion and should be a focus of scientific discovery and explanation. Variation is not a bug to be explained away as error or designated as epiphenomenal to the nature of emotion.

[3] Before *Origin*, a species was defined as a physical type (i.e., with a set of unchanging physical characteristics or features passed down through the generations). This typological characterization fundamentally underestimates within-category variation (in its phenotypic and genetic features) and overestimates between-category variation (Gelman & Rhodes, 2012; Mayr, 2004). One of Darwin's greatest theoretical innovations in *Origin* (Darwin, 1859/2001) was to revolutionize the concept of a *species*, characterizing it as a biopopulation of highly variable individuals (instead of as a group of creatures who share a set of co-occurring biological features; Mayr, 2004). Since Darwin, the concept of a "species" has been characterized by what category members do (i.e., functionally), rather than on the basis of a shared gene pool or a set of physical features. For example, a species is sometimes defined as a reproductive community, and members of different species can be virtually indistinguishable from one another in their physical features but they do not interbreed. Fundamentally, this translates into the insight that a biological category (a "species") is a *conceptual* category, rather than a typological one: a species is a population of physically unique individuals whose similarities are defined functionally, not physically. Believing that a species is a physical type (i.e., an unchanging, biological constant) was a major barrier to discovering evolution and natural selection (Mayr, 2004) and continues to be a major barrier to understanding how evolution and natural selection work (Gelman & Rhodes, 2012).

[4] The classical and construction views do not differ in whether emotion categories are considered biological categories. They differ in terms of what biological processes are relevant to the category—ANS (or neural) patterns (the classical view) or the biology of meaning-making that give ANS changes their emotional functions (the constructionist view; Barrett, 2017a, 2017c).

Table 2
Predictions for Fingerprint and Population Hypotheses

Hypothesis	Consistency	Variation	Specificity	Moderators	Theories
Emotion Fingerprint Hypothesis					
Strong Version	Significant mean change within an emotion category in the *predicted* direction across studies	Small standard deviation	A pattern that is unique to a single emotion category	The pattern of results will be the same across contexts as long as: (a) the emotion was potently induced and (b) nothing impeded the expression of the emotion (e.g., display rules or emotion regulation).	e.g., Ekman, 1972; Ekman & Davidson, 1994; Ekman & Cordaro, 2011; Roseman, 1996, 2001, 2011
Weak Version	Significant mean change within an emotion category in *any* direction across studies	Small standard deviation	A pattern that is unique to a single emotion category	Same as the strong version.	e.g., Levenson, 1992, 2011, 2014
Emotion Population Hypothesis	Significant change within an emotion category across studies is possible, but substantial variation within categories could result in no significant changes.	Large standard deviation	Either of the following: (a) no unique pattern to a single emotion category, (b) a specific pattern that is an abstract, statistical representation of the category (i.e., does not appear in every instance of the category; exists in the presence of a high standard deviation).	Autonomic features are linked to situated action. Because very few studies explicitly manipulate the features of the context (and most studies are attempting to induce what the classical view presumes is the prototypic instance of emotion), moderators (other than for induction quality) are not expected to significantly account for much variance.	e.g., Barrett, 2006b, 2013, 2017a, 2017c; Barrett, Wilson-Mendenhall, & Barsalou, 2015; Clore & Ortony, 2008; Russell, 2003

Scientists often mistake the theory of constructed emotion as a "dimensional" approach, meaning that emotions can be reduced to general affective dimensions of valence and arousal (cf., Barrett, 2015a).[5] Affective dimensions are indeed a low dimensional space for describing emotions, but no constructionist theory of emotion has ever claimed that these two properties alone provide a *sufficient* account of all the important differences among emotion categories. Constructionist hypotheses differ from classical view hypotheses not because one is dimensional and the other is categorical. When it comes to the issue of variability, the two approaches differ in one basic way: Is there limited ANS variation around a fingerprint (the classical view) or substantial variation that is meaningfully tied to the situation (the constructionist view)? What holds emotional instances together in a category is not their ANS similarities; it is the fact that a variety of ANS changes and actions can be used in the service of the same emotion-related goal, because each one is situated to the requirements of a specific situation.

The idea that biological categories are also conceptual categories holds insights for the science of emotion. First, if we discard

[5] The dimensions themselves can differ across theories and have been referred to as hedonic valence and physiological arousal (Barrett & Bliss-Moreau, 2009; Russell, 1980; Russell & Barrett, 1999), positive and negative activation (Watson & Tellegen, 1985; Watson, Wiese, Vaidya, & Tellegen, 1999), positive and negative affect (Cacioppo & Berntson, 1999; Cacioppo, Gardner, & Berntson, 1999), approach and withdrawal (Davidson, 1992; Lang, Bradley, & Cuthbert, 1990; Lang & Davis, 2006), or tense and energetic activation (Thayer, 1989).

the notion of a category as necessarily representing a physical type, then an emotion category can still be a biological category with neither specific nor consistent ANS patterns. If a single ANS pattern exists for an emotion category, it is a *stereotype* or a *statistical summary* and need not be found in nature (by analogy, see Mayr, 2004 for a similar discussion of the "species" concept). The constructionist definition of a category is consistent with recent evidence from cognitive psychology: A category is a group of instances that are treated as similar for some goal or function (Murphy, 2002) and the prototype of a category is the ideal instance that best suits the function or goal, whether or not it exists in nature (e.g., Barsalou, 1993; Voorspoels, Vanpaemel, & Storms, 2011; for a discussion linking the species concept to modern cognitive science research on concepts and categories, see Barrett, 2017a). By this same logic, a constructionist view considers an emotion category as a group of highly variable instances that are treated as similar in the service of some goal, but whose ANS features can vary from instance to instance in a situated way (as a person's brain tailors their physical responses to the environment in the service of that goal). The evolutionary significance of population thinking and variation is supported by computational investigations in modern evolutionary biology (e.g., Whitacre, 2010; Whitacre & Bender, 2010; Whitacre, Rohlfshagen, Bender, & Yao, 2010).[6]

Second, variation in ANS activity is not only expected in different situations, but it can also be observed in instances of an emotion category within the same situation because of the biological reality of *degeneracy*. Degeneracy refers to the capacity for biologically dissimilar systems or processes to give rise to an identical function in the same situation (Edelman & Gally, 2001; Marder & Taylor, 2011; Tononi, Sporns, & Edelman, 1999). Degeneracy is a property of all biological systems, from the systems inside cells to the entire organism (for examples, see Edelman & Gally, 2001). For the purposes of this article, degeneracy refers to the population of unique ANS configurations for instances belonging to the same emotion category within the same situation.[7]

The constructionist hypothesis, then, is that there will be *low consistency* in the ANS patterns observed across different instances of each emotion category as well as overlap in the ANS features across emotion categories (*low specificity*). Even when a specific pattern can be derived for each emotion category using pattern classification methods, for example, this abstract, statistical representation will not necessarily be observed in any specific instance of the category (as we recently demonstrated with a mathematical simulation of brain imaging studies, see Clark-Polner, Johnson, & Barrett, 2016).[8] By analogy, the average middle class American family consisted of 3.13 people in 2014,[9] but no actual family has 3.13 members.[10]

Prior Empirical Tests of the Emotion Fingerprint Hypotheses

The autonomic nervous system is comprised of the sympathetic and parasympathetic branches, the activation of which are generally associated with utilizing energy resources (e.g., fight or flight) or replenishing energy resources (e.g., rest or digest), respectively. The enteric nervous system, that, in addition to the sympathetic and parasympathetic branches, controls gastrointestinal function, is a third component of the ANS. Because there is little data on enteric function in emotion, however, we will not consider it further here (Berntson, Cacioppo, & Quigley, 1991; Stemmler, 2004).[11] In emotion research, the most commonly assessed indices of ANS activity are derived from cardiovascular, respiratory and electrodermal (i.e., sweat gland) responses (see Table 3 for measurements, abbreviations, and definitions). Tests of the emotion fingerprint hypothesis using ANS measures have been conducted in two ways, either by quantitative reviews of the research literature coupled with univariate meta-analytic techniques or by pattern classification analyses of single-experiment data to search for patterning in the ANS responses associated with each emotion category. We review these briefly.

Prior meta-analyses of emotion studies. Recent qualitative reviews have made a case for the existence of emotion fingerprints

[6] In fact, for species, unique individuals are desirable and natural selection operates because individuals vary (Whitacre, 2010; Whitacre & Bender, 2010; Whitacre, Rohlfshagen, Bender, & Yao, 2010).

[7] In evolutionary biology, it is well known that natural selection favors systems with degeneracy. Such systems are high in complexity; they represent a lot more information with greater efficiency than do modular systems (like those proposed by the classical view) and they can create new information by generative combination, allowing them to quickly adapt to changing environmental demands (Whitacre & Bender, 2010; Whitacre, Rohlfshagen, Bender, & Yao, 2010).

[8] Using a mathematical simulation, we recently showed that a pattern which successfully distinguishes the brain maps from neuroimaging studies of one emotion category from the brain maps from neuroimaging studies of other emotion categories (with an accuracy greater than that which might be expected by chance) does not produce the brain state (i.e., the neural "fingerprint" or essence that is found in every instance, or even any, instance of the category. Pattern classifiers work on the logic of population thinking, where the pattern that correctly diagnoses category members is an abstract, statistical summary (Clark-Polner et al., 2017).

[9] U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplements, 1955, 1960, 1965 and 1970 to 2014.

[10] In past research, any category-specific pattern observed has been mistakenly interpreted as evidence for the classical view (Clark-Polner, Johnson, & Barrett, 2016; Clark-Polner, Wager, Satpute, & Barrett, 2017). Scientists often misunderstand population thinking (Kragel & LaBar, 2013, 2015; Rainville, Bechara, Naqvi, & Damasio, 2006; Saarimäki et al., 2016), drawing essentialist conclusions which can interfere with understanding of evolution, natural selection, and scientific inference more generally (Gelman & Rhodes, 2012). Population thinking may be more easily understood by recalling basic statistical concepts about distributions in which a population of responses is characterized by a mean and variance. The mean only describes most individual responses within the distribution if the variance is small. If the variance is large, then the mean is an abstract summary of the population that does not really convey much information about individual cases, and does not allow you to predict individual cases very well. Pattern classification uses algorithms to develop patterns that distinguish categories from one another, but correct classification of an instance requires only that it is statistically closer to one pattern than another. Every instance of anger, for example, can be correctly classified using an abstract ANS pattern but still possess none of the changes evident in that pattern (i.e., the ANS pattern is not *the* body state for anger and it may not even be *a* body state for anger).

[11] Some theorists (Levenson, 2011, 2014) suggest that measurement of the enteric nervous system (ENS) is integral for bodily fingerprints of emotion to fully emerge. Unfortunately, activity of the ENS is extremely difficult to measure in the laboratory and the gastrointestinal system measures available (e.g., electrogastrogram) are also subject to parasympathetic control, making it difficult to disentangle parasympathetic and enteric effects.

Table 3
Autonomic Nervous System Dependent Variables

Measurement	Body system	Measurement system (units)	Description
Cardiac Output (CO)	Cardiovascular	Cardiac Impedance & ECG (liters, minute)	Amount of blood circulated in the body per unit of time (liters, minute).
Diastolic Blood Pressure (DBP)	Cardiovascular	Auscultatory, Oscillometric, or arterial tonometric, most common (mmHg)	Lowest pressure exerted by circulating blood on the walls of blood vessels during each cardiac cycle.
Finger Pulse Amplitude (FPA)	Cardiovascular	Photoplethysmograph (arbitrary units)	Height or amplitude of the pulse waveform detected in the finger. Measure of dilation, constriction of the blood vessels in the finger.
Finger Pulse Volume (FPV)	Cardiovascular	Photoplethysmograph (ml)	Change in the volume of blood in the finger with each heart beat.
Heart Rate (HR)	Cardiovascular	ECG (beats, minute)	Number of beats per unit of time.
Heart Rate Variability (HRV)	Cardiovascular	Derived from the heart period from ECG (units vary by method)	Variation in heart period (or rate) as a function of central respiratory drive or peripheral respiratory afferent input. The measure here is specifically the high-frequency HRV; also called respiratory sinus arrhythmia (RSA).
Interbeat Interval, Heart Period (IBI, HP)	Cardiovascular	ECG (ms)	Time between heart beats (inverse of heart rate).
Mean Arterial Pressure (MAP)	Cardiovascular	Auscultatory, Oscillometric, or arterial tonometric, most common (mmHg)	Average arterial pressure during a cardiac cycle.
Pulse Transit Time (PTT)	Cardiovascular	Photoplethysmograph (ms)	Time between contraction of the left ventricle and the highest or systolic point in the pulse wave.
Pre-ejection Period (PEP)	Cardiovascular	Cardiac Impedance & ECG (ms)	The time interval from the beginning of electrical stimulation of the heart to the opening of the aortic valve.
Stroke Volume (SV)	Cardiovascular	Cardiac Impedance & ECG (ml)	Volume of blood ejected from the heart with each beat.
Systolic Blood Pressure (SBP)	Cardiovascular	Auscultatory, Oscillometric, or arterial tonometric, most common (mmHg)	Peak pressure exerted by circulating blood on the walls of the blood vessels during the cardiac cycle.
Total Peripheral Resistance (TPR)	Cardiovascular	Cardiac Impedance & ECG (units vary by method)	Sum of the resistance of all peripheral vasculature in the systemic circulation.
Skin Conductance Amplitude (SCA)	Electrodermal	Average μSiemens	Measure of the change in electrical conductance of the skin due to additional sweat in the eccrine ducts. Reported as the amplitude of response (during a skin conductance response or SCR) over a period of time (determined by researcher). Amplitude is the average of all nonzero SCRs.
Skin Conductance Magnitude (SCM)	Electrodermal	Average μSiemens	Measure of the change in electrical conductance of the skin due to additional sweat in the eccrine ducts. Reported as the magnitude of response (during a skin conductance response or SCR) over a period of time (determined by researcher). Magnitude is the average of all potential SCRs, even if they are zero.
Skin Conductance Level (SCL)	Electrodermal	Average μSiemens	Measure of the electrical conductance of the skin per unit of time due to higher amount of sweat in the eccrine ducts.
Skin Conductance Responses (SCRs)	Electrodermal	Number of responses over a period of time (determined by researcher)	Measure of the change in electrical conductance of the skin due to additional sweat in the eccrine ducts. Reported as numbers of responses over a period of time (e.g., responses per minute).
Temperature (Temp)	Other		
Expiratory Time (TE)	Respiratory	Respiration belt or inductive plethysmography (sec)	Average exhalation time during a respiratory cycle.
Inspiratory Time (TI)	Respiratory	Respiration belt or inductive plethysmography (sec)	Average inhalation time during a respiratory cycle.
Respiratory rate (RR)	Respiratory	Respiration belt or inductive plethysmography (cycles, minute)	Number of breaths. (inhalation, exhalation cycles) taken within a period of time (usually minutes).
Tidal Volume (TV)	Respiratory	Respiration belt or inductive plethysmography (mL)	Volume of air moved into or out of lungs during quiet breathing.

Note. The measures in bold were included in our meta-analysis because they were sufficiently powered across studies ($k > 2$ within one emotion category). Because heart rate (HR) and interbeat interval (IBI) are inversely related, we switched the direction of the effect sizes when IBI was extracted (instead of HR). If both IBI and HR were reported, we extracted the HR data (for full descriptions of each ANS measure, see Berntson, Cacioppo, & Quigley, 1991; Cacioppo et al., 2000; Larsen, Berntson, Poehlmann, Ito, & Cacioppo, 2008).

in ANS responding (Friedman, 2010; Kreibig, 2010),[12] although quantitative (i.e., meta-analytic) reviews tell a different story. Prior univariate meta-analytic reviews have found limited evidence for either consistency or specificity in ANS responding during emotion. For example, Cacioppo et al. (2000, 22 studies) concluded that there was relatively little evidence of emotion specificity across the studies they examined, and noted low consistency across studies in ANS responses (compared with baseline) and sufficient variability that one could not interpret the findings as evidence for fingerprints. Stemmler (2004) also assessed consistency and specificity, specifically for ANS patterns of anger and fear categories (eight of which also appeared in the Cacioppo et al., 2000 meta-analysis), and found some evidence of consistency but also uncovered substantial variation. Both variation (Cacioppo, Berntson, Klein, & Poehlmann, 1997; Cacioppo et al., 2000) and context (Stemmler, 2004) emerged as main features of these meta-analyses. Their findings showed weak support for the emotion fingerprint hypothesis and are more consistent with the emotion population hypothesis.[13] See supplemental materials, Appendix A for a comparison of prior meta-analyses, including ANS effects within emotion categories (see also, Quigley & Barrett, 2014).

Multivariate pattern classification. Pattern classification and, in particular, multivariate pattern classification (MPCA), is considered by some to offer a better test of the fingerprint hypothesis on both theoretical and statistical grounds (e.g., Friedman, 2010). Theoretically, the classical theories of emotion have hypothesized that each emotion category is associated with a unified (i.e., specific) pattern of change across ANS systems. Statistically, multivariate pattern classification allows researchers to examine multiple ANS features simultaneously without violating the dependency assumptions that underlie the General Linear Model. A growing number of studies use MPCA to investigate distinctions between different emotion categories based on features of behavior, ANS activity, and brain activity, but they typically do not draw any specific conclusions about how or why emotion categories differ from one another (e.g., Baucom et al., 2012; Kragel & LaBar, 2013; Nyklíček, Thayer, & Van Doornen, 1997; Park, Jang, Chung, & Kim, 2013; Petrantonakis & Hadjileontiadis, 2010; Rainville, Bechara, Naqvi, & Damasio, 2006; Sitaram et al., 2011; Stephens, Christie, & Friedman, 2010; Takahashi, 2004).[14]

Many of these authors interpreted their findings as support for the ANS emotion fingerprint hypothesis (Kragel & LaBar, 2013; Nyklíček, Thayer, & Van Doornen, 1997; Stephens, Christie, & Friedman, 2010). However, pattern classification, even when successful, does not reveal *a single* body state for an emotion category. For statistical reasons, it is best interpreted in a way that is consistent with the emotion population hypothesis, as a stereotype or abstract statistical summary that itself need not exist in nature (for a discussion, see Clark-Polner, Johnson, & Barrett, 2016; Clark-Polner, Wager, Satpute, & Barrett, 2017). Furthermore, the multivariate patterns themselves do not replicate across studies, even when they use the same stimuli (e.g., music and films) and experimental methods. For example, Stephens et al. (2010) and Kragel and LaBar (2013) identified multivariate ANS patterns that successfully distinguished one emotion category from another within their experiments, but the patterns were not consistent with each other, across experiments. Because these experiments used identical induction methods and stimuli, it is difficult to claim that the variation observed is epiphenomenal (e.g., due to the methods).

Summary. Taken together, previous meta-analyses and pattern classification analyses explicitly designed to search for ANS fingerprints of emotion instead uncovered evidence more consistent with the emotion population hypothesis. If science is the quantification of doubt (Gee, 2014) then these findings call into question the emotion fingerprint hypothesis and highlight the need for an analysis strategy to specifically test the emotion population hypothesis.

Primary Research Aims

In this article, we used multilevel meta-analysis, combined with MPCA, to test whether the growing body of research on ANS changes during instances of emotion is more consistent with the emotion fingerprint hypothesis (associated with the classical approach to emotion) or with the emotion population hypothesis (associated with our theory of constructed emotion). We first used multilevel meta-analysis to investigate whether there are *consistent* and *specific* ANS features that correspond to each emotion category or whether significant variation in ANS responses within a category is normative. We next examined whether meta-analytic multivariate pattern classification analysis (MPCA) would allow us to identify a ANS pattern that distinguish one emotion category from another. In a final analysis, we investigated whether methodological moderators could explain any of the within-category variability in ANS features. Evidence for one theory or another does not rest on whether differences between categories are found, but on the nature of the differences, as well as on the degree of within-category variation.

[12] Kreibig (2010) reported a review of 134 studies summarizing physiological responses during 16 different emotion categories (see her Table 2) and concluded that there was strong evidence of ANS fingerprints of emotion. She did not report a meta-analysis but rather an un-weighted "modal" response. Crucially, there was no adjustment for study sample size or other quality indicators, and no quantitative definition of "modal" response. Standard meta-analytic practices (e.g., estimating the variances among sample sizes) were not reported. Furthermore, this review did not emphasize comparisons between emotions and could not definitively answer whether there is a consistent and specific ANS response pattern for each emotion. In addition, Kreibig (2010) suggested there might be "subforms" of emotions like sadness or disgust. For example, she proposed a contamination-related "core" disgust and a mutilation-related "body-boundary" disgust. This terminology raises the theoretical question of whether these proposed sub-forms are considered different emotion categories (e.g., Scarantino, 2015) or instead contextually-based variations of the same "basic" emotion "family." We assume it is the former, because Kreibig (2010) reported observing distinct ANS patterns.

[13] Lench, Flores, and Bench (2011) conducted a meta-analysis that examined physiological outcomes across emotion categories. In their analysis, physiological effect sizes were not reported for individual ANS measures. They were combined into an overall physiological outcome effect size. This makes it impossible to assess patterns in physiology across emotion categories (for additional discussion of this meta-analysis, see Lindquist, Siegel, Quigley, & Barrett, 2013). The Lench et al. (2011) meta-analysis did provide an important contribution to the literature on emotion in their evaluation of the influence of individual induction procedures (see Table 4, p. 14 in Lench, Flores, & Bench, 2011).

[14] Other prior studies used similar approaches to distinguish between just two emotions and a neutral condition (Kolodyazhniy et al., 2011; Kreibig, Wilhelm, Roth, & Gross, 2007; Sinha, 1996) or to distinguish between a small number of laboratory-induced stressor states (Allen, Boquet Jr., & Shelley, 1991).

It is important to note that our ability to directly test the population hypothesis (i.e., that variation is meaningfully tied to the context as part of an emotional episode, or that degeneracy exists within the same context) is limited by the published studies available for meta-analysis. Few studies utilize biological measurements to assess metabolic demands or systematically manipulate the fine-grained features of the context (Stemmler, 2004; Stemmler, Aue, & Wacker, 2007; Wilson-Mendenhall, Barrett, & Barsalou, 2015; Wilson-Mendenhall, Barrett, Simmons, & Barsalou, 2011). No study that we know of examines the existence of ANS degeneracy for a given emotion category within the same situation. Our meta-analysis provides the most definitive test, to date, of substantial and reliable variation in ANS responses within a single emotion category, particularly under experimental conditions that were designed to minimize such variation. Our findings support the hypothesis that ANS variation is not epiphenomenal to the nature emotion. But our ability to fully understand the nature of ANS variation (in relation to varying metabolic and situational demands) is necessarily limited by the literature available.

Method

Selection of Studies

Erika H. Siegel, Molly K. Sands, and Paul Condon authors served as primary coders who, along with three trained undergraduate research assistants, searched multiple databases for published research on ANS responses during emotion (including Web of Knowledge, PsychInfo and Google Scholar). Searches included English-language articles published from 1950 through December 2013 (see, Figure 1). We searched broad and narrow variations of terms related to emotion, including each of six emotion categories (e.g., "emotion," "happiness," "anger") and physiological responses (e.g., "physiology," "autonomic nervous system," "cardiac," "blood pressure;" contact the corresponding author for a complete list of search terms). This yielded a starting database of 1,203 source articles whose abstracts were then reviewed for initial inclusion by three trained research assistants (who were not the primary coders) to assess whether each article met our general search criteria. In addition, we placed a call for unpublished research using electronic listservs, including the Society for Personality and Social Psychology. We also sent specific electronic requests for unpublished research to ~100 authors who already had studies included in our database and were experts in the psychophysiological measurement of emotion. We received responses from four research groups who were willing to share unpublished data (one research group later changed their mind). Unfortunately, none of the three remaining unpublished data sets met our specific search criteria.

The primary coders then completed a second, more detailed review of the remaining studies to determine inclusion in the meta-analysis. Articles had to meet the following criteria to be included: (a) the study manipulated emotion in the laboratory (e.g., no experience sampling or ambulatory physiological studies were included); (b) the study reported at least one peripheral physiological measure as a dependent variable; (c) the study reported data from a nonclinical, adult sample;[15] (d) the study included a baseline comparison condition for assessing ANS change (because change from baseline in each ANS measure constituted our "effect size"); and (e) participants were not instructed to regulate or change their emotional experience during the induction procedure (e.g., we excluded studies where participants were instructed to regulate their response during an anger induction). All articles were included unless they clearly did not meet search criteria (e.g., the sample comprised children, there was no emotion manipulation). Six-hundred and 44 individual studies (from 641 full-text articles) were reviewed and coded by the primary coders, 440 individual studies were eventually excluded during a second review process, leaving 204 unique studies from 195 full-text articles (Table 4; note that several articles report data from more than one emotion category). The most common reasons for exclusion after coding were not being able to extract an effect size (195), the emotion was compared with a neutral condition (vs. a baseline, 183), the induction included emotion regulation as part of the task (43), no emotion was manipulated (10), participants' ANS physiology was manipulated by something other than an emotion induction (e.g., a shot of norepinephrine or running on a treadmill [six]), or participants were physically posed and/or instructed how to feel during an emotion induction (three).

Data Extraction

We used a coding procedure designed by the primary coders and senior authors. Articles were randomly assigned to a primary coder and independently coded. For articles with multiple studies from different samples of participants, we coded each study separately. See supplemental materials, Appendix B for a list of the moderators that were coded for each study and how each moderator was operationalized and supplemental materials, Appendix C for a detailed description of induction procedures.

The primary coders conducted the initial coding and senior authors served as secondary reviewers for more difficult cases (e.g., interpreting less commonly used ANS measures or providing details about unfamiliar emotion induction procedures). Final coding decisions were made by consensus in difficult or ambiguous cases. To assess interrater reliability, all three primary coders coded a little over 10% ($N = 67$) of the 644 coded studies. We calculated an intraclass correlation coefficient for continuous variables and a kappa coefficient for categorical variables. The intraclass correlation coefficient for the two continuous variables (mean age and percent male) was 1.0. The kappa coefficients for the subjective moderators (rigor of emotion induction procedure and rigor of physiological recording) were 0.85 and 0.94, respectively. The kappa coefficient for all categorical variables was 1.0. We contacted study authors when there was missing information (e.g., no mean reported for the baseline or we were unable to tell if participants were videore-

[15] We excluded studies on older adults since autonomic nervous system changes that occur in later adulthood would make it difficult to interpret our results if data from older individuals were included (see, Lipsitz, 2004; Mendes, 2010; Pfeifer et al., 1983; Straub, Miller, Schölmerich, & Zietz, 2000). We did include studies in which older adults were part of a broader sample (e.g., where the age range for the study was 21–68 years of age). Also, we considered samples to be "clinical" (and thus we excluded them) if participants were either: (a) diagnosed by a clinician or (b) the mean group outcome on a self-report survey (e.g., the Beck Depression Inventory) indicated a clinically significant mental health issue.

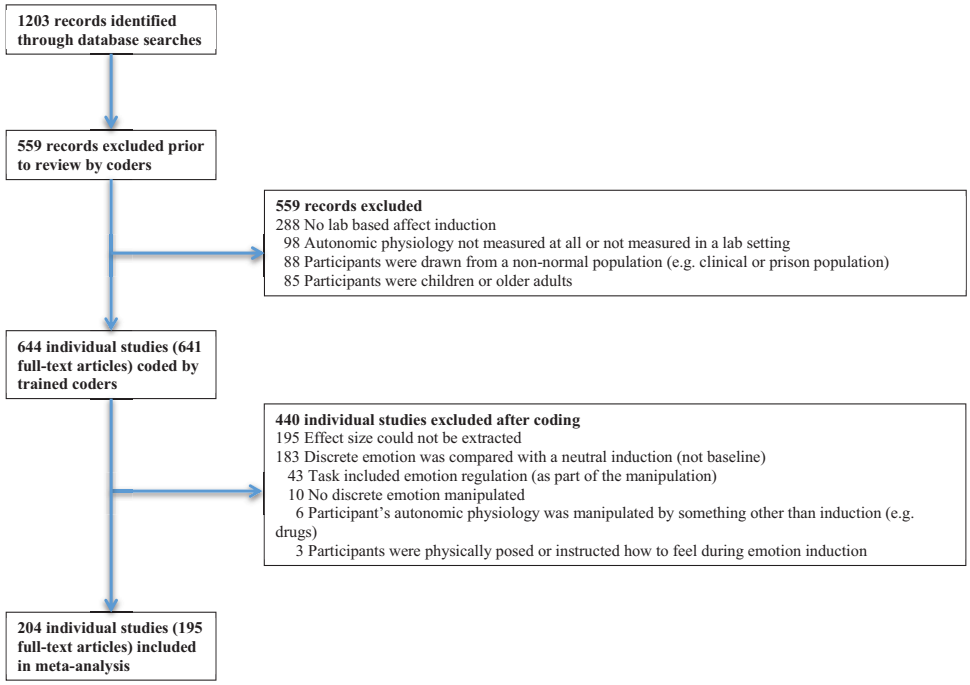

Figure 1. Flowchart describing identification and screening of articles and studies. See the online article for the color version of this figure.

corded). Of the 103 authors we contacted, we received 69 responses. Of those, 33 were able to provide information, resulting in an additional 122 effect sizes.

Emotion Category Designation

When assigning each study to an emotion category, we used the authors' designations whenever possible. For example, if the authors indicated that they were manipulating fear, we coded it as fear. When different authors referred to a similar task as a different emotion (e.g., one said it was a "fear" induction, the other an "anger" induction), we assigned the effect sizes according to the authors' designation, because it is possible to modify an experimental procedure slightly to achieve different effects. If the authors labeled an induction procedure as "stress," and the induction procedure was not obviously associated with any emotion category (e.g., a social stress task in which a participant completed a cognitive load task with another participant), the article was excluded from the meta-analysis. If the authors labeled an induction procedure as "stress" (i.e., the author did not assign an emotion category label) but used a task that was consistently and obviously labeled by other researchers as one (and only one) emotion category (e.g., a social evaluative task with harassment), then we assigned it to that emotion category (in this case, "anger"). This resulted in the following numbers of effect sizes from nine emotion categories (plus neutral): anger (286), fear (111), happy (99), sad (87), disgust (51), surprise (eight), excitement (two), awe (one), shame (one), and neutral (105). Of these categories, only anger, fear, happy, sad, disgust, and neutral were sufficiently powered to be included in the meta-analysis (for a total of 739 effect sizes included in our analyses, see Table 4).

Table 4
Number of Coded Effect Sizes by Emotion Category

Category	Effect sizes	Studies	Participants
Anger	**286**	**76**	**11,820**
Disgust	**51**	**16**	**2,651**
Fear	**111**	**47**	**4,373**
Happy	**99**	**26**	**2,899**
Sad	**87**	**29**	**3,059**
Neutral	**105**	**53**	**3,976**
Awe	1	1	37
Excitement	2	2	51
Shame	1	1	99
Surprise	8	3	73

Note. Categories in bold were sufficiently powered to be included in our meta-analyses. Summing the number of bolded studies in this table for each category included in the analysis results in 247 total. This differs from the total number of studies reported in Figure 1 ($N = 204$) and in the text because several papers report data from more than one emotion category.

Outcome Variables

Our meta-analyses included effect sizes from 20 ANS measures (for a complete list, see Table 5; for descriptions of the measures, see Table 3). Table 5 includes a list of effect sizes for each ANS measure and Figure 2 presents the number of effect sizes for each ANS measure and each emotion category. We excluded articles using facial electromyography (fEMG) because these measures do not capture autonomic nervous system

Table 5
Number of Coded Effect Sizes by Autonomic Nervous System Variable

Body system	Autonomic nervous system measurement	Effect sizes	Studies	Sample size
Cardiovascular	Heart Rate (HR)	244	110	10,050
	Heart Rate Variability (HRV)	36	18	1,958
	Pulse Transit Time (PTT)	3	3	174
	Cardiac Output (CO)	12	4	679
	Stroke Volume (SV)	8	5	386
	Total Peripheral Resistance (TPR)	12	5	679
	Diastolic Blood Pressure (DBP)	107	49	4,219
	Systolic Blood Pressure (SBP)	111	51	940
	Mean Arterial Pressure (MAP)	2	2	82
	Temperature (TEMP)	6	3	232
Respiratory	Respiratory Rate (RR)	33	16	835
	Expiratory Time (TE)	6	2	122
	Inspiratory Time (TI)	11	2	202
	Tidal Volume (TV)	3	2	73
Electrodermal	Skin Conductance Level (SCL)	87	46	3,750
	Skin Conductance Responses (SCR)	58	32	1,750

Note. If we sum the number of studies for each autonomic nervous system variable in this table we get 350. This differs from the number of total studies reported in Figure 1 ($N = 195$) because most studies include data from more than one Autonomic Nervous System (ANS) variable.

activity and because the two most frequently measured muscles: corrugator supercilii and zygomaticus major are generally considered to correspond to changes in hedonic valence (pleasantness and unpleasantness) rather than to specific emotion categories (Tassinary, Cacioppo, & Vanman, 2007).[16]

Preliminary analyses revealed that effect sizes comparing ANS responses during an emotion induction versus a neutral induction (e.g., participants watched a "fear" film and a "neutral" film with the neutral induction used as the comparison, i.e., fear induction-neutral induction) produced systematically different results from effect sizes comparing an emotion induction to a preinduction baseline (e.g., fear induction-baseline). Thus, the two types of effect sizes could not be included in the same analysis. We opted to use effect sizes for a neutral condition relative to a baseline (neutral induction-baseline, Table 4, Figure 2), so that we could compare changes from baseline for both emotion inductions and neutral inductions.

Effect Size Extraction

We extracted data from ANS measures obtained during emotion inductions and baseline periods to calculate the effect size which represents the standardized mean change from baseline (Lipsey & Wilson, 2001). Mean change from baseline was our only effect size statistic. We did not include effect sizes comparing two inductions directly (e.g., fear vs. disgust) for two reasons. First, using a change score from baseline to task is standard in psychophysiology because it adjusts for individual differences in baseline physiology and provides a better measure of how physiology changes during a task for a given person. Second, extracting effect sizes in this way meant that each effect size was a standardized and interpretable metric of change and we could include all of the effect sizes in the same multivariate analysis, thereby allowing us to test for specific fingerprints of emotion categories across all physiological measures.

If means and standard deviations of ANS measures during an induction and a baseline were provided, we first calculated standardized differences (Cohen's *d*; Cohen, 1969) between baseline and induction. If the means, sample standard deviations, and number of participants in the sample were available, we pooled the standard deviations of the mean values and calculated *d*. We then transformed the differences into Pearson product–moment correlation coefficients (Cohen, 1969, 1988, 1992; Rosenthal, 1991; Rosenthal, 1994). For ease of interpretation, mean effect sizes then were converted back to *d* from *r*.

If means and standard deviations were not reported in the original study, then effect sizes were calculated from statistical tests (typically *t* tests or *F* tests with one degree of freedom) that compared the ANS measure during baseline to that during the emotion induction or used *p* values and degrees of freedom. In some instances, we were able to extract data by hand from clearly labeled graphs that included standard errors or standard deviations. If a study reported that a comparison was not statistically significant or did not provide specific *p* values or significance tests, we attempted to contact the author and, if unsuccessful, we excluded that study from the analysis.

Calculating Standard Deviations of the Difference

When the standard deviation of the mean was reported for both the baseline and induction, we calculated standard deviations by pooling the standard deviations using the formula, $s_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2) - 2}}$, where s_1 represents the standard deviation of the mean of the baseline condition and s_2 represents the standard deviation of the mean of the emotion condition (n_1 and n_2 are the sample sizes for the baseline and emotion conditions, respectively). Fifty-three percent of the included effect sizes were calculated using this method (in which coders calculated pooled

[16] There is one exception. Some authors have suggested that the levator labii superioris may be activated preferentially during experiences of "disgust" versus other discrete emotions (e.g., Vrana, 1993; Whitton, Henry, Rendell, & Grisham, 2014). Unfortunately, we were unable to identify a sufficient number of studies that measured levator labii activity during other discrete emotion states to permit comparison.

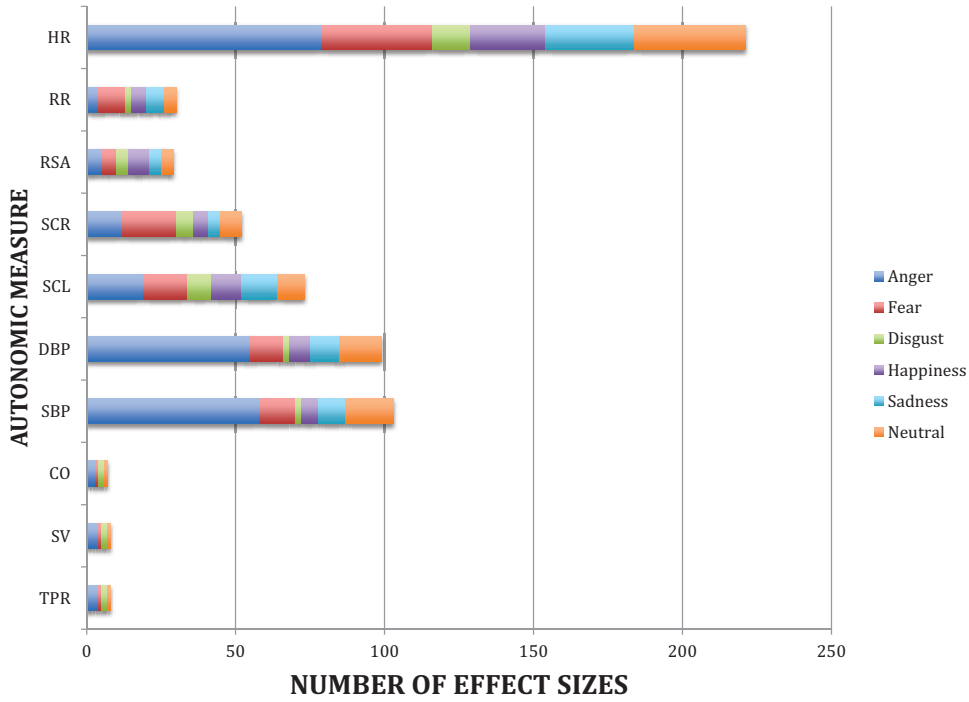

Figure 2. Number of effect sizes for each autonomic nervous system measure by emotion category. See the online article for the color version of this figure.

standard error). For the remaining effect sizes (47%), both mean differences and standard deviations of the differences were directly reported in the original papers. To be sure that our pooled standard deviation calculation method did not introduce systematic differences in variability, we tested whether calculation method was a significant moderator of effect sizes, and it was not ($p = .21$). See Table 6 for a complete list of studies included in the analysis along with effect sizes and moderator codes.

Meta-Analytic Procedures

Only studies that met all of the inclusion criteria were included in the meta-analysis. We extracted at least one effect size from 195 studies, yielding 739 unique effect sizes. See Tables 4 and 5 for a list of effect sizes by emotion category and ANS measure and Table 6 for the full list of articles included in the analysis. Six separate meta-analyses were performed, one for each emotion category (anger, disgust, fear, happy, sad) and for neutral, using the same procedure for each analysis. Supplemental materials, Appendix B provides an overview and description of the moderators included in these analyses. Restricted maximum likelihood estimation was used to estimate the analyses, as implemented in the SAS MIXED procedure (Littell et al., 2006).

To compare effect sizes across studies, we transformed extracted effect sizes into Fisher's *z* (Cohen, Cohen, West, & Aiken, 2003). Effect sizes were then weighted by the inverse of their variance (i.e., sampling error), so that larger studies contributed more to the aggregate effect size estimate than smaller studies (Lipsey & Wilson, 2001; Raudenbush & Bryk, 2002).

Multilevel models. For each meta-analysis (one for each emotion category), we used a three-level meta-analytic technique for dependent effect sizes described by Van den Noortgate and colleagues (2013, 2015, 2003). Each study (denoted by the subscript *k*) may have one or more outcomes (denoted by the subscript *j*) measured in a sample of participants. A three -level meta-analysis accounts for possible variation at each of these three levels: Observed effect sizes may vary due to differences between studies (level 3), due to differences between outcomes studied within the same study (level 2), and due to sampling variation (level 1). Two of these variances are present in a traditional meta-analysis: (a) *sampling variance* (i.e., differences between observed effect sizes and population effect sizes); and (b) *between-study variance* (i.e., systematic differences between the population effect sizes from different studies). The three-level meta-analysis quantifies a third source of variance in meta-analyses with dependent effect sizes (c) *between-outcome/within-study variance* (i.e., systematic differences between the effect sizes from different outcomes within the same study).

Consistent with other multilevel modeling approaches, we started with a random-effects model (REM), this is a model that decomposes the variance into variance at each of the levels without trying to explain variation by including predictors. The REM reads as:

$$ES_{jk} = \beta_0 + V_k + U_{jk} + e_{jk}$$

in which ES_{jk} is the observed effect size for outcome *j* in study *k*; β_0 is the overall mean effect size across all conditions and studies; V_k refers to the random deviation of the mean effect in study *k* from the overall mean effect size β_0, U_{jk} refers to the deviation of the effect for outcome *j* in study *k* from the mean effect in study *k*; and e_{jk} is the residual due to sampling fluctuations, indicating the deviation of the observed effect size from the population effect size for outcome *j* in study *k*. All three error terms, V_k, U_{jk}, and e_{jk} were assumed to be

Table 6
List of All Citations Included With Effect Sizes and Moderator Codes

						Moderators												
						Sample		Study					Quality					
Citation	Year	*N*	Emotion	ANS	Effect size (Z)	Sample type	Gender	Induction type	Eval	Video recorded	Presence of others	Emotion words	BL duration	No. emotions	Induction duration	Manip. check	Rigor: Induction	Rigor: Physio	Other meta-analysis
Aue, Cuny, Sander, & Grandjean	2010	42	Anger	HR	.11	1		1	0	0	0	1		4	1	1	1		
				SCL	.24	1		1	0	0	0	1		4	1	1	1		
Abel, Larkin, & Edens Group 1	1995	66	Anger	DBP	.14	1	1	5	1	0	1	1	4	2	4	1	1	2	
				HR	.31	1	1	5	1	0	1	1	4	2	4	1	1	2	
				SBP	.19	1	1	5	1	0	1	1	4	2	4	1	1	2	
Abel, Larkin, & Edens Group 2	1995	66	Anger	DBP	.29	1	1	5	1	0	1	1	4	2	4	1	1	2	
				HR	.24	1	1	5	1	0	1	1	4	2	4	1	1	2	
				SBP	.22	1	1	5	1	0	1	1	4	2	4	1	1	2	
Adenauer	2010	19	Neutral	HR	−.01	2	1	3	0	0	0	1		3	1	1	2	2	
Adsett, Schottstaedt, & Wolf Group 1	1962	33	Anger	CO	.19	2	0	4	1	0	1	1	4	1	3	1	3	3	Stemmler
				DBP	.21	1	0	4	1	0	1	1	4	1	4	1	3	3	
				HR	.08	1	0	4	1	0	1	1	4	1	4	1	3	3	
				SBP	.16	1	0	4	1	0	1	1	4	1	4	1	3	3	
				SV	−.05	2	0	4	1	0	1	1	4	1	3	1	3	3	
				TPR	−.02	2	0	4	1	0	1	1	4	1	3	1	3	3	
Adsett, Schottstaedt, & Wolf Group 2	1962	33	Anger	DBP	.16	1	0	4	1	0	1	1	4	1	4	1	3	3	
				HR	.02	1	0	4	1	0	1	1	4	1	4	1	3	3	
				SBP	.11	1	0	4	1	0	1	1	4	1	4	1	3	3	
Allen et al.	1996	98	Neutral	HR	−.02	1	1	1	0	0	1	1	4	3	3	1	1	1	
Anderson, Linden, & Habra	2006	142	Anger	DBP	.65	1	1	4	1	0	1	1	4	1	4	1	3	3	
				HR	.48	1	1	4	1	0	1	1	4	1	4	1	3	3	
Andrews et al.	2007	142	Anger	SBP	.51	1	2	4	1	1	1	1	4	1	4	1	1	3	
Anttonen & Surakka	2009	60	Anger	HR	.18	1	1	3	0	1	1	1	3	4	2	1	2	3	
			Happy	HR	.18	1	1	3	0	1	1	1	3	4	2	1	2	3	
Averill	1969	18	Happy	DBP	.16	1	2	2	0	0	0	1	4	3	4	1	1	2	Cacioppo et al.
				SCR	.25	1	2	2	0	0	0	1	4	3	4	1	1	2	
				HR	.08	1	2	2	0	0	0	1	4	3	4	1	1	2	
				FPV	−.17	1	2	2	0	0	0	1	4	3	4	1	1	2	
				RR	.02	1	2	2	0	0	0	1	4	3	4	1	1	2	
				SBP	−.02	1	2	2	0	0	0	1	4	3	4	1	1	2	
	1969	18	Sad	DBP	.44	1	2	2	0	0	0	1	4	3	4	1	1	2	
				SCR	.24	1	2	2	0	0	0	1	4	3	4	1	1	2	
				FPV	−.18	1	2	2	0	0	0	1	4	3	4	1	1	2	
				HR	.22	1	2	2	0	0	0	1	4	3	4	1	1	2	
				RR	.01	1	2	2	0	0	0	1	4	3	4	1	1	2	
				SBP	.24	1	2	2	0	0	0	1	4	3	4	1	1	2	
	1969	18	Neutral	DBP	.00	1	2	2	0	0	0	1	4	3	4	1	1	2	
				SCR	−.09	1	2	2	0	0	0	1	4	3	4	1	1	2	
				FPV	−.07	1	2	2	0	0	0	1	4	3	4	1	1	2	
				HR	.16	1	2	2	0	0	0	1	4	3	4	1	1	2	
				RR	.00	1	2	2	0	0	0	1	4	3	4	1	1	2	
				SBP	−.05	1	2	2	0	0	0	1	4	3	4	1	1	2	
Ax	1953	16	Anger	DBP	−.06	1	1	5	1	0	1	1	4	2	4	1	1	2	Cacioppo et al.; Stemmler
				SCR	.44	1	1	5	1	0	1	1	4	2	4	1	1	2	
				HR	−.14	1	1	5	1	0	1	1	4	2	4	1	1	2	
				SBP	.03	1	1	5	1	0	1	1	4	2	4	1	1	2	
				SCL	.17	1	1	5	1	0	1	1	4	2	4	1	1	2	
Azevedo	2005	40	Disgust	HR	.06	1	2	3	0	0	0	0	3	3	3	1	1	1	
			Neutral	HR	−.01	1	2	3	0	0	0	0	3	3	3	1	1	1	
Balaban & Taussig	1994	42	Disgust	SCL	.20	1	1	3	0	1	0	1	1	4	1	1	2	2	
			Fear	SCL	.27	1	1	3	0	1	2	1	1	4	1	1	2	3	

(*table continues*)

Table 6 (*continued*)

						Moderators												
						Sample		Study					Quality					
Citation	Year	*N*	Emotion	ANS	Effect size (Z)	Sample type	Gender	Induction type	Eval	Video recorded	Presence of others	Emotion words	BL duration	No. emotions	Induction duration	Manip. check	Rigor: Induction	Rigor: Physio	Other meta-analysis
Baldaro et al.	1996	42	Neutral	HRV	−.06	1	1	2	0	0	0	0	3	2	3	1	1	2	
Baldaro et al. Study 1	2001	40	Disgust	HR	.07	1	2	3	0	0	0	0	3	3	3	1	1	1	
Baldaro et al. Study 2	2001	98	Disgust	HR	−.05	2	1	2	0	0	0	1	2	1	2	1	2	3	
Baumgartner, Esslen, & Janke	2006	24	Happy	HR	.00	1	1	4	0	1	1	1	3	4	2	1	2	3	
				RR	.06	1	1	2	0	0	0	1	2	2	3	1	2	1	
Beckham et al.	2002	53	Anger	DBP	.13	2	2	1	0	0	1	1	4	1	1	1	3	2	
				HR	.02	2	2	1	0	0		1	4	1	1	1	3	2	
				SBP	.05	2	2	1	0	0		1	4	1	1	1	3	2	
Betensky & Contrada	2010	63	Sad	DBP	.33	1	0	4	1	0	1	1	3	1	3	1	1	2	
				HR	.42	1	0	4	1	0	1	1	3	1	3	1	1	2	
				SBP	.50	1	0	4	1	0	1	1	3	1	3	1	1	2	
Blechert et al.	2006	42	Fear	SCR	.45	1	1	6	0	0	0	0	3	1	3	0	2	3	
				HR	.08	1	1	6	0	0	0	0	3	1	3	0	2	3	
				HRV	.01	1	1	6	0	0	0	0	3	1	3	0	2	3	
				RR	.23	1	1	6	0	0	0	0	3	1	3	0	2	3	
				Te	−.10	1	1	6	0	0	0	0	3	1	3	0	2	3	
				Ti	−.11	1	1	6	0	0	0	0	3	1	3	0	2	3	
				Vt	−.09	1	1	6	0	0	0	0	3	1	3	0	2	3	
				SCL	.25	1	1	6	0	0	0	0	3	1	3	0	2	3	
				PTT	.00	1	1	6	0	0	0	0	3	1	3	0	2	3	
Boiten	1996	16	Anger	RR	.00	1	1	3	1	1	1	0	2	6	1	0	2	2	
				Te	−.21	1	1	3	1	1	1	0	2	6	1	0	2	2	
				Ti	−.17	1	1	3	1	1	1	0	2	6	1	0	2	2	
			Disgust	Ttot	.02	1	1	3	1	1	1	0	2	6	1	0	2	2	
				Te	−.06	1	1	3	1	1	1	0	2	6	1	0	2	2	
				Ti	−.07	1	1	3	1	1	1	0	2	6	1	0	2	2	
			Fear	Ttot	−.11	1	1	3	1	1	1	0	2	6	1	0	2	2	
				Te	−.20	1	1	3	1	1	1	0	2	6	1	0	2	2	
				Ti	−.14	1	1	3	1	1	1	0	2	6	1	0	2	2	
			Happy	Ttot	−.24	1	1	3	1	1	1	0	2	6	1	0	2	2	
				Te	−.16	1	1	3	1	1	1	0	2	6	1	0	2	2	
				Ti	−.13	1	1	3	1	1	1	0	2	6	1	0	2	2	
			Sad	Ttot	.04	1	1	3	1	1	1	0	2	6	1	0	2	2	
				Te	−.13	1	1	3	1	1	1	0	2	6	1	0	2	2	
				Ti	−.13	1	1	3	1	1	1	0	2	6	1	0	2	2	
Bongard	1997	16	Anger	DBP	−.23	1	0	4	1	0	1	1	3	1	4	0	2		
				HR	−.17	1	0	4	1	0	1	1	3	1	4	0	2		
				SBP	−.22	1	0	4	1	0	1	1	3	1	4	0	2		
Burns, Evon, & Strain-Saloum	1999	61	Anger	DBP	.18	1	1	4	1	0	1	1	4	3	3	0	2	2	
				HR	.21	1	1	4	1	0	1	1	4	3	3	0	2	2	
				SBP	.30	1	1	4	1	0	1	1	4	3	3	0	2	2	
			Neutral	DBP	.27	1	1	4	1	0	1	1	4	3	3	0	2	2	
				HR	.27	1	1	4	1	0	1	1	4	3	3	0	2	2	
				SBP	.84	1	1	1	1	0	1	1	4	3	3	0	2	2	
Castaneda & Sergerstrom	2004	97	Fear	SCR	−.16	1	1	1	0	0	0	1	2	1	3	1	1	1	
				HR	−.08	1	1	1	0	0	0	1	2	1	3	1	1	1	
				HRV	.08	1	1	1	0	0	0	1	2	1	3	1	1	1	
			Anger	HR	.39	1	2	4	1	0	1	1	3	3	3	1	1	2	
				SBP	.08	1	2	4	1	0	1	1	3	3	3	1	1	2	
Chen, Gilligan, Coups, & Contrada	2005	62	Anger	HR	.28	1	2	4	1	0	1	1	3	3	3	1	1	2	
				SBP	.04	1	2	4	1	0	1	1	3	3	3	1	1	2	
Chen, Lewin, & Craske	1996	23	Fear	HR	.08	1	1	6	1	0	1	1	2	1		1	1	2	
			Neutral	HR	.08	1	1	6	1	0	1	1	2	1		1	1	2	

Table 6 (*continued*)

						Moderators												
						Sample		Study					Quality					
Citation	Year	*N*	Emotion	ANS	Effect size (Z)	Sample type	Gender	Induction type	Eval	Video recorded	Presence of others	Emotion words	BL duration	No. emotions	Induction duration	Manip. check	Rigor: Induction	Rigor: Physio	Other meta-analysis
Collett et al.	1997	30	Anger	SCL	.69	2	1	4	0	0	1	1		1		1	2	3	Cacioppo et al.
			Disgust	SCL	.02	2	1	2	0	0	0	1	2	1	2	1	2	3	
			Fear	SCL	.15	2	1	1	0	0	0	1	3	3	1	1	2	3	
			Happy	SCL	.37	1	2	2	1	0	1	0	3	4	4	1	2	2	
			Sad	SCL	.27	1	2	2	1	0	1	0	3	4	4	1	2	2	
Cornwell et al.	2006	23	Neutral	HR	.09	1	1	6	1	0	1	1	2	1		1	1	2	
Creech & Smith	2011	53	Fear	HR	.00	1	1	1	0	0	0	0	3	1	3	0	2	3	
				SCL	.19	1	1	1	0	0	0	0	3	1	3	0	2	3	
Cuthbert et al.	1996	24	Neutral	HR	−.04	1	1	1	0	0	0	0		2	1	1	2	2	
				SCL	.04	1	1	1	0	0	0	0		2	1	1	2	2	
Cuthbert et al.	2003	24	Fear	HR	.15	1	1	1	0	0	0	0		2	1	1	2	2	
				SCL	.07	1	1	1	0	0	0	0		2	1	1	2	2	
Davidson & Schwarz	1976	20	Anger	HR	.50	1	1	1	0	0	0	1		3	3	0	2	2	
Davydov, Zech, & Luminet	2011	23	Sad	HR	−.10	1	0	2	0	1	1	1	3	1	4	1	2	1	
				SCL	.08	1	0	2	0	1	1	1	3	1	4	1	2	1	
de Jong, van Overveld, & Peters	2011	60	Disgust	HR	.10	1	1	2	0	0	0	1	3	2	3	1	2	1	
				HRV	.04	1	1	2	0	0	0	1	3	2	3	1	2	1	
				SCL	.15	1	1	2	0	0	0	1	3	2	3	1	2	1	
Delgado et al.	2009		Fear	HR	.06														
				RR	.02														
Demaree & Schmeichel	2006	15	Disgust	SCR	.08	1	1	2	0	0	0	1	2	1	3	1	3	3	
				HR	−.14	1	1	2	0	0	0	1	2	1	3	1	3	3	
				HRV	.05	1	1	2	0	0	0	1	2	1	3	1	3	3	
				RR	.04	1	1	2	0	0	0	1	2	1	3	1	3	3	
Dimberg	1986	28	Fear	SCR	.32	1		3	0	0	0	1		2	1	0	2	3	
				HR	.30	1		3	0	0	0	1		2	1	0	2	3	
Ditto, Miller, & Maurice	1987	42	Anger	DBP	.16	1	0	1	1	0	1	1	2	3	3	0	2	3	
				HR	.15	1	0	1	1	0	1	1	2	3	3	0	2	3	
				SBP	.22	1	0	1	1	0	1	1	2	3	3	0	2	3	
			Fear	DBP	.18	1	0	1	1	0	1	1	2	3	3	0	2	3	
				HR	.16	1	0	1	1	0	1	1	2	3	3	0	2	3	
				SBP	.16	1	0	1	1	0	1	1	2	3	3	0	2	3	
Dorr, Brosschot, Sollers, & Thayer	2007	50	Anger	DBP	.42	1	2	4	1	1	1	0		1	4	1	1	1	
				HR	.34	1	2	4	1	1	1	0		1	4	1	1	1	
				SBP	.46	1	2	4	1	1	1	0		1	4	1	1	1	
Dowd, Zautra, & Hogan	2010	62	Anger	DBP	.27	1	2	4	1	0	1	1	3	3	3	1	1	2	
				HR	.21	1	2	4	1	0	1	1	3	3	3	1	1	2	
				SBP	.11	1	2	4	1	0	1	1	3	3	3	1	1	2	
			Neutral	DBP	−.03	1	0	4	1	0	1	1	2	1	3	1	2	2	
				HR	−.05	1	0	4	1	0	1	1	2	1	3	1	2	2	
				SBP	−.02	1	0	4	1	0	1	1	2	1	3	1	2	2	
Dunsmoor, Martin, & Labar	2012	24	Fear	SCR	.76	2		6	0	0	0	0		2	2	0	1	2	
Earle, Linden, & Wenberg	1999	60	Anger	HR	.62	2	2	5	1	1	1	0	3	2	3	1	2	2	
Ekman, Levenson, & Friesen	1983	16	Anger	HR	.25	2		3	0	1	0	1	1	6	1	1	1	3	Cacioppo et al.
			Disgust	HR	−.05	2		3	0	1	0	1	1	6	1	1	1	3	
			Fear	HR	.13	2		3	0	1	0	1	1	6	1	1	1	3	
			Happy	HR	.09	2		3	0	1	0	1	1	6	1	1	1	3	
			Sad	HR	.09	2		3	0	1	0	1	1	6	1	1	1	3	
Elesser, Heuschen, Pundt, & Sartory	2006	39	Fear	HR	−.04	2	1	3	0	0	0	1		2	1	1	2	2	
	2006	39	Neutral	HR	−.08	2	1	3	0	0	0	1		2	1	1	2	2	

(*table continues*)

Table 6 (*continued*)

						Moderators												
						Sample		Study					Quality					
Citation	Year	*N*	Emotion	ANS	Effect size (Z)	Sample type	Gender	Induction type	Eval	Video recorded	Presence of others	Emotion words	BL duration	No. emotions	Induction duration	Manip. check	Rigor: Induction	Rigor: Physio	Other meta-analysis
Elesser, Sactory, & Tackenberg	2004	29	Neutral	HR	.03	1	1	2	0	0	0	1	2	2	3	1	2	1	
Evans & Moran	1987	48	Fear	HR	.04	2		3	0	0	0	1	4	3	2	1	3	2	
Fang & Myers	2001	62	Anger	DBP	.19	1	2	4	1	0	1	1	3	3	3	1	1	2	
				HR	.29	1	2	4	1	0	1	1	3	3	3	1	1	2	
				SBP	.12	1	2	4	1	0	1	1	3	3	3	1	1	2	
			Neutral	HR	.36	1	2	4	1	0	1	1	3	3	3	1	1	2	
				SBP	.15	1	2	4	1	0	1	1	3	3	3	1	1	2	
Favaron et al.	2010	29	Fear	SCL	.01	2	1	1	0	0	0	1	3	3	1	1	2	3	
Felsten	1995	34	Anger	DBP	.39	1		4	1	0	1	1		1	4	1	1	2	
				HR	.31	1		4	1	0	1	1		1	4	1	1	2	
				SBP	.57	1		4	1	0	1	1		1	4	1	1	2	
Felsten	1996	39	Anger	DBP	.22	1	2	4	1	0	1	1	3	1	4	1	1	2	
				HR	.34	1	2	4	1	0	1	1	3	1	4	1	1	2	
				SBP	.38	1	2	4	1	0	1	1	3	1	4	1	1	2	
Fontana & McLaughlin	1998	33	Anger	DBP	.40	1	0	4	1	0	1	1	4	1	4	1	3	3	
				HR	.24	1	0	4	1	0	1	1	4	1	4	1	3	3	
				SBP	.32	1	0	1	0	0		1	3	2	1	1	2	2	
Foster & Webster	2001	10	Anger	HR	.31	1	0	1	0	0		1	3	2	1	1	2	2	
				SCR	.38	1	0	1	0	0		1	3	2	1	1	2	2	
Foster & Webster	2001	10	Happy	HR	.97	1	1	1	0	0	1	1		5	3	1	1	1	
Foster, Smith, & Webster	1999	36	Anger	HR	.23	1	0	1	0	0	1	1	3	2	1	1	2	2	
				SCL	.78	1	0	1	0	0	1	1	3	2	1	1	2	2	
Foster, Webster, & Smith	1997	16	Anger	HR	.04	1	1	1	0	0	1	1		5	3	1	1	1	
			Fear	HR	.00	1	1	1	0	0	1	1		5	3	1	1	1	
				SCL	−.04	1	1	1	0	0	1	1		5	3	1	1	1	
			Happy	HR	.03	1	1	1	0	0	1	1		5	3	1	1	1	
				SCL	−.03	1	1	1	0	0	1	1		5	3	1	1	1	
			Sad	HR	.00	1	1	1	0	0	1	1		5	3	1	1	1	
				SCL	−.03	1	1	1	0	0	1	1		5	3	1	1	1	
Frazer, Larkin, & Goodie	2002	53	Anger	DBP	−.26	1	0	4	1	0	1	1	3	1	4	0	2	1	
				HR	−.26	1	0	4	1	0	1	1	3	1	4	0	2	1	
				SBP	−.25	1	1	2	0	1	0	1	3	3	4	1	2	1	
Frazier, Strauss, & Steinhauer	2004	53	Neutral	HR	−.14	1	0	4	1	0	1	1	3	1	4	0	2	1	
				SCR	−.25	1	0	4	1	0	1	1	3	1	4	0	2	1	
Frazier, Strauss, & Steinhauer	2002	53	Neutral	HRV	.78	1	1	2	0	0	0	1	2	2	3	1	2	1	
Fredrickson & Levenson Study 1	1998	60	Fear	HP	.05	1	0	2	0	1	0	1	3	1	3	1	2	2	
				FPA	−.03	1	0	2	0	1	0	1	3	1	3	1	2	2	
				PTT	−.04	1	0	2	0	1	0	1	3	1	3	1	2	2	
Fredrickson & Levenson Study 2	1998	72	Sad	FPA	−.09	1	0	1	0	1	0	1	3	1	3	1	2	2	
				HR	−.01	1	0	1	0	1	0	1	3	1	3	1	2	2	
				PTT	.03	1	0	1	0	1	0	1	3	1	3	1	2	2	

Table 6 (*continued*)

						Moderators													
						Sample		Study					Quality						
Citation	Year	*N*	Emotion	ANS	Effect size (Z)	Sample type	Gender	Induction type	Eval	Video recorded	Presence of others	Emotion words	BL duration	No. emotions	Induction duration	Manip. check	Rigor: Induction	Rigor: Physio	Other meta-analysis
Fredrickson et al.	2000	66	Anger	DBP	.03	2	1	1	0	1	1	1	2	1	1	1	1	2	
				HR	.20	2	1	1	0	1	1	1	2	1	1	1	1	2	
				SBP	.15	2	1	1	0	1	1	1	2	1	1	1	1	2	
Funkenstein, King, & Drolette	1954	69	Anger	DBP	.29	1		4	1	0	1	1	4	1		1	1	2	Cacioppo et al.; Stemmler
				HR	.26	1		4	1	0	1	1	4	1		1	1	2	
				SBP	.53	1		4	1	0	1	1	4	1		1	1	2	
Gallo, Smith, & Kirchier	2000	29	Anger	DBP	.82	1	0	4	1	0	1	1	4	1	4	1	1	1	
				HR	.33	1	0	4	1	0	1	1	4	1	4	1	1	1	
				SBP	.80	1	0	4	1	0	1	1	4	1	4	1	1	1	
				SCR	.21	1	0	4	1	0	1	1	4	1	4	1	1	1	
Gambaro & Rabin	1969	60	Anger	DBP	.20	1	2	4	1	0	1	1	4	1	3	1	2	2	
Garcia-Leon	2003	26	Anger	DBP	.48			4											
				HR	.11			4											
				HRV	−.34			4											
				SBP	.54			4											
Gehricke & Fridlund	2002	20	Happy	HR	.10	1	2	1	0	0	1	1	2	4	2	1	2	1	
				SCL	.00	1	2	1	0	0	1	1	2	4	2	1	2	1	
			Sad	HR	.07	1	2	1	0	0	1	1	2	4	2	1	2	1	
				SCL	.00	1	2	1	0	0	1	1	2	4	2	1	2	1	
Gendolla & Silvestrini Study 1	2011	44	Anger	DBP	.04	1	1	3	1	0	0	1	3	2	3	1	2	1	
				HR	.08	1	1	3	1	0	0	1	3	2	3	1	2	1	
				SBP	.16	1	1	3	0	0	0	1	3	2	3	1	2	1	
			Happy	DBP	.04	1	1	3	1	0	0	1	3	2	3	1	2	1	
				HR	.10	1	1	3	1	0	0	1	3	2	3	1	2	1	
				SBP	.07	1	1	3	1	0	0	1	3	2	3	1	2	1	
			Anger	DBP	.12	1	1	3	1	0	0	1	3	2	3	1	2	1	
				HR	.15	1	1	3	1	0	0	1	3	2	3	1	2	1	
Gendolla & Silvestrini Study 2	2011	40	Anger	SBP	.05	1	1	3	0	0	0	1	3	2	3	1	2	1	
			Happy	DBP	.28	1	1	3	1	0	0	1	3	2	3	1	2	1	
				HR	.18	1	1	3	1	0	0	1	3	2	3	1	2	1	
				SBP	.38	1	1	3	0	0	0	1	3	2	3	1	2	1	
			Sad	DBP	.14	1	1	3	1	0	0	1	3	2	3	1	2	1	
				HR	.18	1	1	3	0	0	0	1	3	2	3	1	2	1	
				SBP	.36	1	1	3	0	0	0	1	3	2	3	1	2	1	
Gerra et al.	1996	20	Neutral	DBP	.36	1	0	2	0	0	0	1		2	4	1	2	3	
				HR	−.06	1	0	2	0	0	0	1		2	4	1	2	3	
				SBP	−.09	1	0	2	0	0	0	1		2	4	1	2	3	
Giesen & McGlynn	1977	17	Fear	HR	.46	1	0	1	0	0	1	0	3	2	1	0	3	2	
				SCL	.30	1	0	1	0	0	1	0	3	2	1	0	3	2	
Gomez & Danuser	2010	37	Neutral	CO	.35	1	1	3	0	0	0	0	4	3	2	1	2	2	
				DBP	−.06	1	1	3	0	0	0	0	4	3	2	1	2	2	
				HR	−.01	1	1	3	0	0	0	0	4	3	2	1	2	2	
				SBP	−.12	1	1	3	0	0	0	0	4	3	2	1	2	2	
				SCL	.04	1	1	3	0	0	0	0	4	3	2	1	2	2	
				SV	−.07	1	1	3	0	0	0	0	4	3	2	1	2	2	
				TPR	.04	1	1	3	0	0	0	0	4	3	2	1	2	2	
Gomez, Zimmermann, Guttormsen-Schar, & Danuser	2005	20	Happy	HR	−.06	1	1	2	0	0	0	1	2	2	4	1	1	1	
				RR	−.15	1	1	2	0	0	0	1	2	2	4	1	1	1	
				SCL	.02	1	1	2	0	0	0	1	2	2	4	1	1	1	
Goswami & Lackner	2010	15	Anger	CO	.21	2	2	4	1	0	1	1	4	1	3	1	3	3	

(*table continues*)

Table 6 (*continued*)

						Moderators												
						Sample		Study					Quality					
Citation	Year	*N*	Emotion	ANS	Effect size (Z)	Sample type	Gender	Induction type	Eval	Video recorded	Presence of others	Emotion words	BL duration	No. emotions	Induction duration	Manip. check	Rigor: Induction	Rigor: Physio	Other meta-analysis
				DBP	.12	2	2	4	1	0	1	1	4	1	3	1	3	3	
				HR	.27	2	2	4	1	0	1	1	4	1	3	1	3	3	
				MAP	.20	2	2	4	1	0	1	1	4	1	3	1	3	3	
				SBP	.19	2	2	4	1	0	1	1	4	1	3	1	3	3	
				SV	.04	2	2	4	1	0	1	1	4	1	3	1	3	3	
				TPR	−.05	2	2	4	1	0	1	1	4	1	3	1	3	3	
Green & Addis	2012	35	Sad	HR	.00	1	2	2	0	0	0	1	2	1	3	1	2	3	
Gross & Levenson	1997	185	Happy	HR	.08	1	0	2	0	0	0	1	2	3	2	1	3	3	
				SCL	.16	1	0	2	0	0	0	1	2	3	2	1	3	3	
			Sad	HR	.13	1	0	2	0	0	0	1	2	3	2	1	3	3	
			Neutral	HR	.13	1	0	2	0	0	0	3	2	3	2	1	3	3	
				SCL	.14	2	1	2	0	1	0	1		1	3	1	2	2	
Gruber	2011	23	Happy	SCR	−.10	2	1	2	0	1	0	1		1	3	1	2	2	
				HR	−.01	2	1	2	0	1	0	1		1	3	1	2	2	
				HRV	−.14	2	1	2	0	1	0	1		1	3	1	2	2	
Hamm, Cuthbert, Globisch, & Vaitl	1997	9	Happy	HR	−.31	1	0	3	0	0	0	1	1	3	2	1	2	3	
			Neutral	HR	.37	1	0	3	0	0	0	1	1	3	2	1	2	3	
Hayashi	2009	8	Fear	HR	−.15	2	0	2	0	0	1	0		1	3		2	1	
Hernandez, Larken, & Whited	2009	16	Anger	DBP	−.25	1	0	4	1	0	1	1	3	1	4	0	2		
				HR	−.23	1	0	4	1	0	1	1	3	1	4	0	2		
				SBP	−.27	1	0	4	1	0	1	1	3	1	4	0	2		
Herpertz et al.	2000	27	Neutral	HR	.46	2	1	2	0	1	0	1		1	3	1	2	2	
				SCR	−.27	2	1	2	0	1	0	1		1	3	1	2	2	
Herrald & Tomaka	2002	99	Anger	HR	.13	1	1	4	1	0	1	1	3	4	3	1	3	3	
Herring, Burleson, Roberts, & Devine	2011	39	Happy	HR	−.16	1	1	2	0	0	0	1	2	2	3	1	2	1	
				HRV	−.06	1	1	2	0	0	0	1	2	2	3	1	2	1	
				RR	.15	1	1	2	0	0	0	1	2	2	3	1	2	1	
Hess et al.	1992	30	Anger	HR	.63	1	0	1	0	1	1	1	1	3	1	1	1	2	Cacioppo et al.
				SCL	−.31	1	0	1	0	1	1	1	1	3	1	1	1	2	
			Happy	HR	.85	1	0	1	0	1	1	1	1	3	1	1	1	2	
			Sad	HR	.07	1	0	1	0	1	1	1	1	3	1	1	1	2	
				SCL	−.21	1	0	1	0	1	1	1	1	3	1	1	1	2	
Holmes & Will	2011	37	Anger	DBP	.64	2	0	4	0	0	1	0		1	3	1	1	1	
				HR	.32	2	0	4	0	0	1	0		1	3	1	1	1	
				SBP	.73	2	0	4	0	0	1	0		1	3	1	1	1	
Ishida & Okada	2006	32	Fear	HR	.00	2		2	0	0	0	1	4	1	3	1	3	2	
				HRV	−.07	2		2	0	0	0	1	4	1	3	1	3	2	
Janssen, Spinhoven, & Brosschot	2001	28	Anger	DBP	.10	1	1	4	1	0	1	1	3	1	3	1	2	2	
				HR	.18	1	1	4	1	0	1	1	3	1	3	1	2	2	
				SBP	.07	1	1	4	1	0	1	1	3	1	3	1	2	2	
				SCL	.08	1	1	4	1	0	1	1	3	1	3	1	2	2	
Johnsen, Thayer, & Hugdahl	1995	24	Anger	HR	−.01	1	1	5	1	0	1	1	4	2	4	1	1	2	
			Happy	HR	.03	1	2	2	1	0	1	0	3	4	4	1	2	2	
Johnson	1989	27	Anger	DBP	.19			4											
				HR	.36			4											
				SBP	.48			4											
Jonassaint et al.	2009	109	Anger	CO	.05	2	2	1	0	0	1	1	3	1	4	1	2	2	
				DBP	.29	2	2	1	0	0	1	1	3	1	4	1	2	2	
				HR	.10	2	2	1	0	0	1	1	3	1	4	1	2	2	

Table 6 (*continued*)

						Moderators												
						Sample		Study					Quality					
Citation	Year	*N*	Emotion	ANS	Effect size (Z)	Sample type	Gender	Induction type	Eval	Video recorded	Presence of others	Emotion words	BL duration	No. emotions	Induction duration	Manip. check	Rigor: Induction	Rigor: Physio	Other meta-analysis
				SBP	.39	2	2	1	0	0	1	1	3	1	4	1	2	2	
				TPR	.03	2	2	1	0	0	1	1	3	1	4	1	2	2	
Kaiser & Roessler	1970	20	Neutral	SCR	.03	1		2	0	0	0	1		2	4	1	2	2	
Kelly et al.	2008	23	Fear	HR	.07	1	1	6	1	0	1	1	2	1		1	1	2	
Khalfa et al.	2008	49	Happy	DBP	.20	1	1	5	0	1	0	1	1	2	2	1	3	2	
				HR	.00	1	1	5	0	1	0	1	1	2	2	1	3	2	
				RR	.06	1	1	5	0	1	0	1	1	2	2	1	3	2	
				SCR	.20	1	1	5	0	1	0	1	1	2	2	1	3	2	
			Sad	DBP	.03	1	1	5	0	1	0	1	1	2	2	1	3	2	
				HR	.00	1	1	5	0	1	0	1	1	2	2	1	3	2	
				RR	.09	1	1	5	0	1	0	1	1	2	2	1	3	2	
				SCR	−.01	1	1	5	0	1	0	1	1	2	2	1	3	2	
Kreibg, Wilhelm, Roth, & Gross	2007	33	Fear	DBP	.23	1	1	2	0	0	0	1	3	3	4	1	1	1	
				HR	.08	1	1	2	0	0	0	1	3	3	4	1	1	1	
				RR	.19	1	1	2	0	0	0	1	3	3	4	1	1	1	
				SBP	.15	1	1	2	0	0	0	1	3	3	4	1	1	1	
				SCL	.18	1	1	2	0	0	0	1	3	3	4	1	1	1	
				SCR	.37	1	1	3	0	0	0	1	3	3	4	1	1	1	
			Sad	DBP	.12	1	1	2	0	0	0	1	3	3	4	1	1	1	
				HR	−.08	1	1	2	0	0	0	1	3	3	4	1	1	1	
				RR	−.02	1	1	2	0	0	0	1	3	3	4	1	1	1	
				SBP	.05	1	1	2	0	0	0	1	3	3	4	1	1	1	
				SCL	.14	1	1	2	0	0	0	1	3	3	4	1	1	1	
				SCR	.30	1	1	2	0	0	0	1	3	3	4	1	1	1	
Kring & Neale	1996	18	Happy	SCR	.23	2	2	2	0	1	1	1	4	4	3	1	1	1	
			Neutral	SCR	.07	2	2	2	0	1	1	1	4	4	3	1	1	1	
Kunzmann & Gruhn	2005	96	Sad	HR	−.03	2	1	2	0	0	1	1		2		1	1	1	
				SCL	−.06	2	1	2	0	0	1	1		2		1	1	1	
Labouvie-Vief, Lumley, Jain, & Heinze	2003	113	Anger	HR	.19	2	1	1	1	1	1	1	4	4	3	1	1	1	
			Fear	HR	.16	2	1	1	1	1	1	1	4	4	3	1	1	1	
			Happy	HR	.10	2	1	1	1	1	1	1	4	4	3	1	1	1	
			Sad	HR	.15	2	1	1	1	1	1	1	4	4	3	1	1	1	
Lane et al.	2009	12	Disgust	HRV	−.07	2	3	1	0	1	0	1	3	3	2	1	2	2	
			Happy	HRV	−.02	2	3	1	0	1	0	1	3	3	2	1	2	2	
			Sad	HRV	−.01	2	3	1	0	1	0	1	3	3	2	1	2	2	
Larkin	2004	13	Anger	DBP	.47	2	1	4	0	1	1	1	4	2		0	2	1	
				HR	.29	2	1	4	0	1	1	1	4	2		0	2	1	
				SBP	.83	2	1	4	0	1	1	1	4	2		0	2	1	
			Neutral	DBP	.83	2	1	4	0	1	1	1	4	2		0	2	1	
				HR	.68	2	1	4	0	1	1	1	4	2		0	2	1	
				RR	.45	2	1	4	0	1	1	1	4	2		0	2	1	
				SBP	.16	2	1	4	0	1	1	1	4	2		0	2	1	
Lavoie, Miller, Conway, & Fleet	2001	21	Anger	CO	.30	2	0	4	1	0	1	1	4	1	3	1	3	3	
				DBP	.21	2	0	4	1	0	1	1	4	1	3	1	3	3	
				HR	.48	2	0	4	1	0	1	1	4	1	3	1	3	3	
				SBP	.54	2	0	4	1	0	1	1	4	1	3	1	3	3	
				SV	−.07	2	0	4	1	0	1	1	4	1	3	1	3	3	
				TPR	−.02	2	0	4	1	0	1	1	4	1	3	1	3	3	
Lawson & Bulik	1995	11	Fear	HR	−.07			6						1					
				SBP	.15			6						1					
Lee & Lang	2009	80	Anger	HR	−.07	1	1	2	0	0	0	0	1	4	2	1	1	2	
				SCL	.14	1	1	2	0	0	0	0	1	4	2	1	1	2	
			Fear	HR	−.05	1	1	2	0	0	0	0	1	4	2	1	1	2	
				SCL	.07	1	1	2	0	0	0	0	1	4	2	1	1	2	

(*table continues*)

Table 6 (*continued*)

						Moderators												
						Sample		Study					Quality					
Citation	Year	*N*	Emotion	ANS	Effect size (Z)	Sample type	Gender	Induction type	Eval	Video recorded	Presence of others	Emotion words	BL duration	No. emotions	Induction duration	Manip. check	Rigor: Induction	Rigor: Physio	Other meta-analysis
			Happy	HR	−.02	1	1	2	0	0	0	0	1	4	2	1	1	2	
				SCL	.08	1	1	2	0	0	0	0	1	4	2	1	1	2	
			Sad	HR	−.02	1	1	2	0	0	0	0	1	4	2	1	1	2	
				SCL	.02	1	1	2	0	0	0	0	1	4	2	1	1	2	
Leon, Nouwen, & Sheffield	2007	58	Anger	DBP	.63	1	0	4	1	1	1	1	4	1	3	1	2	2	
				HR	.56	1	0	4	1	1	1	1	4	1	3	1	2	2	
				SBP	.63	1	0	4	1	1	1	1	4	1	3	1	2	2	
Levenson, Ekman, & Friesen	1990	41	Anger	HR	.86	1	1	3	1	1	1	1	2	6	1	1	1	3	Cacioppo et al.; Stemmler
				SCL	.23	1	1	3	1	1	1	1	2	6	1	1	1	3	
				Temp	.47	1	1	3	1	1	1	1	2	6	1	1	1	3	
			Disgust	HR	.07	1	1	3	1	1	1	1	2	6	1	1	1	3	
				SCL	.32	1	1	3	1	1	1	1	2	6	1	1	1	3	
				Temp	.13	1	1	3	1	1	1	1	2	6	1	1	1	3	
			Fear	HR	.91	1	1	3	1	1	1	1	2	6	1	1	1	3	
				SCL	.33	1	1	3	1	1	1	1	2	6	1	1	1	3	
				Temp	−.10	1	1	3	1	1	1	1	2	6	1	1	1	3	
			Happy	HR	.43	1	1	3	1	1	1	1	2	6	1	1	1	3	
				SCL	.42	1	1	3	1	1	1	1	2	6	1	1	1	3	
				Temp	.25	1	1	3	1	1	1	1	2	6	1	1	1	3	
			Sad	HR	.70	1	1	3	1	1	1	1	2	6	1	1	1	3	
				SCL	.24	1	1	3	1	1	1	1	2	6	1	1	1	3	
				Temp	.13	1	1	3	1	1	1	1	2	6	1	1	1	3	
Levenson, Ekman, Heider, & Friesen	1992	46	Anger	HR	.28	2	2	3	1	1	1	1		6	1	1	3	3	Cacioppo et al.; Stemmler
				SCL	.10	2	2	3	1	1	1	1		6	1	1	3	3	
			Disgust	HR	.03	2	2	3	1	1	1	1		6	1	1	3	3	
				SCL	.16	2	2	3	1	1	1	1		6	1	1	3	3	
			Fear	HR	.16	2	2	3	1	1	1	1		6	1	1	3	3	
				SCL	.09	2	2	3	1	1	1	1		6	1	1	3	3	
			Happy	HR	.10	2	2	3	1	1	1	1		6	1	1	3	3	
				SCL	.04	2	2	3	1	1	1	1		6	1	1	3	3	
			Sad	HR	.27	2	2	3	1	1	1	1		6	1	1	3	3	
				SCL	.08	2	2	3	1	1	1	1		6	1	1	3	3	
Lin, Lin, Lin, & Huang	2011	37	Neutral	HR	−.06	1	1	4	1	0	1	0	4	1	3	0	2	2	
				RR	−.06	1	1	4	1	0	1	0	4	1	3	0	2	2	
				SCR	.08	1	1	4	1	0	1	0	4	1	3	0	2	2	
Lobbestael, Arntz, & Wiers	2008	16	Anger	DBP	.36	1	1	5	1	0	1	1	4	2	4	1	1	2	
				HR	.17	1	1	5	1	0	1	1	4	2	4	1	1	2	
				SBP	.26	1	1	5	1	0	1	1	4	2	4	1	1	2	
				SCL	.00	1	1	5	1	0	1	1	4	2	4	1	1	2	
				SCR	.00	1	1	5	1	0	1	1	4	2	4	1	1	2	
Lobbestael, Arntz, Cima, & Chakhssi	2009	35	Anger	DBP	.41	2	1	1	0	0	1	1	4	1	4	1	2	2	
				HR	.46	2	1	1	0	0	1	1	4	1	4	1	2	2	
				SBP	.35	2	1	1	0	0	1	1	4	1	4	1	2	2	
				SCL	.21	2	1	1	0	0	1	1	4	1	4	1	2	2	
				SCR	.20	2	1	1	0	0	1	1	4	1	4	1	2	2	
Marci, Glick, Loh, & Dougherty	2007	9	Anger	HR	.26	2	1	1	0	0	1	1	2	3	2	1	3	3	
				HRV	.22	2	1	1	0	0	1	1	2	3	2	1	3	3	
				SCL	.21	2	1	1	0	0	1	1	2	3	2	1	3	3	
			Happy	HR	.23	2	1	1	0	0	1	1	2	3	2	1	3	3	

Table 6 (*continued*)

						Moderators												
						Sample		Study					Quality					
Citation	Year	*N*	Emotion	ANS	Effect size (Z)	Sample type	Gender	Induction type	Eval	Video recorded	Presence of others	Emotion words	BL duration	No. emotions	Induction duration	Manip. check	Rigor: Induction	Rigor: Physio	Other meta-analysis
				HRV	.15	2	1	1	0	0	1	1	2	3	2	1	3	3	
				SCL	.09	2	1	1	0	0	1	1	2	3	2	1	3	3	
			Sad	HR	.17	2	1	1	0	0	1	1	2	3	2	1	3	3	
				HRV	.19	2	1	1	0	0	1	1	2	3	2	1	3	3	
				SCL	.03	2	1	1	0	0	1	1	2	3	2	1	3	3	
Mataix-Cols et al.	2008	14	Disgust	SCR	.10	2	0	4	1	0	1	1	4	1	3	1	3	3	
Mauss & Butler	2010	28	Anger	HR	.33	2	1	3	0	0	1	0		3	4	1	2	2	
Mauss, Cook, & Gross	2007	67	Anger	HR	.02	1	0	4	1	0	1	1	3	1	3	1	1	1	
				CO	.25	1	0	4	1	0	1	1	3	1	3	1	1	1	
				HR	.20	1	0	4	1	0	1	1	3	1	3	1	1	1	
				MAP	.30	1	0	4	1	0	1	1	3	1	3	1	1	1	
				TPR	−.05	1	0	4	1	0	1	1	3	1	3	1	1	1	
McClelland, Jones, & Gregg	2009	99	Anger	DBP	.45	1	1	4	0	0	1	1	3	2	2	1	3	2	
				HR	.08	1	1	4	0	0	1	1	3	2	2	1	3	2	
				SBP	.49	2	1	1	0	0	0	1	3	3	1	1	2	3	
McTeague et al.	2009	76	Fear	HR	−.04	2	1	1	0	0	0	1	3	3	1	1	2	3	
				SCL	.05	2	1	1	0	0	0	1	3	3	1	1	2	3	
			Neutral	HR	.05	2	1	1	0	0	0	1	3	3	1	1	2	3	
				SCL	−.05	2	1	1	0	0	0	1	3	3	1	1	2	3	
McTeague et al.	2010	75	Anger	HR	−.04	2	1	1	0	0	0	1	3	3	1	1	2	3	
				SCL	.00	2	1	1	0	0	0	1	3	3	1	1	2	3	
			Fear	HR	.09	2	1	1	0	0	0	1	3	3	1	1	2	3	
				SCL	.07	2	1	1	0	0	0	1	3	3	1	1	2	3	
			Neutral	HR	−.05	2	1	1	0	0	0	1	3	3	1	1	2	3	
				SCL	.00	2	1	1	0	0	0	1	3	3	1	1	2	3	
Meissner, Muth, & Herbert	2011	31	Disgust	HR	.09	2	1	3	0	0	0	1	3	1	3	1	3	3	
				HRV	.29	2	1	3	0	0	0	1	3	1	3	1	3	3	
				SCR	.03	2	1	3	0	0	0	1	3	1	3	1	3	3	
Merckelbach, Vanhout, Vandenhout, & Mersch	1989	18	Anger	SCR	.22	2		3											
Moon & Eisler	1983	10	Anger	DBP	.00	1		4	0	0	0	1	3	3	1	1	2	3	
				SBP	.81	1		4	0	0	0	1	3	3	1	1	2	3	
Moore	2009	21	Anger	HRV	−.13	2	0	4	1	1	1	0	3	1	2	0	3	2	
		13	Neutral	HRV	−.13	2	0	4	1	1	1	0	3	1	2	0	3	2	
Muth et al.	1999	20	Fear	HR	−.24			4						1					
Najstrom	2006	79	Anger	SCR	−.08	1	1	4	1	0	1	1	4	1	4	1	2	1	
Neumann & Waldstein	2001	36	Anger	HR	.46	2	1	4	0	0	1	1		1		1	2	3	
				SBP	.16	1	1	3	0	0	0	1	3	2		0	2	2	
Neumann et al.	2011	99	Anger	DBP	.41	1	1	4	1	0	1	1	4	1	4	1	2	1	
				HR	.39	1	1	4	1	0	1	1	4	1	4	1	2	1	
				SBP	.36	2	1	4	0	0	1	1		1		1	2	3	
Neumann et al.	2004	36	Anger	HR	.71			4											
Nyklicek et al.	1997	26	Anger	DBP	.04	1	1	5	0	0	0	1	3	5	4	1	2	2	
				HR	.15	1	1	5	0	0	0	1	3	5	4	1	2	2	
				HRV	−.18	1	1	5	0	0	0	1	3	5	4	1	2	2	
				RR	.39	1	1	5	0	0	0	1	3	5	4	1	2	2	
				SBP	.27	1	1	5	0	0	0	1	3	5	4	1	2	2	
			Happy	DBP	.00	1	1	5	0	0	0	1	3	5	4	1	2	2	

(*table continues*)

Table 6 (*continued*)

						Moderators												
						Sample		Study					Quality					
Citation	Year	*N*	Emotion	ANS	Effect size (Z)	Sample type	Gender	Induction type	Eval	Video recorded	Presence of others	Emotion words	BL duration	No. emotions	Induction duration	Manip. check	Rigor: Induction	Rigor: Physio	Other meta-analysis
				HR	.01	1	1	5	0	0	0	1	3	5	4	1	2	2	
				HRV	−.22	1	1	5	0	0	0	1	3	5	4	1	2	2	
				RR	.67	1	1	5	0	0	0	1	3	5	4	1	2	2	
				SBP	.29	1	1	5	0	0	0	1	3	5	4	1	2	2	
			Sad	DBP	.04	1	1	5	0	0	0	1	3	5	4	1	2	2	
				HR	.20	1	1	5	0	0	0	1	3	5	4	1	2	2	
				HRV	−.11	1	1	5	0	0	0	1	3	5	4	1	2	2	
				RR	.31	1	1	5	0	0	0	1	3	5	4	1	2	2	
				SBP	.19	1	1	5	0	0	0	1	3	5	4	1	2	2	
			Neutral	DBP	.19	1	1	5	0	0	0	1	3	5	4	1	2	2	
				HR	−.08	1	1	5	0	0	0	1	3	5	4	1	2	2	
				HRV	.25	1	1	5	0	0	0	1	3	5	4	1	2	2	
				RR	−.03	1	1	5	0	0	0	1	3	5	4	1	2	2	
				SBP	.04	1	1	5	0	0	0	1	3	5	4	1	2	2	
Oxley et al.	2008	46	Fear	SCL	−.04	1	1	1	0	0	1	1		5	3	1	1	1	
Palm & Ohman	1992	24	Anger	DBP	−.27	1		4	1	0	1	1	3	1	3	1	1	2	
				HR	−.20	1		4	1	0	1	1	3	1	3	1	1	2	
				SBP	−.24	1		4	1	0	1	1	3	1	3	1	1	2	
Panayiotou et al.	2011	53	Fear	HR	.08	2	1	2	0	1	0	1		1	3	1	2	2	
Papousek, Freudenthaler, & Schulter	2008	60	Happy	HR	.05	1	2	2	0	0	0	1	2	2	2	2	2	2	
				HRV	−.08	1	2	2	0	0	0	1	2	2	2	2	2	2	
			Sad	HR	.08	1	2	2	0	0	0	1	2	2	2	2	2	2	
				HRV	−.11	1	2	2	0	0	0	1	2	2	2	2	2	2	
Pauls & Stemmler	2003	40	Anger	DBP	.39	2	0	4	1	1	1	1	3	1	4	1	1	1	
				HR	.77	2	0	4	1	1	1	1	3	1	4	1	1	1	
				HRV	.48	2	0	4	1	1	1	1	3	1	4	1	1	1	
				SBP	.96	2	0	4	1	1	1	1	3	1	4	1	1	1	
				SCR	.36	2	0	4	1	1	1	1	3	1	4	1	1	1	
			Fear	DBP	.26	2	0	4	1	1	1	1	3	1	4	1	1	1	
				HR	.97	2	0	4	1	1	1	1	3	1	4	1	1	1	
				HRV	.83	2	0	4	1	1	1	1	3	1	4	1	1	1	
				SBP	.81	2	0	4	1	1	1	1	3	1	4	1	1	1	
				SCR	.73	2	0	4	1	1	1	1	3	1	4	1	1	1	
Pointer et al.	2012	50	Anger	DBP	.36	2		1	0	0	1	1	3	2	4	1	1	3	
				HR	.38	2		1	0	0	1	1	3	2	4	1	1	3	
				SBP	.34	2		1	0	0	1	1	3	2	4	1	1	3	
Powch & Houston	1996	109	Neutral	DBP	.34	1	0	4	1	1	1	1	3	1	4	1	2	2	
				HR	.38	1	0	4	1	1	1	1	3	1	4	1	2	2	
				SBP	.01	1	0	4	1	1	1	1	3	1	4	1	2	2	
Prkachin, Mills, Zwall, & Husted	2001a	50	Anger	DBP	−.24	1	1	1	0	0	1	1	4	3	3	1	2	2	Stemmler
				HR	−.16	1	1	1	0	0	1	1	4	3	3	1	2	2	
				SBP	−.20	1	1	1	0	0	1	1	4	3	3	1	2	2	
			Neutral	DBP	−.20	1	1	1	0	0	1	1	4	3	3	1	2	2	
				HR	−.15	1	1	1	0	0	1	1	4	3	3	1	2	2	
				SBP	−.13	1	1	1	0	0	1	1	4	3	3	1	2	2	
Prkachin, Mills, Zwall, & Husted	2001b	98	Disgust	CO	−.07	2	1	2	0	0	0	1	2	1	2	1	2	3	
				DBP	.07	2	1	2	0	0	0	1	2	1	2	1	2	3	
				SBP	.03	2	1	2	0	0	0	1	2	1	2	1	2	3	
				SV	−.07	2	1	2	0	0	0	1	2	1	2	1	2	3	
				TPR	.02	2	1	2	0	0	0	1	2	1	2	1	2	3	
Quartana & Burns	2010	34	Anger	DBP	.61	2	1	4	0	0	0	1	2	1	2	1	2	3	
				HR	.23	2	1	4	0	0	0	1	2	1	2	1	2	3	

Table 6 (*continued*)

						Moderators													
						Sample		Study					Quality						
Citation	Year	*N*	Emotion	ANS	Effect size (Z)	Sample type	Gender	Induction type	Eval	Video recorded	Presence of others	Emotion words	BL duration	No. emotions	Induction duration	Manip. check	Rigor: Induction	Rigor: Physio	Other meta-analysis
				SBP	.37	2	1	4	0	0	0	1	2	1	2	1	2	3	
			Fear	DBP	.38	2	1	4	0	0	0	1	2	1	2	1	2	3	
				HR	.25	2	1	4	0	0	0	1	2	1	2	1	2	3	
				SBP	.42	2	1	4	0	0	0	1	2	1	2	1	2	3	
Radstaak et al.	2011	103	Anger	DBP	.53	2	1	4	0	0	0	1	2	1	2	1	2	3	
				HR	.31	2	1	4	0	0	0	1	2	1	2	1	2	3	
				SBP	.58	2	1	4	0	0	0	1	2	1	2	1	2	3	
Ribeiro, Teixeira-Silva, Pompeia, & Bueno	2007	24	Happy	SCL	.11	1	1	3	0	0	0	0		4	1	1	1	1	
			Neutral	SCL	.09	1	1	3	0	0	0	0		4	1	1	1	1	
Richman et al.	2007	164	Anger	DBP	.29	2	1	1	0	0	1	1	3	1	3	1	3	2	
				HR	.25	2	1	1	0	0	1	1	3	1	3	1	3	2	
				SBP	.37	2	1	1	0	0	1	1	3	1	3	1	3	2	
Ritz, Dahme, & Claussen	1999	20	Disgust	HP	−.05	1	1	3	0	0	0	1	2	3	3	1	2	2	
				HRV	.00	1	1	3	0	0	0	1	2	3	3	1	2	2	
				RR	.03	1	1	3	0	0	0	1	2	3	3	1	2	2	
				SCL	−.01	1	1	3	0	0	0	1	2	3	3	1	2	2	
			Happy	HR	−.07	1	1	3	0	0	0	1	2	3	3	1	2	2	
				HRV	.01	1	1	3	0	0	0	1	2	3	3	1	2	2	
				RR	.09	1	1	3	0	0	0	1	2	3	3	1	2	2	
				SCL	.00	1	1	3	0	0	0	1	2	3	3	1	2	2	
Roberts & Weerts	1982	16	Anger	DBP	.25	1	1	1	0	0	0	1	3	3	2	1	2	2	Cacioppo et al.; Stemmler
				HR	.24	1	1	1	0	0	0	1	3	3	2	1	2	2	
				SBP	.30	1	1	1	0	0	0	1	3	3	2	1	2	2	
			Fear	DBP	.06	1	1	1	0	0	0	1	3	3	2	1	2	2	
				SBP	.27	1	1	1	0	0	0	1	3	3	2	1	2	2	
			Neutral	DBP	.27	1	1	1	0	0	0	1	3	3	2	1	2	2	
				HR	.04	1	1	1	0	0	0	1	3	3	2	1	2	2	
				SBP	−.01	1	1	1	0	0	0	1	3	3	2	1	2	2	
Rochman & Diamond	2008	27	Anger	HRV	.07	2	1	2	0	1	0	1	2	1	3	1	1	1	
				SCL	.19	2	1	2	0	1	0	1	2	1	3	1	1	1	
				Temp	−.09	2	1	2	0	1	0	1	2	1	3	1	1	1	
			Sad	HRV	.06	2	1	2	0	1	0	1	2	1	3	1	1	1	
				SCL	.14	2	1	2	0	1	0	1	2	1	3	1	1	1	
Rohrmann, Hopp, & Quirin	2008	100	Disgust	CO	−.04	2	1	2	0	0	0	1	2	1	2	1	2	3	
				DBP	.02	2	1	2	0	0	0	1	2	1	2	1	2	3	
				HR	.01	2	1	2	0	0	0	1	2	1	2	1	2	3	
				SBP	.07	2	1	2	0	0	0	1	2	1	2	1	2	3	
				SCL	.03	2	1	2	0	0	0	1	2	1	2	1	2	3	
				SV	−.01	2	1	2	0	0	0	1	2	1	2	1	2	3	
				TPR	.09	2	1	2	0	0	0	1	2	1	2	1	2	3	
Rottenberg et al.	2002	8	Sad	HR	.06	2	1	2	0	1	0	1	2	1	3	1	1	1	
				HRV	−.03	2	1	2	0	1	0	1	2	1	3	1	1	1	
				RR	.22	2	1	2	0	1	0	1	2	1	3	1	1	1	
				TV	−.04	2	1	2	0	1	0	1	2	1	3	1	1	1	

(*table continues*)

Table 6 (*continued*)

						Moderators												
						Sample		Study					Quality					
Citation	Year	*N*	Emotion	ANS	Effect size (Z)	Sample type	Gender	Induction type	Eval	Video recorded	Presence of others	Emotion words	BL duration	No. emotions	Induction duration	Manip. check	Rigor: Induction	Rigor: Physio	Other meta-analysis
Sargent, Flora, & Williams	1999	17	Anger	DBP	.08	1	1	4	1	0	1	1	2	1	3	1	2	2	
				HR	.07	1	1	4	1	0	1	1	2	1	3	1	2	2	
				SBP	.22	1	1	3	0	0	0	0		3	1	0	2	1	
Sarlo et al.	2005	46	Neutral	HR	1.08	1	1	3	0	0	0	0		3	1	0	2	1	
				SCR	.18	1	1	3	0	0	0	0		3	1	0	2	1	
Sarlo et al.	2008	17	Neutral	HR	−.08	1	1	4	1	0	1	1	4	1	4	1	3	3	
Schachter	1957	48	Anger	CO	1.04	2	1	4	0	0	1	0	4	3		1	2	2	Cacioppo et al.; Stemmler
				DBP	.35	2	1	4	0	0	1	0	4	3		1	2	2	
				HR	1.24	2	1	4	0	0	1	0	4	3		1	2	2	
				RR	1.50	2	1	4	0	0	1	0	4	3		1	2	2	
				SBP	.47	2	1	4	0	0	1	0	4	3		1	2	2	
				SCL	−.34	2	1	4	0	0	1	0	4	3		1	2	2	
				SV	1.02	2	1	4	0	0	1	0	4	3		1	2	2	
				TPR	.04	2	1	4	0	0	1	0	4	3		1	2	2	
			Fear	CO	1.76	2	1	4	0	0	1	0	4	3		1	2	2	
				DBP	.33	2	1	4	0	0	1	0	4	3		1	2	2	
				HR	1.74	2	1	4	0	0	1	0	4	3		1	2	2	
				RR	1.80	2	1	4	0	0	1	0	4	3		1	2	2	
				SBP	.50	2	1	4	0	0	1	0	4	3		1	2	2	
				SCL	−.33	2	1	4	0	0	1	0	4	3		1	2	2	
				SV	1.38	2	1	4	0	0	1	0	4	3		1	2	2	
				TPR	−.32	2	1	4	0	0	1	0	4	3		1	2	2	
Schneider, Gur, Jaggi, & Gur	1994	12	Happy	HR	.66	2	1	2	0	1	0	1	2	1	3	1	1	1	
			Sad	HR	.55	2	1	2	0	1	0	1	2	1	3	1	1	1	
Schwartz, Weinberger, & Singer	1980	32	Anger	DBP	.94	2	1	1	0	1	1	1		6		1	2	2	Cacioppo et al.; Stemmler
				HR	1.64	2	1	1	0	1	1	1		6		1	2	2	
				SBP	1.17	2	1	1	0	1	1	1		6		1	2	2	
			Fear	DBP	.48	2	1	1	0	1	1	1		6		1	2	2	
				HR	1.62	2	1	1	0	1	1	1		6		1	2	2	
				SBP	1.10	2	1	1	0	1	1	1		6		1	2	2	
			Happy	DBP	.50	2	1	1	0	1	1	1		6		1	2	2	
				HR	1.08	2	1	1	0	1	1	1		6		1	2	2	
				SBP	.86	2	1	1	0	1	1	1		6		1	2	2	
			Sad	DBP	.22	2	1	1	0	1	1	1		6		1	2	2	
				HR	1.29	2	1	1	0	1	1	1		6		1	2	2	
				SBP	1.06	2	1	1	0	1	1	1		6		1	2	2	
			Neutral	DBP	1.10	2	1	1	0	1	1	1		6		1	2	2	
				HR	.13	2	1	1	0	1	1	1		6		1	2	2	
				SBP	.02	2	1	1	0	1	1	1		6		1	2	2	
Shapiro, Goldstein, & Jamner	1995	165	Anger	DBP	.29	2	1	1	0	0	1	1	3	1	3	1	3	2	
				HR	.25	2	1	1	0	0	1	1	3	1	3	1	3	2	
				SBP	.37	2	1	1	0	0	1	1	3	1	3	1	3	2	
Shiota et al.	2011	37	Happy	HR	−.05	1	1	3	0	1	1	1	3	4	2	1	2	3	
			Neutral	HR	.50	1	1	3	0	1	1	1	3	4	2	1	2	3	

Table 6 (*continued*)

						Moderators												
						Sample		Study					Quality					
Citation	Year	*N*	Emotion	ANS	Effect size (Z)	Sample type	Gender	Induction type	Eval	Video recorded	Presence of others	Emotion words	BL duration	No. emotions	Induction duration	Manip. check	Rigor: Induction	Rigor: Physio	Other meta-analysis
Silvestrini & Gendolla	2007	13	Neutral	SCL	−.05	1	1	3	1	0	0	1	3	2	3	1	2	1	
		15	Sad	DBP	.08	1	1	3	1	0	0	1	3	2	3	1	2	1	
				HR	.16	1	1	3	0	0	0	1	3	2	3	1	2	1	
				SBP	.16	1	1	3	1	0	0	1	3	2	3	1	2	1	
Silvestrini & Gendolla	2011	36	Happy	DBP	.79	1	1	2	0	0	0	1	4	3	4	1	2	2	
				HR	.50	1	1	2	0	0	0	1	4	3	4	1	2	2	
				SBP	.00	1	1	2	0	0	0	1	4	3	4	1	2	2	
			Sad	DBP	.80	1	1	2	0	0	0	1	4	3	4	1	2	2	
				HR	.69	1	1	2	0	0	0	1	4	3	4	1	2	2	
				SBP	.00	1	1	2	0	0	0	1	4	3	4	1	2	2	
		13	Neutral	DBP	.80	1	1	2	0	0	0	1	4	3	4	1	2	2	
				HR	−.02	1	1	2	0	0	0	1	4	3	4	1	2	2	
				SBP	−.12	1	1	2	0	0	0	1	4	3	4	1	2	2	
Simic	2012	15	Fear	RR	.25	1	1	2	1	0	0	1		3	4	1	1	2	
Sinha & Parsons	1996	27	Anger	DBP	.46	2	2	1	1	0		1	3	5	2	1	1	1	Stemmler
				HR	.50	2	2	1	1	0		1	3	5	2	1	1	1	
				SBP	.49	2	2	1	1	0		1	3	5	2	1	1	1	
				SCL	.05	2	2	1	1	0		1	3	5	2	1	1	1	
			Fear	DBP	.04	2	2	1	1	0	1	1	3	5	2	1	1	1	
				HR	.54	2	2	1	1	0	1	1	3	5	2	1	1	1	
				SBP	.62	2	2	1	1	0	1	1	3	5	2	1	1	1	
				SCL	.05	2	2	1	1	0	1	1	3	5	2	1	1	1	
			Neutral	DBP	.05	2	2	1	1	0	1	1	3	5	2	1	1	1	
				HR	.08	2	2	1	1	0	1	1	3	5	2	1	1	1	
				SBP	.08	2	2	1	1	0	1	1	3	5	2	1	1	1	
				SCL	.07	2	2	1	1	0	1	1	3	5	2	1	1	1	
Sinha, Lovalla, & Parsons	1992	27	Anger	HR	.34														Cacioppo et al.; Stemmler
			Fear	HR	.02	1	1	3	0	0	0	0	4	3	1	0	2	1	
			Happy	HR	.09	1	2	2	1	0	1	0	3	4	4	1	2	2	
			Neutral	HR	.06	1	1	3	0	0	0	1	4	3	4	1	2	1	
			Sad	HR		1	2	2	1	0	1	0	3	4	4	1	2	2	
Smith, McHugo & Lanzetta	1986	8	Anger	HR	.09	1	2	2	1	0	1	0	3	4	4	1	2	2	
				SCL	.14	1	2	2	1	0	1	0	3	4	4	1	2	2	
			Happy	HR	.00	1	2	2	1	0	1	0	3	4	4	1	2	2	
				SCL	.19	1	2	2	1	0	1	0	3	4	4	1	2	2	
			Neutral	HR	.03	1	2	2	1	0	1	0	3	4	4	1	2	2	
				SCL	.09	1	1	2	0	0	0	1	4	3	4	1	2	2	
		7	Sad	HR	.14	1	1	2	0	0	0	1	4	3	4	1	2	2	
				SCL	.42	1	0	5	0	0	0	1	4	3	3	1	2	1	
Sokhadze	2007	29	Disgust	HR	−.19														
				HRV	.22	1	2	2	1	0	1	0	3	4	4	1	2	2	
				SCR	.69	1	0	5	0	0	0	1	4	3	3	1	2	1	
Spiegel & Zelin	1973	42	Anger	SBP	.17	1	2	4	1	0	1	1		3	4	1	2	3	
Stark et al.	1999	48	Fear	DBP	−.05	1	0	5	0	0	0	1	4	3	4	1	2	1	
				HP	.07	1	0	5	0	0	0	1	4	3	4	1	2	1	
				HRV	.00	1	0	5	0	0	0	1	4	3	4	1	2	1	
				RR	.00	1	0	5	0	0	0	1	4	3	4	1	2	1	
				SBP	−.05	1	1	3	1	1	1	0	2	6	1	0	2	2	

(*table continues*)

Table 6 (*continued*)

						Moderators												
						Sample		Study					Quality					
Citation	Year	*N*	Emotion	ANS	Effect size (Z)	Sample type	Gender	Induction type	Eval	Video recorded	Presence of others	Emotion words	BL duration	No. emotions	Induction duration	Manip. check	Rigor: Induction	Rigor: Physio	Other meta-analysis
Sugawara, Tarumi, & Tanaka	2010	17	Happy	DBP	.03	2	1	2	0	0	0	0	4	1	4	1	1	2	
				HR	−.03	2	1	2	0	0	0	0	4	1	4	1	1	2	
				SBP	.06	2	1	2	0	0	0	0	4	1	4	1	1	2	
			Neutral	DBP	.09	2	1	2	0	0	0	0	4	1	4	1	1	2	
				HR	.06	2	1	2	0	0	0	0	4	1	4	1	1	2	
				SBP	−.03	1	1	5	1	0	1	1	4	2	4	1	1	2	
Sullivan	1981a	13	Anger	DBP	1.12	2		3	0	1	0	1	1	6	1	1	1	3	
				HR	.68	1	1	5	1	0	1	1	4	2	4	1	1	2	
				SBP	1.14	2	1	2	0	0	0	0	4	1	4	1	1	2	
Sullivan	1981b	13	Anger	DBP	.72			4											
				HR	1.36	1	0	4	1	0	1	1	4	1	4	1	3	3	
				SBP	.85	1	0	4	1	0	1	1	4	1	4	1	3	3	
Thomas	1982	20	Anger	HR	.35	1	1	1	0	0	0	1		3	3	0	2	2	
		23	Neutral	HR	.35	1	1	6	1	0	1	1	2	1		1	1	2	
Tsai, Chentsova-Dutton, Freire-Bebeau, & Przymus	2002	98	Anger	SCL	.10	1	1	1	0	1	1	1	2	5	3	1	1	2	
			Disgust	SCL	.09	1	1	1	0	1	1	1	2	5	3	1	1	2	
			Happy	SCL	.09	1	1	1	0	1	1	1	2	5	3	1	1	2	
			Sad	SCL	.07	1	1	1	0	1	1	1	2	5	3	1	1	2	
Uchiyama	1992	6	Anger	DBP	.02	1	1	4	0	0		1	3	3	2	1	1	2	Stemmler
				HR	.24	1	1	4	0	0		1	3	3	2	1	1	2	
				RR	−.05	1	1	4	0	0		1	3	3	2	1	1	2	
				SBP	.02	1	1	4	0	0		1	3	3	2	1	1	2	
				SCR	.58	1	1	4	0	0		1	3	3	2	1	1	2	
			Fear	DBP	.03	1	1	4	0	0		1	3	3	2	1	1	2	
				HR	.16	1	1	4	0	0		1	3	3	2	1	1	2	
				RR	−.02	1	1	4	0	0		1	3	3	2	1	1	2	
				SBP	.17	1	1	4	0	0		1	3	3	2	1	1	2	
				SCR	.76	1	1	4	0	0		1	3	3	2	1	1	2	
			Happy	DBP	.00	1	1	4	0	0		1	3	3	2	1	1	2	
				HR	.07	1	1	4	0	0		1	3	3	2	1	1	2	
				RR	.03	1	1	4	0	0		1	3	3	2	1	1	2	
				SBP	.10	1	1	4	0	0		1	3	3	2	1	1	2	
				SCR	.53	1	1	4	0	0		1	3	3	2	1	1	2	
Van den Hout, De Jong, & Kindt	2000	38	Anger	SCR	.31	2	1	6	0	0	0	1		3	1	1	2	2	
			Fear	SCR	.17	2	1	6	0	0	0	1		3	1	1	2	2	
van Overveld et al.	2009	69	Disgust	HRV	−.05			1	0	0	0	1		3	1	1	2	2	
				SCL	.02	1		1	0	0	0	1		3	1	1	2	2	
Velasco & Bond	1998	39	Anger	HR	.10	1		1	0	0	0	1		3	1	1	2	2	
				SCL	.16			1	0	0	0	1		3	1	1	2	2	
				SCR	.11	1		1	0	0	0	1		3	1	1	2	2	
		39	Fear	HR	.06	1		1	0	0	0	1		3	1	1	2	2	
				SCL	.19	1		1	0	0	0	1		3	1	1	2	2	
				SCR	.19	1		1	0	0	0	1		3	1	1	2	2	
		39	Neutral	HR	−.02	1		1	0	0	0	1		3	1	1	2	2	
				SCL	.11	1		1	0	0	0	1		3	1	1	2	2	
				SCR	.29	1	1	5	1	0	1	1	4	2	4	1	1	2	
Vrana	2009	50	Anger	DBP	.06	1	1	1	0	0	0	1		4	1	1	1	2	
				HR	.08	1		1	0	0	0	1		4	1	1	1	2	
				SBP	.09														
Vrana	1994	48	Disgust	HR	−.11	1	1	5	1	0	1	1	4	2	4	1	1	2	

Table 6 (*continued*)

						Moderators												
						Sample		Study					Quality					
Citation	Year	*N*	Emotion	ANS	Effect size (Z)	Sample type	Gender	Induction type	Eval	Video recorded	Presence of others	Emotion words	BL duration	No. emotions	Induction duration	Manip. check	Rigor: Induction	Rigor: Physio	Other meta-analysis
Vrana et al.	1994	50	Anger	HR	.13	1	1	1	0	0		1		3		1	2	2	
				SCL	.02	1	1	1	0	0		1		3		1	2	2	
		50	Disgust	HR	.15	1	1	1	0	0	1	1		3		1	2	2	
				SCL	−.02	1	1	1	0	0	1	1		3		1	2	2	
Why & Johnston Study 1	2008	58	Anger	CO	.00			1											
				DBP	.30	1	2	1	0	0	1	1	3	1	3	1	3	2	
				HR	.19	1	2	1	0	0	1	1	3	1	3	1	3	2	
				SBP	.25	1	2	1	0	0	1	1	3	1	3	1	3	2	
				TPR	.13			1											
Why & Johnston Study 2	2008	59	Anger	CO	.00	1	2	1	1	0	1	1	3	2	3	1	1	2	
				DBP	.26	1	2	1	1	0	1	1	3	2	3	1	1	2	
				HR	.14	1	2	1	1	0	1	1	3	2	3	1	1	2	
				SBP	.22	1	2	1	1	0	1	1	3	2	3	1	1	2	
				TPR	.11			1											
			Neutral	DBP	.11	1	2	1	1	0	1	1	3	2	3	1	1	2	
				HR	.21	1	2	1	1	0	1	1	3	2	3	1	1	2	
				SBP	.13	1	2	1	1	0	1	1	3	2	3	1	1	2	
Yogo et al.	1995	24	Anger	DBP	.14	1	2	1	1	0	1	1	3	2	3	1	1	2	
				SBP	.17	1	2	1	1	0	1	1	3	2	3	1	1	2	
			Happy	DBP	.44	1	2	1	1	0	1	1	3	2	3	1	1	2	
				SBP	.43	1	2	1	1	0	1	1	3	2	3	1	1	2	
			Neutral	DBP	.01	1	2	1	1	0	1	1	3	2	3	1	1	2	
				SBP	.01	1	2	1	1	0	1	1	3	2	3	1	1	2	

independently and normally distributed with zero mean. Note that the sampling variance for each outcome was known *before* the meta-analyses were conducted: the sampling variance of Fisher's z is equal to $\frac{1}{N-3}$, with N being the sample size of the study. Therefore, only β_0, the overall mean effect size, and the between and within-study variance components, σ^2_V and σ^2_U, were estimated in the meta-analysis. A between-study variance (σ^2_V) estimate larger than zero means that the differences between studies in the observed effect sizes is larger than what would be expected by chance (i.e., more than just random fluctuation between studies that is due to sampling variance). Therefore, a significant between-study variance reflects the fact that the observed effect sizes in some studies are systematically larger than in other studies. A within-study variance (σ^2_U) estimate larger than zero means that the observed effects for an ANS measure within a single study varies across participants.

We extended this REM by including a single, categorical predictor variable to denote ANS measurement type (e.g., HR, DBP, etc.). The categorical variable, when used in a multilevel model, allowed us to model differences between ANS measurement types within studies, without assuming that the measurement type explains all variance between outcomes within studies. In our analyses, we tested the difference between ANS type, and estimated the overall effect for each ANS measurement type (i.e., one effect size for HR, for DBP, etc.).

Heterogeneity. We investigated whether the variance between the observed effect sizes was larger than what would be expected on the basis of sampling variance alone (Hedges, 1982; Rosenthal & Rubin, 1982). If the effect sizes are heterogeneous it means: (a) the mean effect size *does not* represent individual effect sizes for studies within the population such that (b) moderators of the effect sizes may be present (e.g., induction type) and/or (c) the effect sizes reflect real, contextual or situated ANS changes that are consistent with the hypothesis that an emotion category is a population of variable, situated events. Studies do not usually measure or report fine-grained contextual details, so we were able to meta-analytically assess Option 2 but not Option 3. A multilevel meta-analysis model can include multiple moderator variables, without assuming that all variability between studies and between conditions can be explained by the included moderators (Van Den Noortgate & Onghena, 2003).

We computed several indices of heterogeneity. We first computed Cochran's Q statistic, which is distributed as a chi-square with $k - 1$ degrees of freedom, where k = number of effect sizes (Cochran, 1954; Hedges & Olkin, 1985) and calculated as $Q = \sum_{i=1}^{k} w_i (ES_i - \overline{ES})^2$, where ES_i is the individual effect size for $i = 1$ to k, $\overline{ES}$ is the weighted mean effect size, and w_i is the individual weight for ES_i (in our case the inverse of the variance). A statistically significant Q indicates heterogeneity in effect sizes. Note, however, that a nonsignificant Q does not always provide great confidence that the population distribution is homogeneous (Gavaghan, Moore, & McQuay, 2000). In an analysis with a small number of effect sizes, especially if they are based on small sample size studies, the Q-statistic may be nonsignificant even when there is considerable variability among the effect sizes. This is a potential limitation to being able to fully test the emotion population hypothesis.

Conversely, a misleading inflation of Q can also occur if the number of studies is particularly large. To account for this, we also calculated H^2 (Higgins & Thompson, 2002) which is defined as $H^2 = Q/k - 1$. The H^2 statistic describes the relative difference between the observed Q and its expected value in the *absence* of heterogeneity. Thus, H^2 does not systematically increase with the number of studies included in the comparison. This allowed us to compare variability across mean effect sizes for different emotion categories and ANS outcomes where there were variable numbers of studies. According to Higgins and Thompson (2002), there are no universal rules for quantifying how much heterogeneity is present. However, they suggest these general guidelines: $H^2 = 1$ suggests a homogeneous population, H^2 values >1 suggest the presence of heterogeneity, and $H^2 > 1.5$ suggests substantial heterogeneity.

Finally, we computed the percent of variability in effect sizes due to heterogeneity using the I^2 statistic (Higgins & Thompson, 2002), which is defined as: $I^2 = \frac{H^2 - 1}{H^2} \times 100\%$. When Q and H^2 suggest the presence of variability, I^2 represents the amount of variability in effect sizes that is accounted for by heterogeneity as a proportion of the total variability. According to Higgins and Thompson's (2002) general guidelines, mild heterogeneity would be suggested by an $I^2 = 30\%$ of the variability in effect sizes, moderate heterogeneity by an I^2 between 30% and 50%, and notable heterogeneity when I^2 is >50% of the variability. We also calculated the proportion of the total observed variance that can be regarded as sampling variance for a given measure across participants within a study (level 1), within-study variance in various ANS measures (level 2) and variance across studies (level 3). Because the sampling variance depends on the study size, we computed these estimates using the median sample size of the studies in our database.

Moderator Analyses

Sample, study, and quality moderators. We conducted moderator analyses to test whether features of the sample or experimental context influenced the observed effect sizes. We used three classes of moderator variables in our analyses (as described in supplemental materials, Appendix B): *sample* characteristics (whether the sample was students, community participants, etc., and gender make-up of the sample), *study* characteristics (induction type, evaluation, use of video recording, presence of others, and presence of emotion words),[17] and *quality* moderators (baseline duration, number of emotions induced in the experiment, induction duration, use of manipulation checks, the

[17] To answer the question of whether presence of a manipulation check could have altered the mean effect size for an induction, we first examined Lench, Flores, and Bench (2011) who reported on experimental features that could have impacted the potency of an induction. They noted that very few studies included this type of information. For example, they noted (p. 841) "Attempts were made to code several characteristics that may influence the efficacy of emotion elicitations, but this information was available for only a small subset of studies; therefore, these characteristics could not be included as potential moderators. These included ethnicity, the reliability of the chosen outcome measure(s), and individual difference characteristics (e.g., depression, extraversion)." As a consequence, they instead used variables such as the proportion of female participants in an experiment, the country in which the study was conducted, the type of sample (community members vs. students), the presence of a cover story for the induction, whether participants were tested in a group or individually, and *whether or not participants were excluded on the basis of their response to the emotion induction* which seemed closest to the idea of a manipulation check. If there was no mention of excluding participants, the authors assumed all participants were included; only 12% of studies actually report excluding participants. Their results, (on their page 843), indicate that many of these methodological considerations did matter to the *overall* effectiveness of emotion inductions, but they did not test whether they accounted for variance within an induction type, nor whether they impacted the consistency and specificity of autonomic nervous system responses.

rigor of the emotion induction, and the rigor of the physiological recording).[18]

We conducted three moderator analyses for each of five emotion categories and the neutral category (one analysis for each class of moderators). Each analysis included the categorical code for ANS measure type, as before, plus the moderators of interest (sample, study, or quality variables) and the two-way interaction term for each moderator by the categorical variable indicating type of ANS measure. The moderator by ANS measure type interaction tested whether the effect of the moderator was different for each ANS measure type (i.e., did the moderator differentially effect estimates of HR, DBP, etc.). As an example, the model that tested whether the sample characteristics influenced effect sizes in the *anger* category included two moderator variables (gender composition, sample type), plus the categorical variable indicating type of ANS measure (ANS measure type), plus two, 2-way interaction terms (gender × ANS measure type *and* sample type x ANS measure type). SAS was used to compute one regression estimate (plus a *p* value) for each level of the categorical interaction term, corresponding to the effect of sample type on HR effect size, the effect of sample type on DBP effect size, and so forth). No higher order (3-way) interaction terms (such as Gender × Sample Type × ANS Measure Type) were investigated due to power and sample size limitations.

Estimating publication bias. Publication bias occurs when only studies reporting results at conventional levels of statistically significant are published, while others are not (Rosenthal, 1979). We used three methods to search for and estimate the degree of publication bias in the available data.

Funnel plots. To assess potential bias in our meta-analytic database, we started by creating funnel plots of the data (one for each emotion category). A funnel plot is a scatterplot with effect size on the *x*-axis and sample size on the *y*-axis that assesses the degree of publication bias (Light & Pillerner, 1984). Studies with smaller sample sizes will usually show more variability in effect size so the scatterplot should be wide at the bottom (i.e., where sample size is smallest) narrowing as the sample size increases,[19] creating a funnel-like shape. Deviations from this expected form suggests either (a) the presence of moderators or (b) some publication bias in our sample of studies.

We used the Egger Regression method (Egger, Davy Smith, Schneider, & Minder, 1997) to assess the amount of asymmetry in our funnel plots. This method estimates the degree of funnel plot asymmetry by calculating the intercept from the ordinary least square (OLS) regression of the effect size against the precision of the effect size estimation (the inverse of its standard error). In a nonbiased distribution, the Y-axis of the regression line originates at zero, the further the result is from zero, the greater the evidence of publication bias.

Trim and fill analyses. We conducted trim and fill analyses for each emotion category to more carefully probe potential asymmetries in the shape of our funnel plots. Trim and fill is a nonparametric method for estimating the number of missing studies in a population (Duval & Tweedie, 2000) that adjusts for missing studies and estimates a revised effect size that is approximately correct, substantially improving the coverage of effect size confidence intervals (Duval, 2005). We present data from trim and fill analyses and also note the imputed data in our funnel plots.

Sensitivity analyses. The trim and fill method, while reliable, makes strong assumptions about the symmetry of effect size distributions (e.g., all of the missing studies are those with the smallest effect sizes) which can lead to overcorrection. An alternative, more exploratory approach is a sensitivity analysis that uses weights to model the likelihood of a study being published (based on a criterion like statistical significance). We used the method described by Hedges and Vevea (1996) and Vevea and Woods (2005) to compute a sensitivity analysis because it performs well in random-effects designs, can accommodate the inclusion of moderators, and allows for the flexibility to explore specific conditions of publication bias (e.g., is there bias in only one-tailed or both tails of the distribution?). As we did in the moderator analyses, we included the categorical moderator, ANS measure type, to account for intrinsic differences in effect sizes across ANS measures. To interpret the results of these analyses we are interested in the extent to which the population effect size changes under different selection conditions. If a pattern of selection tends to favor the publication of significant effects, and the data available to meta-analyze represents effects that survived that process, then we are interested in whether the population mean effect size would be altered under different selection conditions.

We examined two-tailed selection models (i.e., effect sizes near zero are less likely to be observed, but significant correlations in either direction are favored). A more traditional take on the classical view might use a one-tailed selection model because for most ANS variables, negative correlations would usually go against the specific hypotheses, but a careful read of Kreibig (2010) convinced us that a more modern take on the classical view includes bidirectional hypotheses (see Table 1 of Quigley & Barrett, 2014). Thus, we did not model one-tailed selection (i.e., effect sizes near zero are unlikely and significant correlations in only one direction are favored).

Multivariate Pattern Classification Analysis

To determine whether there is specificity in the *pattern* of ANS features for each emotion category, we employed a multivariate pattern classification analysis (MPCA). MPCA allowed us to examine multiple ANS features at one time without violating dependency assumptions.[20] Given the extent of missing data (i.e., not all emotion categories had an equivalent number of effect sizes for each ANS variable), we were not able to run a single multiclass pattern classifier. Instead, these data were more suited to conduct-

[18] The limitations of our database prevented us from directly testing whether impediments to the *expression* of emotion (e.g., due to display rules or emotion regulation) moderated effect sizes during instances of emotion. Unfortunately, very few of the articles included in this analysis either measured or quantified these potential alterations to the expression of emotion. We addressed this to the best of our ability by testing whether the self-reported experience of emotion (in the form of a manipulation check) moderated the effect sizes, but this was an imperfect approximation.

[19] The largest samples should give the best estimates of the population effect and thus should cluster more tightly around the true population effect.

[20] In cases where a study included data from the same ANS measure (e.g., HR) from the same participants across more than one emotion category (e.g., data from both fear and anger inductions), we randomly selected data from *only one* emotion category to include. This was necessary to avoid including the same baseline data in the analysis more than once.

ing binary classifications, so we estimated a multiclass classification problem using a set of binary classifications. We did this using an error-correcting output code (ECOC) multiclass modeling approach (Allwein, Schapire, & Singer, 2001) implemented with a support vector machine (SVM) learning algorithm (Cortes & Vapnik, 1995).[21] With SVM, the algorithm is given a set of training data with each case in the training set marked as belonging to one of two binary categories (e.g., "anger" vs. "sad"). The SVM training algorithm then builds a model based on the two categories. New data are then fed into the training algorithm and predicted to belong to one or the other category (e.g., "anger" vs. "sad"). The success or failure of the classifier is measured by the percentage of cases in which the algorithm correctly predicted each category. Once the model was built from the effect size data used to train it, we then assessed how accurately the model classified a test set of effect sizes compared to chance. Because we had $c = 6$ categories in our study (anger, disgust, fear, happy, sad, neutral), and one classifier can be built to distinguish between any two categories, we therefore had to build C classifiers, or $C = c(c - 1)/2 = 15$ classifiers. These 15 classifiers were run simultaneously, with the constraint that a new case was assigned to an emotion category if (and only if) it maximized the distance between the decision boundary for that category versus all the others.

To perform the analysis, we used a 10-fold cross-validation procedure (Mosteller & Tukey, 1968). We randomly partitioned the data into 10 approximately equal-sized bins. We used nine bins to train the 15 classifiers (simultaneously, as a group) and then we used the one remaining group as the validation (test) group. We performed this across 10 separate iterations so that each bin in turn served as the validation (test) group. This allowed us to examine the sensitivity of the classifier group (the ability of each classifier to discriminate between the target category and the other categories) and also its precision (the liberal or conservative propensity of the classifier to categorize studies as the target emotion category when all else is equal; similar to the concept of "bias" in signal detection theory.[22] Due to the sparseness of the dataset (i.e., not every study recorded every ANS dependent measure), we used a matrix completion algorithm to impute missing values.

Treatment of missing values. To perform a pattern classification analysis, each case (here, a study) must include data from most (if not all) of the measures of interest. In our dataset, the majority of experiments did not report an effect size for every ANS dependent variable. In fact, most studies included fewer than half of the available ANS measures and, as a result, more than 50% of the effect size values were missing in the database. To account for this sparseness, imputation was necessary. To decrease bias during imputation (given the sparseness of our dataset), we only included studies in which more than one effect size was measured (e.g., systolic *and* diastolic blood pressure; resulting in data from 241 cases, representing each emotion manipulated in the study, from 190 published articles). We applied a matrix completion algorithm for imputation (Candès & Recht, 2009) to avoid the limitations of typical imputation methods that assume a small, randomly dispersed percentage of missing cases within a data matrix. The advantage of a matrix completion algorithm is that it uses a sampling of entries (however sparse) to estimate an underlying (latent) data matrix. Companies like Netflix, for example, use matrix completion algorithms to recommend movies to users based on their own ratings because users typically rate only very few movies (meaning that only a few scattered entries are available in the data matrix (Wang et al., 2014); completion algorithms of this type are very common in computer science (Cai, Candès, & Shen, 2010; Candès, Li, Ma, & Wright, 2011; Candès & Tao, 2010; Halko, Martinsson, & Tropp, 2011; Lin, Chen, & Ma, 2010; Wright et al., 2009).

Matrix completion reliably recovers missing entries if the data matrix in question has a low-rank structure (Candès & Recht, 2009). A low-rank structure means that every element can be written as the linear combination of a small number of latent factors. A full-rank matrix, by contrast, has independent elements as well as linearly independent rows and columns. If the classical view of emotion is true, then our observed data matrix, denoted as *X*, would have a low-rank structure because the features in our dataset (i.e., autonomic nervous system measures) should be caused by a smaller number of latent factors (i.e., the different emotion categories). This is a corollary of classical measurement theory—changes in observed variables are correlated because they have a common, latent cause—in this case, the latent construct would be anger, sadness, or fear, and so forth (for a discussion of classical measurement theory applied to the study of emotion, see Barrett, 2006a, 2011b; Coan, 2010). Thus, according to the classical view of emotion, our observed data matrix, *X*, should be a low-rank matrix.

To test whether our observed data matrix, designated as *X*, had a low-rank structure, we completed the following steps. Our null hypothesis for this test was that our observed data matrix was full-rank (i.e., that the elements could not be explained by a small set of latent factors). To test our hypothesis, we had to estimate a full-rank, null hypothesis matrix, *R,* that we could compare to our observed data matrix, *X,* to assess whether *X* was full-rank or low-rank in structure. To create our null hypothesis matrix (*R*), we independently sampled its entries from a Gaussian distribution whose mean and variance were the same as those of entries in our observed matrix, *X*. We knew that the observed entries of *R* made up a full-rank matrix because its entries (as well as rows and

[21] Before settling on SVM, we also tested other learning algorithms, specifically unsupervised classification methods. In unsupervised pattern classification, data labels are not provided to the algorithm; rather, the algorithm looks for patterns in unlabeled data. We performed an unsupervised classification procedure using a k-means approach with the number of clusters set to be at four to six (representing the emotion categories). The resulting cluster labels were very different from the class labels, meaning the effect sizes that made up each cluster did not correspond to emotion categories; the normalized mutual information between cluster and class labels (i.e., the overlap) was around 0.07. This is because the different classes (i.e., emotion categories) in our dataset did not follow Gaussian distributions, which is the key assumption of k-means clustering, and not all features (i.e., individual ANS measures) were useful in distinguishing different classes. We also attempted other unsupervised clustering techniques, such as spectral clustering and hierarchical clustering, but they performed equally poorly to the k-means approach (i.e., the clusters defined by the algorithm did not map on to the true emotion category labels).

[22] This lack of precision in the classifier is analogous to bias as described in signal detection theory. In our MPCA, correct detection of anger was fairly high (i.e., the classifier was relatively good at predicting when a case was anger). However, the classifier's false alarm rate was also high (nearly as high as the true positive or hit rate) indicating that the classifier also frequently predicted that cases were anger that, in fact, belonged to another category. Thus, despite its relative success at predicting true cases of anger it was still a poor anger classifier because it also had a "liberal" anger bias.

columns) were statistically independent (i.e., could not be explained by a small number of latent factors). Note, the location of the observed entries in our random, null hypothesis matrix, R, were the same as they were in our observed matrix, X. We then applied the matrix completion algorithm outlined in Candès & Recht (2009) to our observed matrix, X and our null hypothesis matrix, R. The outputs are denoted as X^* and R^* respectively. We then repeated this process 10 times to avoid the randomness of a single run. During each run of the completion algorithm, for both our imputed observed matrix, X^* and our imputed random matrix, R^*, we tested how well a small number of latent factors accounted for the variance in our matrices, by computing an explained-variance-ratio (see, Eq. 12.69 in Murphy, 2012). If a small number of factors accounted for a large portion of the variance (creating a high ratio value), then this would suggest that the matrix is low-rank. We tested whether X^* had a higher explained-variance-ratio than R^* (which we constructed as full-rank, and therefore, by design needed many latent factors to account for its variance); this test involved comparing the mean and variance of the explained-variance-ratios for X^* and R^* across our 10 runs. We found that the average ratio for X^* far exceeded R^*, indicating that fewer latent factors explained a larger proportion of the variance in X^* than in R^*. In fact, only two latent factors accounted for 87% of the variance in X^* (whereas five latent factors were needed to account for the same amount of variance in R^*). This suggested that our observed matrix, X, was low-rank and therefore the Candès and Recht (2009) matrix completion method was suitable for imputing the missing effect sizes.

Results

The results are presented in four parts. In the first section we provide descriptive results from the meta-analytic database. In the second section, we describe the kinds of ANS measures that were collected over the more than 60-year period covered by the studies we sampled and then examine possible biases in publication. In the third section, we compare the emotion fingerprint hypothesis associated with the classical view of emotion and the emotion population hypothesis associated with our theory of constructed emotion by assessing the consistency or variability in ANS patterns within each emotion category as well as the specificity of ANS patterns across emotion categories. In the fourth section we discuss the results of our moderator analyses and determine if features of the context can explain variability in ANS reactivity.

Descriptive Results

Study characteristics. Study characteristics are presented in Tables 4 and 5. There are 8,443 *unique* individuals across all included studies (meaning that this is the total n across all studies) with many individuals contributing data to more than one effect size for a total of $n = 28{,}788$ data points across all meta-analytic comparisons. Because not all studies reported demographic statistics, we cannot report specific information about age and sex; however, the type of sample (students or members of the community) and the gender composition of each sample (all male, mixed, all female) were included as sample moderators in our analyses (see Table 11 and supplemental materials, Appendix B).

Distributions of effect sizes. Distributions of the effect sizes across emotion categories and ANS variables are presented in Tables 4 and 5, and in Figure 2. Across all emotion categories, heart rate (HR) was measured most frequently, with twice as many effect sizes as for any other measure. Thus, emotion categories and ANS measures were not evenly represented in the published literature (see Tables 4 and 5, and Figure 2). Certain combinations of emotion categories and ANS measures occurred frequently in published experiments (e.g., inducing anger and measuring blood pressure) whereas other combinations were rare or nonexistent (e.g., inducing sadness while measuring stroke volume).

Evaluations of Publication Bias

The assessment of potential emotion fingerprints has been a subject of empirical investigation for more than a century, during which there could have been changes in the emotion categories studied and ANS measures used (e.g., due to technical or scientific cultural shifts) that could subtly influence the meta-analytic outcomes. To assess changes over time, we plotted the number of effect sizes for each emotion category (see Figure 3) and each ANS variable (see Figure 4) over time. Some emotion categories and some ANS variables appeared more frequently in the literature than did others during some time periods, although we did not find any evidence that this variation contributed to systematic shifts in effect sizes over time.

Funnel plots with imputed data. In Figure 5 (panels a–f), we present funnel plots for each emotion category including the imputed data from the trim and fill analyses. Funnel plots for all six emotion categories appeared asymmetric and therefore, trim and fill analyses were conducted for each emotion category.

Egger's regression test. Results from our regression tests indicated that there was significant asymmetry only in the anger category, $t(283) = 2.27$, $p = .01$. This suggests the possible presence of bias in the anger category.

Trim and fill. Results from our trim and fill analysis are presented in Table 7. The anger, fear, and neutral categories required the most imputation with 17%, 19%, and 23% of effect sizes, respectively.

Sensitivity analyses. We estimated the population effect sizes under moderate and severe two-tailed selection bias for each emotion category (see Table 8). We found a reduction in the magnitude of the mean effect size for all six emotion categories when we modeled selection bias. The largest changes occurred when we modeled severe two-tailed selection in the sad category (−53%), followed by happy (−44%), neutral (−44%), disgust (−33%), anger (−31%), and fear (−26%). In addition, variation around the mean increased when we modeled two-tailed selection bias with the largest increases in variation in the happy category (19.5%), followed by sad (16.1%), anger (14.5%), neutral (3.5%), and fear (2.5%). Variation in the disgust category decreased (−73.7%). In three emotion categories—sad, happy and anger—not only did effect size estimates change considerably depending on the selection model used, but heterogeneity also increased. Under conditions of significantly altered effect size and heterogeneity like these, both Vevea and Woods (2005) and McShane, Böckenholt, and Hansen (2016) suggest that publication bias may be driving the estimates, meaning that the true effect size would be even smaller than we report, and the findings more variable, in the absence of publication bias. In the fear condition, effect size magnitude decreased some, and variation did not change appre-

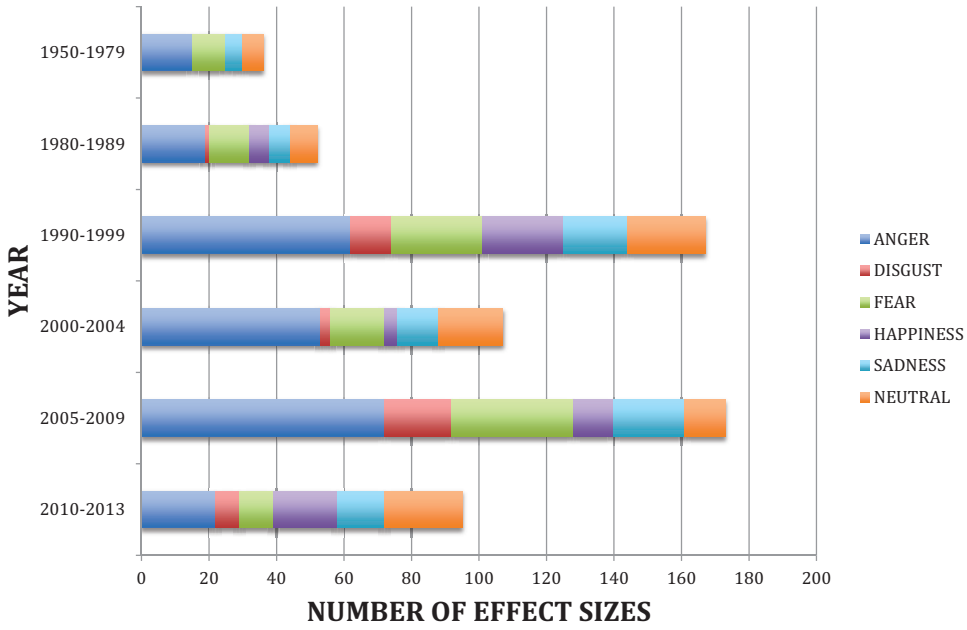

Figure 3. Number of effect sizes for each emotion category as a function of publication date. See the online article for the color version of this figure.

ciably. For the disgust and neutral categories, the findings were more mixed; the effect size would be lower without selection bias than we report, but variation did not change much, and in fact, variation decreased in the disgust category. When estimates and variation are relatively stable, regardless of the selection model assumed, it suggests that publication bias is unlikely to drive effect sizes (McShane, Böckenholt, & Hansen, 2016) suggesting that fear, disgust and neutral category findings are unlikely to be greatly impacted by publication bias.

Testing Fingerprint Versus Population Hypotheses

Evaluation of specificity and consistency. We examined the mean effect sizes across ANS measures to evaluate the specificity of these changes for each emotion category. We then examined several measures of effect size heterogeneity to evaluate the consistency of changes within each emotion category. In Table 9 we present all mean effect sizes and heterogeneity statistics (Cochran's Q, H^2, and I^2), and in Figures 6 and 7 we

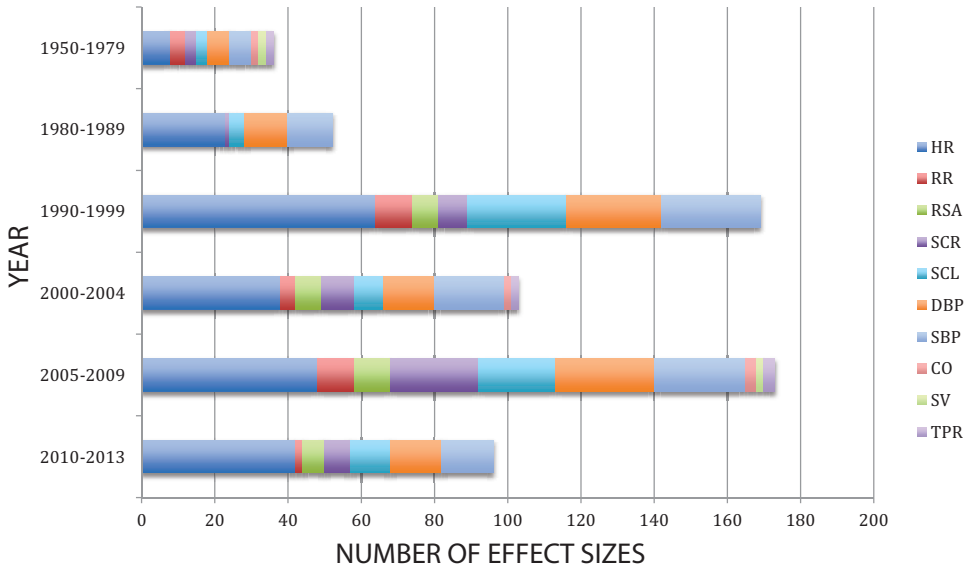

Figure 4. Number of effect sizes for each autonomic nervous system measure as a function of publication date. See the online article for the color version of this figure.

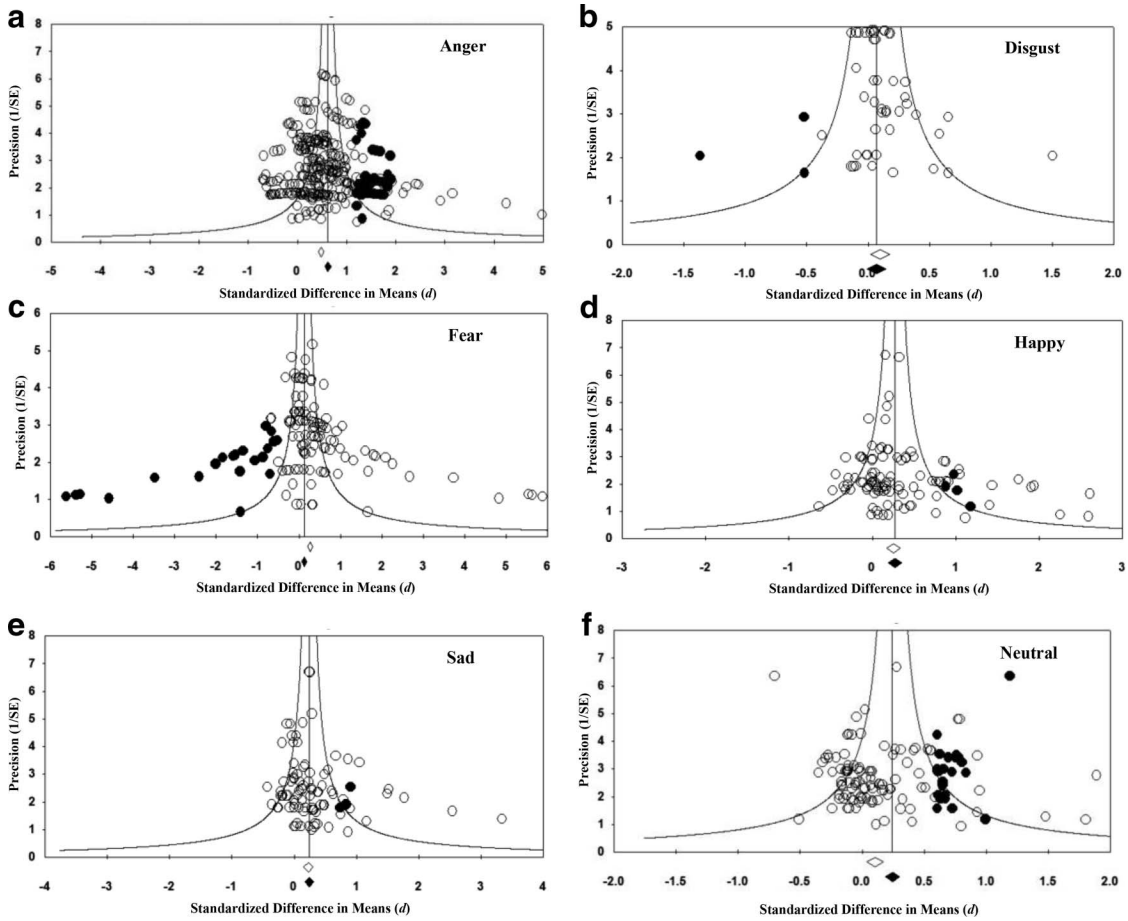

Figure 5. a–f. Funnel plots of individual effect sizes in our meta-analytic database separated by emotion category. In these plots, precision (1/standard error) is on the *y*-axis and estimated effect size on the *x*-axis. Because studies with smaller sample sizes will usually show more variability in effect size, the scatterplot should be the widest at the bottom (i.e., where sample size is smallest) and should progressively narrow as it moves up the *y*-axis (i.e., as the sample size increases) creating a funnel-like shape. Deviations from this expected form suggests either (a) the presence of moderators or (b) some publication bias in our sample of studies. Funnel plots for all six emotion categories contained some amount of bias and required imputation (filled black circles denote the imputed missing studies). The diamonds shown below the *x*-axis illustrate the mean difference (*d*) before (white diamonds) and after (black diamonds) imputation. Data from the trim and fill analysis are reported in Table 5.

depict these mean effect sizes and their 95% CIs for each ANS measure (for individual effect sizes, see Table 6). Overall, we found evidence of substantial variation within each category (lack of consistency) coupled with nonspecific ANS changes that occurred across multiple emotion categories (lack of specificity).

We computed mean effect sizes for 58 combinations of ANS variables and emotion categories. Of these, 35 tests were sufficiently powered to permit meaningful interpretation ($k > 4$). We found significant differences in the mean effect size for 22 of these 35 tests (62.8%). However, the I^2 statistic suggested that heterogeneity explained a moderate to substantial percentage (i.e., $I^2 \geq 30\%$) of the variability in 16 of these 35 effect sizes (45.7%). Consistent with this, the Cochran's Q statistic was significant for 13 of the 35 (59.1%) Q tests where $k > 4$ (see Table 9) and H^2 exceeded one in 18 of the 35 cases (51.4%) suggesting considerable variability in the effect sizes. In general, the pattern of ANS changes for each emotion category can be described in one of three ways: (a) mean ANS changes from baseline across several effect sizes but with substantial variability (anger and fear categories); (b) few changes in mean effect sizes and moderate variability (disgust and neutral categories); and (c) substantial ANS changes from baseline with moderate variability and indications of publication bias (sad and happy categories).

Consistency in anger and fear categories. Instances of anger resulted in large mean effect sizes (increases from baseline) for several of the ANS measures, namely, HR, CO, DBP, SBP, RR, and SCR (see Table 9). For all but one measure (SCR), however, heterogeneity was substantial (as indicated by a minimum of two of these indices: significant Q, $H^2 > 1$, or $I^2 > 50\%$). The SCR effect size for anger was heterogeneous by only

Table 7
Mean Effect Sizes for Each Emotion Category and Imputed Data from Trim and Fill Analysis

Emotion category	*k*	% Imputed	*d* Adjusted	CI	*Q*
Anger					
Actual	286		.48	.42, .54	615.63
Imputed	49	17%	.62	.55, .68	895.22
Disgust					
Actual	51		.10	.02, .18	28.32
Imputed	7	14%	.07	−.01, .14	45.08
Fear					
Actual	111		.40	.28, .52	353.15
Imputed	21	19%	.14	−.01, .29	676.88
Happy					
Actual	99		.27	.18, .36	117.62
Imputed	4	4%	.29	.20, .38	124.72
Sad					
Actual	87		.25	.15, .35	116.46
Imputed	2	2%	.27	.17, .36	121.55
Neutral					
Actual	105		.11	−.33, .19	135.32
Imputed	25	23%	.24	.16, .32	226.31

Note. Actual = Observed mean effect sizes, confidence intervals and *Q* by emotion category; Imputed = Number of effect sizes that were imputed in trim and fill analysis; % Imputed = percentage of effect sizes that required imputation; *d* Adjusted = adjusted effect size after trim and fill.

one of these statistics (H^2). Like anger, instances of fear resulted in substantial increases from baseline in HR, SBP, RR, and SCR, and all four effect sizes were heterogeneous as indicated by at least two of the three heterogeneity statistics (with the fourth one, SCR, just missing the heterogeneity cutoff with an $I^2 = 50\%$). These results suggest that although there are changes from baseline in the mean ANS effect sizes for both anger and fear categories, all but one (SCR for anger) were heterogeneous, suggesting substantial variability in nearly every effect size.

Table 8
Sensitivity Analysis of Selection Bias for Each Emotion Category[1]

Emotion category	Selection condition: Actual *d*	Actual CI	Moderate two-tailed *d*	Moderate two-tailed CI	Severe two-tailed *d*	Severe two-tailed CI
Anger	.51	.30, .72	.43	.23, .65	.35	.12, .58
Disgust	.09	.02, .16	.08	.07, .08	.06	.05, .06
Fear	.47	−.10, 1.08	.41	−.16, 1.02	.35	−.23, .97
Happy	.32	.22, .43	.25	.15, .36	.18	.05, .30
Sad	.3	.16, .43	.22	.08, .36	.14	−.01, .03
Neutral	.12	.06, .19	.09	.03, .16	.07	.001, .14

Note. Sensitivity analyses use weights to model the likelihood of a study being published (based on a criterion like statistical significance; Hedges & Vevea, 1996; Vevea & Hedges, 1995; Vevea & Woods, 2005). We included two-tailed selection models (i.e., effect sizes near zero are less likely to be observed, but significant correlations in either direction are favored) and did not model one-tailed selection (i.e., effect sizes near zero are unlikely and significant correlations in only one direction are favored) because it is unlikely for reactivity across the autonomic nervous system to change in only one direction. Of interest is the extent to which the population effect size changes under different selection conditions. If a pattern of selection tends to favor the publication of significant effects, and the data available to meta-analyze represents effects that survived that process, then it is integral to know whether (and to what extent) the population mean effect size is robust to differences in publication selection conditions. Actual = Observed mean effect sizes and confidence intervals for each category.

[1] The categorical variable ANS measure type was included in every model to account for differences in effect sizes across different autonomic measures. Moderate and Severe = the extent of the selection bias modeled in this analysis.

Consistency in disgust and neutral categories. For disgust, only two ANS measures were sufficiently powered to merit consideration ($k > 4$): SCL, and SCR (see Table 9). Both SCL and SCR had an increased mean effect sizes (relative to baseline; SCL $d = 0.25$, CI [.03, .47] and SCR $d = .48$, CI [.16, .82]), but only SCR was homogeneous, whereas heterogeneity in SCL exceeded the cutoff on two of three indices (see Table 9). For neutral inductions, only SBP had an increased mean effect size (relative to baseline; $d = .32$, CI [.11, .52]) but it also had substantial heterogeneity via all three heterogeneity statistics. Together, these results suggest that there was only one significant increase in mean effect size for instances of either disgust or neutral inductions that was not also highly variable.

Consistency in happy and sad categories. Instances of happiness resulted in increases in mean effect sizes (relative to baseline) in the form of increases from baseline in HR, DBP, RR, SCL, and SCR (see Table 9), and the mean effect sizes were heterogeneous for HR, although both DBP and RR (were heterogeneous by one of the three statistics). DBP, RR, and SCR had the strongest evidence for consistency as indicated by the largest mean effect size increases (DBP $d = .52$, CI [.18, .88]; RR $d = .47$, CI [.13, .83]; SCR $d = .50$, CI [.12, .89]) coupled with little evidence of heterogeneity (i.e., only the H^2 index exceeded 1 for DBP and RR). Instances of sadness (where $k > 4$) resulted in increases from baseline in HR, DBP, SBP, and RR. However, the mean effect sizes were heterogeneous for HR, DBP, and SBP (i.e., for two of the three heterogeneity statistics). Only RR showed evidence of consistency (RR $d = .41$, CI [.02, .82]) without evidence of heterogeneity (see Table 9).

Table 9
Summary of Multilevel Meta-Analysis for Emotion Categories

Emotion category and physiological measure	*k* (*n*)	*d*	95% CI (Lower, Upper)	*Q*	H^2	I^2(%)
Anger						
Heart rate	86 (3,684)	.52***	.39, .66	243.00***	7.00	65.02
Heart rate variability	9 (252)	−.18	−.52, .16		1.03	2.88
Pulse transit time	0					
Cardiac output	8 (398)	.44**	−.11, .87	18.64	2.66	64.44
Stroke volume	4 (105)	.64*	.13, 1.18	32.09***	10.70	90.65
Total peripheral resistance	8 (398)	.74	.43, 1.06	2.32	.33	
Diastolic blood pressure	60 (2,646)	.57***	.42, .72	251.31***	16.17	76.52
Systolic blood pressure	63 (2,782)	.66***	.51, .81	215.39***	11.38	71.68
Mean arterial pressure	2 (82)	.65*	.03, 1.34			
Temperature	2 (68)	.09	−.56, .76		0	
Respiration rate	5 (109)	1.47**	.89, 2.13	40.36***	9.95	90.09
Expiratory time	1 (16)					
Inspiratory time	2 (32)	.57	−.49, 1.80	1.22	1.22	18.30
Tidal volume	0					
Skin conductance level	22 (772)	.19	−.04, .42	47.36*	2.26	55.66
Skin conductance response	14 (476)	.45*	.12, .92	15.96	1.23	15.65
Total	286 (11,820)					
Disgust						
Heart rate	12 (569)	.07	−.15, .28	4.11	.37	
Heart rate variability	4 (180)	.09	−.26, .44	.92	.31	
Pulse transit time	0					
Cardiac output	2 (196)	.01	−.36, .37	.5	.5	
Stroke volume	2 (196)	.03	−.33, .40	.16	.16	
Total peripheral resistance	2 (196)	.22	−.14, .59	.17	.17	
Diastolic blood pressure	2 (200)	.20	−.17, .56	.12	.12	
Systolic blood pressure	2 (200)	.21	−.15, .58	.09	.09	
Mean arterial pressure	0					
Temperature	1 (41)					
Respiration rate	2 (36)	.08	−.68, .86	.02	.02	
Expiratory time	1 (16)					
Inspiratory time	2 (32)	.20	−.66, 1.10	.72	.72	
Tidal volume	0					
Skin conductance level	9 (558)	.25*	.03, .47	4.01	3.88	67.65
Skin conductance response	10 (231)	.48**	.16, .82	7.50	.95	
Total	51 (2,651)					
Fear						
Heart rate	37 (1,555)	.49***	.23, .76	210.81**	5.86	82.92
Heart rate variability	5 (254)	.45	−.11, 1.06	19.62*	4.90	79.61
Pulse transit time	2 (102)	.16	−.71, 1.06	.03	.03	
Cardiac output	1 (48)					
Stroke volume	1 (48)					
Total peripheral resistance	1 (48)					
Diastolic blood pressure	10 (323)	.14	−.31, .60	6.70***	10.94	90.86
Systolic Blood Pressure	11 (335)	.57*	.12, 1.04	23.33*	2.33	57.13
Mean arterial pressure	0					
Temperature	1 (41)					
Respiration rate	6 (193)	.91**	.31, 1.58	98.44***	19.69	94.92
Expiratory time	2 (58)	−.11	−1.13, .88	.10	.10	
Inspiratory time	3 (74)	.17	−.72, 1.09	1.96	.98	
Tidal volume	1 (42)					
Skin conductance level	17 (725)	.19	−.16, .55	14.32	.90	
Skin conductance response	13 (473)	.61**	.20, 1.03	31.67*	1.94	48.56
Total	111 (4,271)					
Happy						
Heart rate	36 (1,177)	.33**	.13, .52	52.80*	1.53	33.72
Heart rate variability	9 (216)	.19	−.17, .55	7.87	.98	
Pulse transit time	0					
Cardiac output	0					
Stroke volume	0					
Total peripheral resistance	0					
Diastolic blood pressure	10 (226)	.52**	.18, .88	13.24	1.47	32.00
Systolic blood pressure	9 (188)	.37	−.01, .76	6.46	.81	
Mean arterial pressure	0					

(*table continues*)

Table 9 (*continued*)

Emotion category and physiological measure	*k* (*n*)	*d*	95% CI (Lower, Upper)	*Q*	H^2	I^2(%)
Temperature	1 (41)					
Respiration rate	9 (231)	.47**	.13, .83	10.25	1.28	21.94
Expiratory time	1 (16)					
Inspiratory time	2 (32)	.33	−.67, 1.41	.55	.55	
Tidal volume	0					
Skin conductance level	13 (574)	.35**	.10, .62	8.35	.70	
Skin conductance response	9 (198)	.50*	.12, .89	7.86	.98	
Total	99 (2,899)					
Sad						
Heart rate	32 (1,424)	.36*	.03, .69	57.48*	1.86	46.06
Heart rate variability	5 (93)	−.02	−.56, .51	.63	.16	
Pulse transit time	1 (72)					
Cardiac output	0					
Stroke volume	0					
Total peripheral resistance	0					
Diastolic blood pressure	9 (275)	.36*	.03, .69	14.86*	1.86	46.17
Systolic blood pressure	8 (237)	.45*	.10, .82	10.79*	1.54	35.10
Mean arterial pressure	0					
Temperature	1 (41)					
Respiration rate	7 (172)	.41*	.02, .82	3.81	.53	
Expiratory time	1 (16)					
Inspiratory time	2 (32)	.11	−.90, 1.15	0	0	
Tidal volume	2 (31)	−.10	−1.05, .84	.05	.05	
Skin conductance level	15 (552)	.18	−.08, .44	5.65	.40	
Skin conductance response	4 (114)	.45*	.12, .92	1.68	.56	
Total	87 (3,059)					
Neutral						
Heart rate	41 (1,641)	.21	−.01, .43	34.62*	.87	
Heart rate variability	4 (134)	−.09	−.52, .35	.03	.01	
Pulse transit time	0					
Cardiac output	1 (37)					
Stroke volume	1 (37)					
Total peripheral resistance	1 (37)					
Diastolic blood pressure	16 (549)	.21	−.07, .43	19.04	1.27	21.20
Systolic blood pressure	18 (674)	.32**	.11, .52	45.77*	2.69	62.80
Mean arterial pressure	0					
Temperature	0					
Respiration rate	4 (258)	−.09	−.40, .60	.16	.05	
Expiratory time	0					
Inspiratory time	0					
Tidal volume	0					
Skin conductance level	11 (515)	−.13	−.39, .12	16.61*	1.66	39.80
Skin conductance response	8 (258)	.15	−.16, .46	8.09	1.16	13.43
Total	105 (3,976)					

Note. *k* (or *N*) = number of independent samples; *n* = number of participants; *d* = Cohen's *d*; CI = 95% confidence interval above and below the mean *r*. *Q* = heterogeneity statistic H^2 = heterogeneity statistic derived from *Q* (reported in the manuscript). Unlike *Q*, H^2 allows for easier comparison across effect sizes. Based on the guidelines laid out in Higgins and Thompson (2002), $H^2 = 1$ suggests a homogeneous population, $H^2 > 1$ suggests the presence of heterogeneity, and $H^2 > 1.5$ suggest substantial heterogeneity. I^2 = percent of variability in effect sizes that is due to heterogeneity; it is shown here when $H^2 > 1$. In general, based on I^2, mild heterogeneity is assumed to account for < 30% of the variability in effect sizes, moderate heterogeneity between 30%–50%, and substantial heterogeneity when $I^2 > 50\%$ (Higgins & Thompson, 2002).
* $p < .05$. ** $p < .01$. *** $p < .001$.

Specificity across emotion categories. None of the mean ANS changes were specific to any emotion category (see Figures 6 and 7). Notably, the mean increases in SCR were similar and large, significant increases from baseline (*d*s = 0.45–0.61) in all but the neutral case ($d = 0.15$). Similarly, the happy category was associated with a mean increase in SCL ($d = 0.35$) that was similar to that for disgust ($d = 0.25$) and also fell within the 95% CIs for the (nonsignificant) mean effect sizes for anger ($d = 0.19$), fear ($d = 0.19$), and sad ($d = 0.18$). Another example of the similarity in mean effect sizes is shown by the RR effect size for the happy ($d = 0.47$) and sad ($d = 0.41$) categories. Finally, Table 9 also reveals that of the 22 significant increases in mean effect sizes (where $k > 4$), only two of them (RR for anger and fear) were ≥ .80 (often considered a large effect size; Cohen, 1988), seven (HR, DBP and SBP for anger, SBP and SCR for fear, and DBP and SCR for happy) were ≥ .50 (medium effect size) and the remaining 13 (two for anger, two for disgust, one for fear, three for happy, four for sad, and one for neutral) were above .20 (small effect size). Considered alongside the notable variability in the mean ANS effect sizes across emotion categories and considering the lower limit of the 95% CIs, only one remained a strong effect

size (RR in anger) and only one remained moderate (SBP in anger), and both of which are also nonspecific they are both also different (with a similar magnitude) from baseline in fear. Thus, the mean changes that occur in ANS features are neither emotion-specific nor robust when considered in light of the extensive variability.

Multivariate pattern classification analysis. Given how many heterogeneous effects and similarities across emotion conditions we found for several ANS measures, we considered the possibility that the traditional, univariate approach was incomplete for testing the emotion fingerprint versus populations views, and that a multivariate test was warranted. We present the results of our multivariate pattern classification analysis in Table 10 in a confusion matrix, which describes the performance of a classification model for test data for which true values are known. Overall performance of the classifiers was low (31.5%). The anger classifier was the most sensitive and was the only classifier to make more correct predictions (i.e., true positives; 51.7% of cases) than incorrect predictions, and even here, the number of false positives (48.3% of cases) were almost as high as the number of true positives.[23] Given that anger was the best classifier, we also calculated a Youden's *J* statistic for the anger classifier. Youden's *J* incorporates both the true positive rate and true negative rate of the classifier and represents this as a number between 0 (for a useless classifier) and 1 (or −1; for a perfect classifier). Even the best classifier, for anger, had a Youden's *J* of only 0.3 and therefore, was closer to zero than to one. Overall, the multivariate pattern classifiers did provide strong evidence of a consistent multivariate pattern for any emotion category.

Moderator Analyses

Full results of our moderator analyses are reported in Table 11 and supplemental materials, Appendix D. Unfortunately, there were too few effect sizes for the disgust category ($k = 51$) for our moderator models to converge. Thus, we report moderator analyses only for the anger, fear, happy, sad, and neutral categories. As predicted by the emotion population hypothesis, moderator effects were not consistent across emotion categories and substantial variability remained even after accounting for the variance in effect sizes due to the moderators.

Variance estimates. Based on variance estimates from multilevel models without any predictor variables, we computed variation in effect sizes, both between studies and between effect sizes from the same study for each emotion category. Quantifying the variation in effect sizes between studies and within studies provided information about how much variance in effect sizes our moderator models would be able to explain (Houben, Van Den Noortgate, & Kuppens, 2015).

In the anger category, we found substantial variation between studies ($\sigma^2_V = 0.043$, $\chi^2(1) = 12.24$, $p < 0.001$) but not between effect sizes from the same study ($\sigma^2_U = 0.007$, $\chi^2(1) = 2.10$, $p > 0.05$). Variance at level 1, the sampling variance, depends on the sample size of the study. Therefore, we looked at the median estimated value, which was 0.03. If we add the level-1 variance (0.03) to the variation between and within studies (0.043 and 0.007, respectively), this suggests that total variance at all levels was 0.08. Thus, for the median study, approximately 53% of the total variance in observed effect sizes was accounted for by variance between studies, approximately 9% by variance between effect sizes of the same study, and approximately 38% by sampling variance. The large proportion of between study variance in the anger category suggests that effect sizes were not consistent across studies and this difference is larger than we would expect based on random fluctuations (i.e., in some studies, the effect sizes were systematically larger than other studies). The low proportion of variance in effect sizes within the same study suggests that within studies, observed effect sizes did not vary much more than could be explained by sampling variance.

For the fear category, we found substantial variation between studies ($\sigma^2 = 0.055$, $\chi^2(1) = 7.48$, $p < 0.01$) and between effect sizes from the same study ($\sigma^2 = 0.055$, $\chi^2(1) = 7.43$, $p < 0.01$). The estimated sampling variance was 0.028. Consequently, approximately 40% of the variance was accounted for by variance between studies, 40% within studies, and 20% by sampling variance. The large amount of between and within study variance in the fear category suggests that some effect sizes were systematically larger than in other studies and that effect sizes differed within studies more than could be explained by sampling variance.

For the happy category, we found substantial variation between studies ($\sigma^2 = 0.041$, $\chi^2(1) = 3.88$, $p < 0.05$) but not between effect sizes from the same study ($\sigma^2 = 0$). The estimated sampling variance was 0.05. Consequently, approximately 41% of the variance was accounted for by variance between studies and 59% by sampling variance. In the happy category, variation across studies was greater than would be expected by random fluctuations but within studies, effect sizes did not differ meaningfully.

For the sad category, we found some variation between studies ($\sigma^2 = 0.03$, $\chi^2(1) = 2.70$, $p < 0.10$) and a small amount of variation in effect sizes from the same study ($\sigma^2 = 0.003$, $\chi^2(1) = 0.25$, $p > 0.1$). The estimated sampling variance was 0.04. Consequently, approximately 40% of the variance was accounted for by variance between studies, 4% within studies, and 56% by sampling variance. In the sad category, variation across studies was greater than would be expected by random fluctuations but within studies, effect sizes did not differ meaningfully.

For the neutral category, we found little variation either between studies ($\sigma^2 = 0.01$, $\chi^2(1) = 0.88$, $p > 0.10$) or between effect sizes from the same study ($\sigma^2 = 0.01$, $\chi^2(1) = 0.69$, $p > 0.10$). The estimated sampling variance was 0.04. Consequently, approximately 15% of the variance was accounted for by variance between studies, 12% within studies, and 73% by sampling variance.

[23] To check whether sampling bias in the literature (e.g., more effect sizes for blood pressure during anger inductions than disgust inductions) was biasing our classifier in a systematic way, we removed several ANS measures one at a time and then re-ran the MPCA. In the first two MPCAs, we removed SBP and DBP, respectively, because they have been most notably oversampled in the literature on anger. We also ran two additional MPCAs, in one removing SCL and in the other removing HR, again because these have been most commonly sampled, especially in studies of anger. Finally, we tested five additional versions of the MPCA in which we used a random number generator to randomly remove from each MPCA one of the other remaining ANS measures; those randomly chosen were CO, RR, HRV, SV, and SCL). Although, changing the data utilized in these MPCAs decreased the overall accuracy of the classifier, the pattern of results remained the same in every analysis (the classifier predicted anger and neutral better than other emotion categories). Further, we did not see an increase in the classification accuracy in any other emotion category suggesting that this pattern of results is relatively robust.

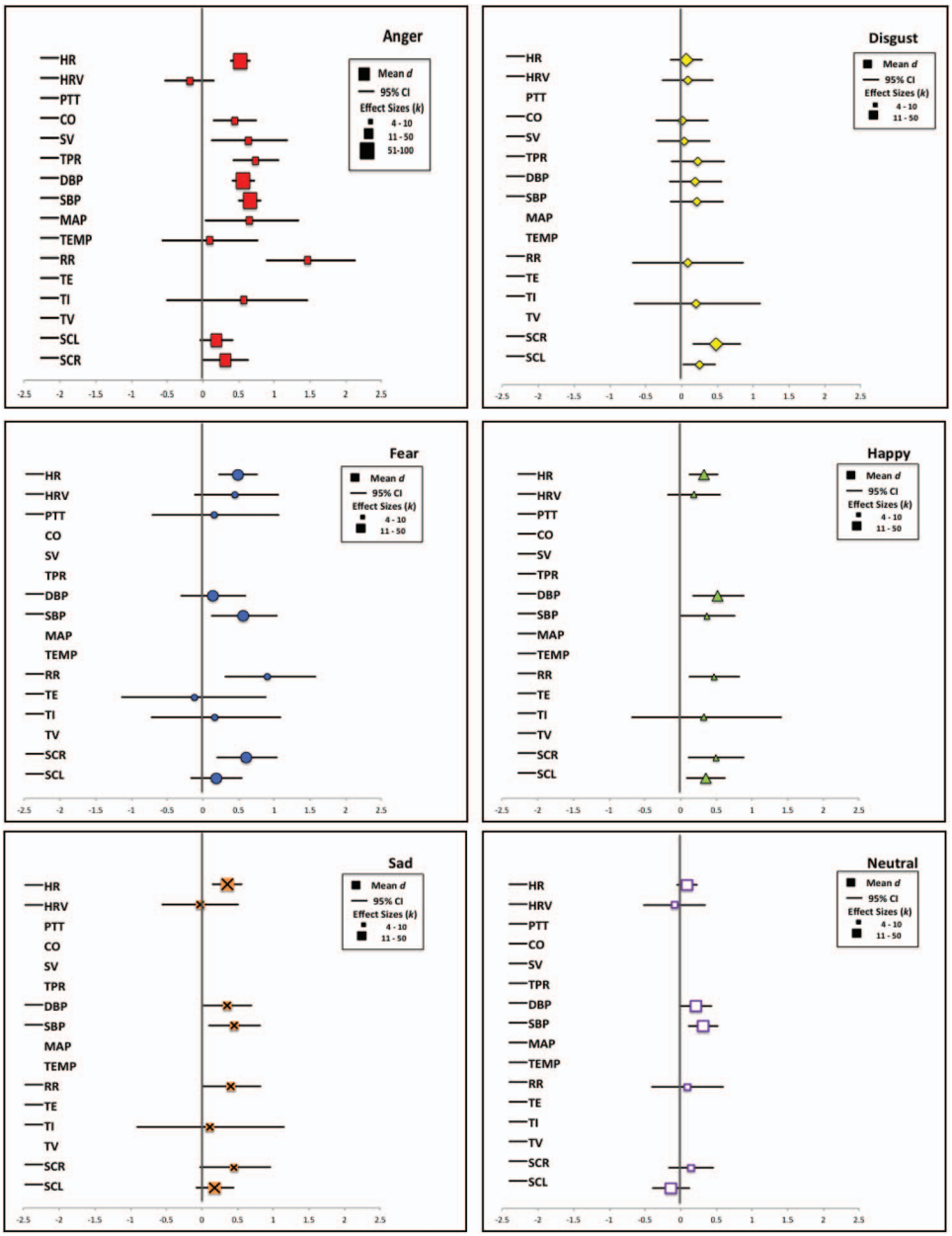

Figure 6. Mean effect sizes plotted for each emotion category with confidence intervals. Larger squares indicate larger numbers of effect sizes in the comparison. See the online article for the color version of this figure.

Sample moderators. Overall, our two sample moderators, gender composition and sample type (students vs. community members), had the strongest impact on effect sizes across emotion categories (see Table 11 and supplemental materials, Appendix D). The gender composition of the sample moderated effect sizes for both the anger and neutral categories. For anger, we found a significant two-way interaction between the ANS outcome measure and gender, $F(16, 156) = 1.98$, $p = .02$. This effect was driven largely by studies with all women which tended to have higher mean DBP ($\beta = 0.99$) and SBP ($\beta = 1.14$) effect sizes than did studies with both men and women (DBP: $\beta = 0.72$, SBP: $\beta = 0.84$) or studies with all men (DBP: $\beta = 0.76$, SBP: $\beta = 0.67$; see Table 11 and supplemental materials, Appendix D). We also uncovered an ANS reactivity by gender interaction in the neutral category, $F(10, 75) = 2.85$, $p = .004$, reflecting the fact that studies with all women had greater increases in SBP ($\beta = 0.47$), DBP ($\beta = 0.99$), and SCR ($\beta = 0.98$) during neutral inductions compared with ANS reactivity in studies with both men and women (SBP: $\beta = 0.44$, DBP: $\beta = 0.11$, SCR: $\beta = -0.002$) or studies with all men (SBP: $\beta = 0.13$, DBP: $\beta = 0.27$, SCR: $\beta = 0.02$).

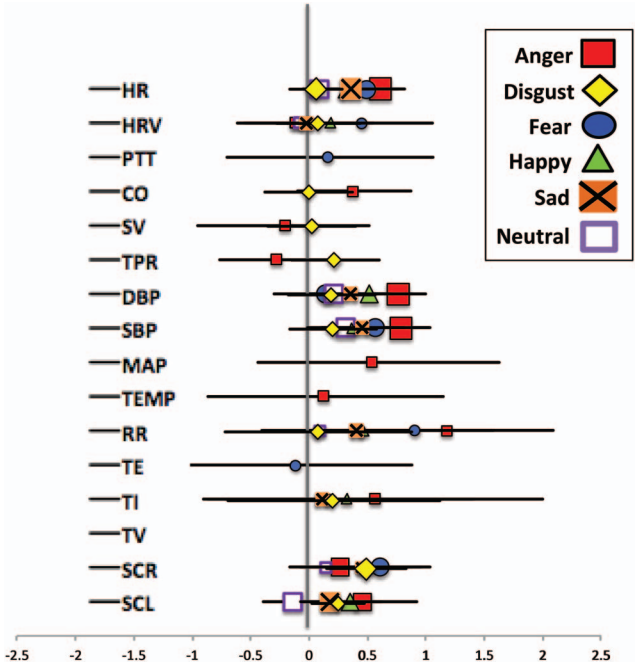

Figure 7. Mean effect sizes and confidence intervals for each of the six emotion categories. Larger shapes indicate greater numbers of effect sizes within the comparison. See the online article for the color version of this figure.

The sample type (i.e., whether participants were undergraduate students or community members) moderated the effect sizes for the anger, fear, and sad categories. Across ANS measures, studies of anger with undergraduate student samples resulted in small decreases from baseline on average ($\beta = -0.11$) whereas samples of community members resulted in increases from baseline ($\beta = 0.20$), $F(1, 52.4) = 5.16$, $p = .03$. Across ANS measures, fear studies with student samples had larger decreases from baseline ($\beta = -0.27$) whereas studies with community members had smaller increases from baseline ($\beta = 0.11$), $F(1, 21.3) = 11.57$, $p = .003$. Further, for fear we also found an ANS reactivity by sample type interaction, $F(6, 59) = 3.93$, $p = .003$. This interaction arose from the fact that studies sampling college students had very small changes from baseline in RR ($\beta = -0.002$) and HR ($\beta = -0.06$) whereas studies with community members had notably larger increases from baseline in both RR ($\beta = 1.53$) and HR ($\beta = 0.83$). We also found an ANS reactivity by sample type interaction for the sad category, $F(5, 64) = 2.60$, $p = .03$. In this case, studies that sampled students showed smaller increases from baseline in RR ($\beta = 0.22$) and HR ($\beta = 0.21$) compared with studies that sampled community members, which showed larger increases from baseline in RR ($\beta = 0.99$) and HR ($\beta = 0.51$).

When we calculated how much of the overall variance both between and within studies was explained by sample moderators (and compared that to the models without predictors) we found that sample moderators accounted for between and within study variance only for the fear category where 4% of the variance was explained.

Study moderators. Study moderators explained no significant variance across emotion categories (see Table 11). Study moderators influenced the overall proportion of unexplained variance only for the fear category where 3% of the variance was explained.

Quality moderators. Quality moderators did not influence many of the effect sizes across emotion categories, but they did account for some variation (see Table 11 and supplemental materials, Appendix D). In the anger category, the rigor of the physiological recording methods moderated the mean effect sizes, $F(2, 30.6) = 2.10$, $p = .04$, such that effect sizes from studies with less rigorous recording procedures were higher ($\beta = 1.20$) than effect sizes from studies with moderate ($\beta = 0.84$) or highly rigorous recording procedures ($\beta = 0.97$). For the happy category, studies with emotion inductions that were less than a minute long had larger ANS increases from baseline ($\beta = 1.62$) compared with a tendency for decreased ANS activity from baseline for inductions that were 1–5 min ($\beta = -0.28$), 5–10 min ($\beta = -0.34$), or more than 10 min ($\beta = -0.38$), $F(3, 11) = 6.57$, $p = .008$. In the neutral category, studies without manipulation checks had larger increases from baseline ($\beta = 0.22$) than studies with manipulation checks ($\beta = 0.04$), $F(1, 8.45) = 7.37$, $p = .03$. Quality moderators influenced the overall proportion of unexplained variance in both the anger and happy categories with 5% and 4% of the variance explained, respectively.

Discussion

Our meta-analysis untangles some of the conceptual confusions that have plagued the science of emotion for over a century. Since the time scholars began pondering the nature of emotion, they have been arguing about emotions' physical nature, and whether variation within an emotion category and similarity across categories is meaningful or error. Consistent with previous meta-analytic reviews, the most robust finding in our analysis was the observation of substantial variation in ANS responding during instances of the same emotion category and for some emotion categories, our sensitivity analyses revealed the possibility that variation was reduced by publication bias. Unlike prior meta-analyses, however, we have offered a theoretical framework that predicts, a priori, variation within a category and similarities in autonomic physiological responses across categories. When the variation in a sample distribution for one emotion category is very large, the central tendency is no longer a good estimate of any one value in the distribution. In other words, variation in effect sizes around the distribution average of effect sizes is meaningful. The emotion populations view hypothesizes that the observed variation *within* emotion categories is expected because it is *intrinsic* to the nature of emotion, and therefore is not reducible to method variance and to processes that are epiphenomenal to emotion. Furthermore, when the statistical moments of multiple distributions (i.e., the central tendency and variance) are very similar, they do not provide specific information about one distribution or another. In other words, the similarity in ANS effect sizes across emotion categories is meaningful. From an emotion populations view we hypothesize that the observed similarity in ANS responding across emotion categories is expected because changes in heart rate, blood pressure, respiration, and so forth must be made meaningful as emotions during the categorization that results from the brain's normal process of predicting sensory inputs and preparing for

Table 10
Confusion Matrix from Support Vector Machine Pattern Classification Analysis

		True Category							
		Anger	Disgust	Fear	Happy	Sad	Neutral	Total No. Predictions	**Correct/Total** Incorrect/Total
Predicted Category	Anger	**46** **19.1%**	3 1.2%	11 4.6%	11 4.6%	9 3.7%	9 3.7%	89	**46/89 (51.7%)** 43/89 (48.3%)
	Disgust	6 2.5%	**5** **2.1%**	7 2.9%	7 2.9%	7 2.9%	6 2.5%	38	**5/38 (13.2%)** 33/38 (86.8%)
	Fear	3 1.2%	3 1.2%	**4** **1.7%**	2 0.8%	3 1.2%	1 0.4%	16	**4/16 (25.0%)** 12/16 (75.0%)
	Happy	4 1.7%	1 0.4%	3 0.8%	**5** **2.1%**	2 2.1%	5 1.2%	20	**5/20 (25.0%)** 15/20 (75.0%)
	Sad	6 2.5%	2 0.8%	8 3.3%	6 2.5%	**14** **5.8%**	6 2.5%	42	**14/42 (33.3%)** 28/42 (66.6%)
	Neutral	10 4.1%	4 1.7%	4 1.7%	8 3.3%	8 3.3%	**2** **0.8%**	36	**2/36 (5.6%)** 34/36 (94.4%)
	Total No. of Cases	75	18	37	39	43	29	**241**	
	Correct/Total True Cases	**46/75 (61.3%)**	**5/18 (27.8%)**	**4/37 (10.8%)**	**5/39 (12.8%)**	**14/43 (32.6%)**	**2/29 (6.9%)**		**76/241 (31.5%)**
	Incorrect/Total True Cases	29/75 (38.7%)	13/18 (72.2%)	33/37 (89.2%)	34/39 (87.2%)	29/43 (67.4%)	27/29 (93.1%)		**165/241 (68.5%)**

Note. Classification of emotion categories using modified support vector machine (SVM) analysis. To avoid potential dependencies in the data, when a study included effect sizes from the same participants across more than one emotion category (e.g., study manipulated fear, anger, and disgust within the same individual) we randomly selected *only one* of the effect sizes to include (resulting in 241 individual cases). The confusion matrix for the five-way classification was based on a series of binary classification analyses (e.g., anger vs fear, sadness vs. disgust, and so on). The first number in each box represents the number of times that classification was made correctly (e.g., anger was predicted correctly 46 times). The percentage in each box represents the number of times the classification occurred divided by all of the cases (e.g. 46, 241 = 19.1%). The true emotion category is represented along the x-axis. The predicted category is represented along the y-axis. Diagonal entries reflect classification accuracy (e.g., the classifier predicted that fear was being induced and in fact, it was). Off-diagonal entries represent incorrect classifications. The boxes along the bottom of the table along the x-axis represent the sensitivity of the classifiers (Correct, Total True Cases) and the false negative rate (Incorrect, Total True Cases). The boxes along the right side of the table along the y-axis represent the precision of the classifiers (i.e., what is the tendency of the classifier to assign any study to a particular emotion category; Correct, Total Predictions) and the false discovery rate (Incorrect, Total Predictions).

action (Barrett, 2006b, 2012, 2017a, 2017c; Barrett & Satpute, 2013; Lindquist & Barrett, 2012).

Evidence for Emotion Populations View: Context Versus Other Sources of Variation

Our ability to provide a strong direct test of the population hypothesis—that metabolic and situational demands produce variation and degeneracy in ANS changes within a single emotion category—is limited by the published studies that were available for meta-analysis. Our meta-analytic results clearly show that ANS variation is not epiphenomenal to the nature of emotion, and therefore is broadly consistent with the population hypothesis. A specific test of this hypothesis awaits studies that manipulate or measure more fine-grained features of the context like situational or metabolic demands. Meta-analyses are useful for suggesting course corrections to the published literature, and we hope that our analyses will be useful for future research on emotion in this regard.

Furthermore, our interpretation of the variation we observed is strengthened by considering our findings in the context of the broader published literature. Our findings, for example, are consistent with meta-analyses of task-related brain activation data, all of which uniformly support the population hypothesis (Kober et al., 2008;

Table 11
Regression Models With Sample, Study, and Quality Moderators for Each Emotion Category

Moderators	Anger			Fear			Happy			Sad			Neutral		
	F	*df*	*p*	*F*	*df*	*p*	*F*	*df*	*p*	*F*	*df*	*p*	*F*	*df*	*p*
Sample moderators															
ANS measure type	2.68	13, 205	<.001	4.59	14, 55.1	<.001	.51	9, 83	.87	1.56	11, 64	.13†	1.37	9, 75	.22
Sample type	5.16	1, 52.4	.03*	11.57	1, 21.3	.003**	1.91	1, 21.4	.18	.52	1, 27.5	.48	.01	1, 38.8	.94
Gender	2.35	2, 129	.09†	.19	2, 16.5	.83	2.1	2, 24.1	.15	.17	2, 17.5	.84	.06	2, 64.9	.94
ANS Measure Type × Gender	1.98	16, 156	.02*										2.85	10, 75	.004**
ANS Measure Type × Sample				3.93	6, 59	.003**				2.6	5, 64	.03*			
Study moderators															
ANS measure type	2.59	13, 218	.002	3.16	14, 62.4	<.001	.47	9, 74	.89	.71	11, 31.9	.72	.61	9, 80	.78
Induction	.71	4, 57.1	.59	1.78	5, 12.1	.19	.34	4, 24.2	.85	.48	4, 23.8	.75	.25	5, 30.3	.94
Evaluation	.76	1, 73.7	.39	1.02	1, 17.7	.33	.51	1, 26.3	.48	.01	1, 43.7	.92	3.72	1, 31.5	.063
Other people	.73	1, 62.3	.40	1.24	1, 21.5	.28	.04	1, 22	.84	.35	1, 19.5	.56	.02	1, 28.1	.88
Video	.49	1, 76.1	.48	1.69	1, 15.1	.21	.27	1, 18.2	.61	.02	1, 20.6	.90	2.48	1, 19.4	.13†
Emotion words	.06	1, 60.6	.80	.06	1, 20.1	.81	.28	1, 18.1	.60	.15	1, 23.1	.70	.01	2, 20.2	.99
Quality moderators															
ANS measure type	3.01	13, 169	<.001	2.19	11, 44.6	.033	.94	9, 51	.50	.63	11, 47	.79	.83	9, 53	.59
Baseline duration	.28	3, 39	.84	.54	3, 5.45	.67	.74	3, 11.7	.55	.6	3, 6.36	.64	2.13	3, 13.1	.15†
No. emotions manip	.4	5, 39.8	.85	.22	5, 4.97	.94	.92	5, 8.9	.51	.27	5, 6.22	.92	.47	4, 10.1	.76
Induction duration	.62	3, 37.9	.61	.66	3, 4.2	.62	6.57	3, 11	.008**	.03	3, 6.44	.99	.33	3, 11.8	.80
Manipulation check	.56	1, 47	.46	.03	1, 5.03	.87	2.72	1, 20.1	.11†	.87	1, 9.47	.38	7.37	1, 8.45	.03*
Rigor induction	.81	2, 44.4	.45	.28	2, 8.63	.76	.73	2, 6.79	.52	.09	2, 5.72	.92	.75	2, 13	.49
Rigor physio record	2.10	2, 30.6	.04*	.59	2, 5.2	.59	.16	2, 9.5	.85	.68	2, 6.66	.54	.45	2, 19.7	.65

Note. Moderator data for disgust is not available because the dataset for disgust was too sparse, and thus, the models would not converge. The sample analysis contained three moderators, the study analysis contained six moderators, and the initial quality analysis contained seven moderators. The variable ANS Measure Type is a categorical variable and was included in every model to account for differences in effect sizes across different autonomic measures. Significant interactions between ANS measure type and moderator variables explicitly examine whether the effect of the moderator differs across different ANS measures. Only significant interactions are reported.
† $p < .15$. * $p < .05$. ** $p < .01$.

Lindquist et al., 2012; Wager et al., 2015); see also Hamman's (2012) discussion of Vytal and Hamann, (2010). Neural fingerprints have not been identified for any emotion category in studies of single neurons (Guillory & Bujarski, 2014), activity in individual brain regions (Lindquist et al., 2012), in the intrinsic connectivity of brain networks (Touroutoglou, Lindquist, Dickerson, & Barrett, 2015), in the dynamic connectivity between networks during intense emotional experiences during tasks (e.g., Raz et al., 2016) or in multivoxel patterns (Clark-Polner, Johnson, & Barrett, 2016; Wager et al., 2015). Studies that claim to find neural fingerprints (Kragel, Knodt, Hariri, & LaBar, 2016; Kragel & LaBar, 2015; Saarimäki et al., 2016) mistakenly interpret the patterns they find as brain states rather than as abstract statistical summaries (Clark-Polner, Wager, Satpute, & Barrett, 2017) and ignore the fact that the details of their patterns do not replicate across studies. But more importantly, these studies, along with many others, provide direct evidence for the theory of constructed emotion's hypothesis that the brain is making sense of variable ANS changes to construct emotion (e.g., Barrett, 2017a, 2017c; Oosterwijk, Touroutoglou, & Lindquist, 2015).

Moreover, behavioral experiments clearly show that the variation within emotion categories is meaningfully tied to context and situational factors and is not merely due to variability in the experimental method. A growing number of studies of emotion are designed to explicitly model and capture heterogeneity within emotion categories both within individuals and across cultures (e.g., Ceulemans, Kuppens, & Van Mechelen, 2012; Gendron, Roberson, van der Vyver, & Barrett, 2014a, 2014b; Hortensius, Schutter, & Harmon-Jones, 2012; Kuppens, Van Mechelen, & Rijmen, 2008; Kuppens et al., 2007; Nezlek, Vansteelandt, Van Mechelen, & Kuppens, 2008; Stemmler, Aue, & Wacker, 2007; Wilson-Mendenhall, Barrett, & Barsalou, 2013, 2015; Wilson-Mendenhall, Barrett, Simmons, & Barsalou, 2011). Even studies of "fear" learning in rodents find evidence of variability in ANS responses and neural circuitry (e.g., Gross & Canteras, 2012; LeDoux, Iwata, Cicchetti, & Reis, 1988). Indeed, the human brain learns emotion categories in a highly situated way (Barrett, 2006b, 2017a, 2017c). Thus, variation is inherent to the nature of emotion, not a moderator that is independent of or incidental to an instance of emotion (Barrett, 2009). Unfortunately, meta-analyses are ill-suited to characterizing intraindividual variability over time due to learning history because most studies of emotion do not include this information and even if they did, self-reports of emotional learning history reflect people's beliefs about themselves rather than their actual learning histories (Robinson & Clore, 2002). The only real way to map intraindividual variability is using idiographic methods associated with experience-sampling, an excellent example of which can be seen in the work of Kuppens and colleagues (e.g., Kuppens, Van Mechelen, & Rijmen, 2008; Kuppens et al., 2007; Nezlek, Vansteelandt, Van Mechelen, & Kuppens, 2008). Using idiographic methodologies in combination with machine learning algorithms, scientists might be able to identify emotion categories *within* individuals using ANS features, which would be an exciting avenue for future study.

Population Thinking Versus Essentialism

Our meta-analysis is not only a test of two emotion theories. It is also a test of the utility of essentialist thinking in the science of emotion. The kind of population thinking that is employed by the theory of constructed emotion (Barrett, 2013; Barrett, 2017a, 2017c) was on display in Darwin's *On the Origin of Species* (Darwin, 1859/2001) and inspired generations of biologists to vanquish essentialism from biology (Mayr, 2004). *Essentialism* is the belief that a category of instances all share an underlying property that conveys category membership, causing them to share observable features, and explain their cohesiveness even when they lack observable similarities (Gelman, 2003).[24] The fingerprint hypothesis is, however, a hypothesis that relies on essentialism. *Psychological essentialism* (Medin & Ortony, 1989) occurs when the essence cannot be identified or when the essence is hypothetical or metaphorical, and is clearly on display in the classical view of emotion. For example, Ekman's hypothetical affect program (e.g., Ekman & Cordaro, 2011), Panksepp's hypothetical FEAR system (e.g., Panksepp & Watt, 2011), and Adolph's "central fear state" or "functional fear state" (Anderson & Adolphs, 2014) are all examples of psychological essentialism.

Support for population thinking (and against essentialism) is particularly strong in our meta-analysis when you consider that the effect size variation in the published experiments on emotion, as summarized in our meta-analyses, comes from individual studies were not designed to evoke variable emotion experiences. Instead, their goal was to use standardized methods to induce stereotypic examples of emotional episodes in the lab (typically on the assumption that each emotion category has an essence). Still, this degree of variability should not be surprising. William James (1884, 1994) wrote about this variation over a century ago explicitly rejecting the idea that a single set of bodily changes could be diagnostic of a given emotion category across instances or individuals. He wrote "Surely there is no definite affection of 'anger' in an 'entitative' sense"[25] (p. 206). If we look around in everyday life, variability in emotion is vividly on display. A person might tremble in fear, jump in fear, scream, or even laugh in fear. In James's words "Fear of getting wet is not the same as fear of a bear" (p. 206).

A Case for Strong Inference

This meta-analysis was guided by our efforts to take a strong inference approach to the science of emotion. In the philosophy of science, strong inference requires testing hypotheses against one another, rather than merely testing either one against the null hypothesis, thereby protecting against confirmation bias. We tested the classical view of emotion against our theory of constructed emotion. In this theory (formerly the conceptual act theory; Barrett, 2006b, 2012, 2013, 2015a, 2017a, 2017c; Barrett, Wilson-Mendenhall, & Barsalou, 2015), an emotion category is a

[24] The biobehavioral fingerprint is considered to be an emotion's innate, universal essence, or the result of the essence. People don't even have to know what the essence is to believe that there is one. This phenomenon, called psychological essentialism (Medin & Ortony, 1989), describes hypothesized classical view constructs, like affect programs (Ekman, 1984; Tomkins, 1962, 1963) and emotion modules (Tooby & Cosmides, 1990). Essentialism is not necessarily a bad thing; some consider it a useful strategy for scientific inquiry because they believe that it mirrors the structure of the real world (i.e., they believe the world is full of natural kind categories; (Bloom, 2000; Kornblith, 1993; Pinker, 1997). Others, however, believe that essentialism is a particularly poor strategy for science to take (e.g., Lewontin, 2000).

[25] In this context, "entitative" is synonymous with essentialist.

heterogeneous collection of instances. To construct a given instance, the brain uses highly situated concept knowledge (past experiences each with some prior probability) to predict and categorize incoming sensory inputs from the world and the body in a highly situated way, preparing a person for situated action. This means that a collection of emotion instances, such as instances of gratitude, fear, or pride, can be biologically and behaviorally unique, but that the brain can treat them as equivalent in some situations (this is what it means to categorize—to treat instances as equivalent for some purpose; Murphy, 2002). In these moments of conceptualized emotion, we are ignoring differences across instances. But we should not base our scientific theories on this intuitive sense that instances *truly* are the same and that variation *truly* is meaningless, forgetting the possibility that grouping might be a function of perception rather than a function of nature. We can use the history of biology as inspiration to resist the lure of essentialism (Barrett, 2013, 2017a).

As scientists, there are several concepts that we can use to build a science using populating thinking. One is degeneracy (Edelman & Gally, 2001; Marder & Taylor, 2011; Tononi, Sporns, & Edelman, 1999). Different proteins can catalyze the same reaction of enzymes (Edelman & Gally, 2001; Tononi, Sporns, & Edelman, 1999), different antibodies can bind to the same antigen (Edelman, 1974), different genotypes can produce the same phenotype (Edelman & Gally, 2001; Tononi, Sporns, & Edelman, 1999), different neurons can give rise to the same intrinsic network (Marder & Taylor, 2011; Tononi, Edelman, & Sporns, 1998; Tononi, Sporns, & Edelman, 1999), and different patterns of network interaction can give rise to the same behavior (Price & Friston, 2002). Degeneracy is strongly related to the capacity of a system to carry information, to be robust to damage, and to be able to evolve (Tononi, Sporns, & Edelman, 1999; Whitacre & Bender, 2010). It is not surprising then that different ANS patterns can give rise to instances of the same emotion category, potentially even within the same study or situation. What we need is an epistemological strategy that allows us to engage in a many (ANS patterns) to many (emotion) mapping. The existence of degeneracy in emotion categories, and potentially even in instances that occur in the same situation, suggests that our traditional statistical methods may be masking important and meaningful variance. The science of emotion treats variance within an emotion category as "error" (in the literal sense, or in the sense that variation across instances of an emotion category is a threat to validity because it reflects some moderating influence that is outside the emotion itself); but degeneracy suggests this variation is "signal"—it's a feature of emotional life, not a bug in our methods. Variability is not simply noise in the data resulting from measurement error, but instead is important because it enhances the capacity of a system to carry information (Marder & Taylor, 2011; Tononi, Edelman, & Sporns, 1998). We must learn to think about concepts of variation, noise, and information differently.

Amid all this variation, we did not find much evidence for consistent or specific patterns of ANS change within each emotion category in either our multilevel univariate models or our multivariate pattern classification analyses, and even where we found some evidence, it is not clear what it indicates in the face of so much variability. In population thinking, the norm or stereotype of the category is an abstract representation and does not necessarily represent the individual members of the category (Mayr, 2004), in the same way that the mean of a distribution does not represent the individual elements of that distribution when there is a large standard deviation. This reasoning means that identifying a classifier for an emotion category with MPCA is necessary, but not sufficient, to provide strong support for the classical approach to emotion. The classifier will never reveal a set of necessary and sufficient features that appear in every, or even in most, instances of the category (i.e., it is not a fingerprint or an essence for the category). To provide strong evidence for the classical view, scientists at a minimum must also show that a highly similar classification pattern replicates from context to context, person to person, and study to study. They must show that the physical elements that make up the pattern are present in instances of the category (rather than being an abstract statistical summary). However, the failure to find significant classifiers is disconfirmatory evidence for the classical view.

Studies have shown that the best example of a category (i.e., its prototype) is not an instance with typical features or the most common features but instead, is an idealized instance of the category for a particular context. The prototype can even be an abstract instance with a combination of features that need not even exist as an instance in nature (Barsalou, 1993, 2005, 2012; Posner & Keele, 1968). In a similar way, the pattern classifier for a category is a statistical summary of the category, but it need not be representative of each and every one of its instances (Barrett, 2017a; Clark-Polner et al., 2016). Consistent with the emotion population hypothesis, successful pattern classification does not require that the pattern be present in every (or even any) individual instance (i.e., the pattern is not an essence; it is an abstract summary).

Prior Meta-Analyses

In one regard, our results did not replicate the prior meta-analyses by Stemmler (2004) and Cacioppo et al. (2000), both of which found some mean autonomic differences between anger and fear categories, albeit amid considerable variation. We found no consistent or specific differences between fear and anger categories in either our multilevel or multivariate analyses. Data from inductions of anger and fear comprised more than half of the effect sizes in our dataset (397 out of 739), so if there were consistent or specific differences in ANS changes from baseline between anger and fear, our analysis should have been sufficiently powered to observe them. Kreibig (2010), Levenson (1992, 2011), and to a lesser extent Stemmler (2004), proposed that key distinctions between fear and anger may lie in having more studies that measure additional variables such as respiration (e.g., lung tidal volume or the ratio of inspiratory to expiratory time; Kreibig, 2010), cardiovascular variables like total peripheral resistance (TPR; Kreibig, 2010; Levenson, 1992; Stemmler, 2004), or gastric variables like the gastric myoelectrical activity that supports peristalsis (Levenson, 2011, 2014). Furthermore, the effects of sympathetic and parasympathetic nervous system activation differ as they innervate different muscles in the body. Researchers neither measure (nor analyze) the autonomic nervous system with this level of granularity and, instead, rely on coarser, more easily acquired measures of the ANS. These hypotheses cannot be ruled out entirely because we did not have sufficient data available to test many of these possibilities. Nonetheless, we doubt that the addition of these variables to future studies will provide the key

evidence for anger and fear fingerprints given other considerations already discussed.

Recently, in an effort to address emerging evidence of variability, there has been an upsurge in the discussion of emotion category "sub-types." For example, Kreibig (2010) and then Harrison, Kreibig, and colleagues (Harrison, Kreibig, & Critchley, 2013) proposed two types of sadness (in deactivated sadness, heart rate decreases; in activated sadness, it increases) and disgust (in core disgust, heart rate goes up; in body-boundary violation, heart rate goes down). Scarantino (2015) and Griffiths (1997) similarly have attempted to parse emotion categories into the "basic" subtypes versus other subtypes. This is how the classical approach takes context into account—by creating finer-grained typologies. However, a typology, no matter how precise, is still a different approach from population thinking.

Limitations

The limitations in the published literature on emotion prevented us from testing other explanations for the effect size variations. For example, we were unable to test whether there were impediments to the expression of emotion (e.g., the unanticipated use of display rules or emotion regulation that led to emotional suppression). Unfortunately, very few of the articles included in this analysis either measured or examined impediments to expressing emotion. We assessed this to the best of our ability by testing whether the presence of manipulation checks moderated the effect sizes, but this was an imperfect approximation, at best.

More importantly, there are currently no objective criteria or "observer-independent" means by which to determine whether or not an emotion has occurred, and therefore it is not possible to determine whether "real" emotions are being elicited in the laboratory. Research shows that various objective, measurable changes that occur during emotion inductions (e.g., facial movements, ANS responses, vocalizations, etc.) do not routinely correlate with one another (i.e., they fail to show coherence; for recent reviews see Barrett, 2006a, 2006b; Mauss & Robinson, 2009; Quigley & Barrett, 2014). This means that there is no objective, gold standard criterion by which we can verify a successful emotion induction (Barrett, 2006a). In most of the studies in our database, manipulation checks took the form of participant self-report which is, in fact, considered the only empirically verified method for reliably indicating whether the target emotion was induced or not (Barrett, 2006b). Few studies in our database measured facial movements, so we did not use the presence of facial action measurement as a way of operationalizing a manipulation check. Moreover, it is not clear that facial actions are a gold standard for assessing a successful manipulation of emotion since there is little scientific evidence that people specifically and consistently scowl when angry, pout when sad, and so on (see Russell, Bachorowski, & Fernández-Dols, 2003).

There also are no scientific criteria that can objectively adjudicate whether some instances of emotion are more "real" than others. Scientists have debated for decades over what counts, and what does not count, as an emotion (see Ortony & Turner, 1990). What is a "real" instance of emotion is a philosophical matter of how to define the scope of inquiry in the first place. To some scientists, some instances of emotion are just not close enough to what they imagine as the prototype of the phenomenon to "count" as an emotion. For them, variation is viewed as a critical barrier to the careful study of ANS fingerprints. For others, however, variation is the point. Regardless, these boundaries cannot be empirically discovered. They are always stipulated, and therefore always subject to dispute (Barrett, 2017b, 2017c).

Our analysis also revealed several areas of sampling bias in the literature. First, we found that studies have not sampled broadly across ANS measures, and that this sampling has changed over time (see Figure 4). More recently accessible ANS variables such as measures of cardiac function that can be derived via impedance cardiography or more detailed measures of respiratory function (e.g., the ratio of inspiratory to expiratory time) have received insufficient attention in the published literature and there were too few effect sizes for most of these measures to be meaningfully interpreted in our meta-analyses. By comparison, many more effect sizes exist for measures like heart rate that have complex ANS underpinnings (e.g., the same change in heart rate can be due to different patterns parasympathetic and sympathetic activation and withdrawal; Berntson, Cacioppo, & Quigley, 1991), which makes it difficult to interpret such measures across studies.

Second, we found that emotions, ANS measures, and induction types have been sampled inconsistently across time (Figures 3 and 4). This is likely due, at least in part, to advances in measurement and physiological recording technology over time. However, these inconsistencies hampered our ability to assess the nature of some of the variability in effect sizes. It would have been interesting, for example, to directly test whether the goals of the emotion induction task predicted ANS responses over and above emotion category designation. This finding would be consistent with both Obrist's cardio-somatic coupling idea (Obrist, Webb, Sutterer, & Howard, 1970) and the emotion population hypothesis, both of which would predict that emotion induction tasks that require more active engagement (e.g., social evaluation) would produce greater variability in ANS responses than more passive tasks (e.g., listening to music). Unfortunately, we could not analyze the data to address this issue due to inconsistent use of induction tasks across emotion categories (e.g., studies of the anger category often used active tasks whereas studies of the sad category often used passive tasks). However, because there was substantial variation in both emotion categories (anger and sad), we can at least conclude that variation is not due solely to either the active or passive task type.

Third, we found there was a lack of granularity in assessing the self-reported experience of individuals in these studies. It would have been interesting, for example, to assess variations in the intensity of feeling angry (or sad or happy, etc.) and the relation of that variation in experiential intensity to ANS reactivity. Unfortunately, in the extant literature, there was not sufficient detail in the reporting of experience to conduct this type of analysis. We tried to account for these outcomes in our moderator analyses, but were unable to operationalize them in a nuanced way. This, combined with the sparseness of the data available, made it difficult to conduct a complete investigation of autonomic consistency and specificity, and also precluded a thorough analysis of the influence of context. Clearly this is an important future research direction.

Fourth, there is substantial variability in the research methodology used across studies within the scientific literature. We ex-

plored the impact of methodological differences using the six "quality" moderators (duration of the baseline, number of emotions manipulated, duration of the emotion induction procedure, presence of a manipulation check, rigor of the induction procedure, and rigor of the physiological recording methods) but we acknowledge that this is a limited sample of methodological differences. For example, there was tremendous variation in the time bin size that researchers used for ANS data. Several of the ANS measures identified in this analysis have predictable response latency, and then a less predictable habituation, adaptation, and recovery slope. Whether these are acknowledged or accounted for varies tremendously by research group. We explored the impact of methodological differences using the "quality" moderators we developed, but these are just estimates and are not a substitute for carefully designed experimental procedures. One important future direction for researchers interested in emotion theory is to design large-scale, comprehensive studies that directly compare emotion theories against one another (e.g., testing the "weak" version of the classical view of emotion against the "strong version") ideally using the same time bins, the same induction methods, identical dependent measures, and a within-subjects design).

The sampling biases in the literature meant that a large portion of our data came from "anger" studies (76 out of 204) with a preponderance of certain ANS measures (i.e., SBP and DBP). In our MPCA analysis, we attempted to address the latter sampling bias by removing ANS measures that were most often sampled in anger studies (e.g., blood pressure variables) and rerunning the classifier. However, removing these measures did not have an impact on the pattern of results nor did it increase the classification accuracy for other emotion categories. This strategy was nonoptimal, but it was what we could accomplish with the extant data and we cannot rule out the possibility that our results would have been different if we had had a less biased literature to draw from.

Final Thoughts and Future Directions

It is our view that the science of emotion would be better served by shifting our goals away from the search for ANS fingerprints. We propose that a more productive strategy would be to map the existing heterogeneity, instead of designing experiments to capture only a stereotype of each emotion category. Further, naturalistic observation is an important part of the scientific paradigm in biology (Mayr, 1988) but has largely been underutilized in psychology because it is expensive, impractical, and computationally difficult. But new technologies make such observations possible (see Barrett, 2012, 2014; Kleckner & Quigley, 2015). It is now possible to utilize the technological and methodological advances afforded by mobile sensing technologies, virtual reality and gaming technologies, and both machine learning algorithms and more sophisticated modeling of social and cultural contexts to begin to design and analyze more naturalistically measured physiological changes during emotional episodes. One very real possibility is that we might discover idiographic regularities for each emotion category (i.e., a given person might have a repertoire of repeatable instances of *anger*, or *fear*, or any other emotion category each of which may have a pattern of ANS change). Discoveries of this sort, however, are only possible if we stop viewing heterogeneity in emotion as epiphenomenal and start recognizing its import as an avenue of empirical inquiry. Moreover, if population-level fingerprints for emotion categories exist, then we will have the means for discovering them using these types of context- and idiographically sensitive investigations.

Another important direction for future investigation is extending this work to test the fingerprints and populations hypotheses in studies of emotion across cultures and languages. Including research in languages other than English was beyond the scope of our meta-analysis, but culture and language are integral to the study of emotion. Research shows that both critically shape the perception and experience of emotion (Barrett, 2017a; Gendron, Lindquist, Barsalou, & Barrett, 2012; Gendron, Roberson, van der Vyver, & Barrett, 2014a). ANS reactivity during instances of emotion across cultures deserves its own treatment and we highly encourage researchers to take up this challenge.

Perhaps the most important observation from our meta-analysis is that there is room for methodological improvement in the science of emotion. For example, Levenson has written convincingly about the many methodological obstacles that serve as sources of error in experiments (e.g., unverified induction methods, inducing emotions of low intensity, narrow assessments of ANS functioning, incomplete characterization of the activity of ANS responses, and poor temporal matching of ANS measurements to the duration and timing of emotion occurrence). Notably, the quality of the experimental methods did not have a robust influence across emotion categories, although it did impact some of the effect sizes. Of the six quality moderators included in the moderator model for each emotion category, one was significant in anger (rigor of physiological recording), one in happy (length of induction), and one in neutral (presence of manipulation checks). Nor did differences in experimental methodology consistently explain the variability. For example, the emotion induction procedure influenced effect sizes only for the neutral category (and marginally for fear). This suggests that methodological features, while important, do not appear to account for much of the variability in ANS reactivity within emotion categories. Our meta-analytic findings highlight the critical importance of deliberate and sustained scientific efforts to observe, map, and better understand the breadth, nature, and function of this variation in emotion categories, both within and across categories. That is at the heart of the kind of population thinking that Darwin introduced in *On the Origin of Species*.

References

Allen, M. T., Boquet, A. J., Jr., & Shelley, K. S. (1991). Cluster analyses of cardiovascular responsivity to three laboratory stressors. *Psychosomatic Medicine, 53,* 272–288. http://dx.doi.org/10.1097/00006842-199105000-00002

Allwein, E. L., Schapire, R. E., & Singer, Y. (2001). Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of Machine Learning Research, 1,* 113–141.

Anderson, D. J., & Adolphs, R. (2014). A framework for studying emotions across species. *Cell, 157,* 187–200. http://dx.doi.org/10.1016/j.cell.2014.03.003

Averill, J. R. (1980). A constructivist view of emotion. *Emotion: Theory, Research, and Experience, 1,* 305–339.

Barrett, L. F. (2006a). 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. (2006b). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10,* 20–46. http://dx.doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F. (2009). Variety is the spice of life: A psychological construction approach to understanding variability in emotion. *Cognition and Emotion, 23,* 1284–1306. http://dx.doi.org/10.1080/02699930902985894

Barrett, L. F. (2011a). Was Darwin wrong about emotional expressions? *Current Directions in Psychological Science, 20,* 400–406.

Barrett, L. F. (2011b). Bridging token identity theory and supervenience theory through psychological construction. *Psychological Inquiry, 22,* 115–127. PMC3142092

Barrett, L. F. (2012). Emotions are real. *Emotion, 12,* 413–429. http://dx.doi.org/10.1037/a0027555

Barrett, L. F. (2013). Psychological construction: The Darwinian approach to the science of emotion. *Emotion Review, 5,* 379–389. http://dx.doi.org/10.1177/1754073913489753

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

Barrett, L. F. (2015). Ten common misconceptions about the psychological construction of emotion. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 45–79). New York, NY: Guilford Press.

Barrett, L. F. (2017a). *How emotions are made*. New York, NY: Houghton Mifflin Harcourt.

Barrett, L. F. (2017b). Functionalism cannot save the classical view of emotion. *Social Cognitive and Affective Neuroscience, 12,* 34–36.

Barrett, L. F. (2017c). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*. Advance online publication. http://dx.doi.org/10.1093/scan/nsx060

Barrett, L. F., & Bliss-Moreau, E. (2009). Affect as a psychological primitive. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 41, pp. 167–218). Burlington, VT: Academic Press.

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology, 58,* 373–403.

Barrett, L. F., & Russell, J. A. (2015). An introduction to psychological construction. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 1–17). New York, NY: Guilford Press.

Barrett, L. F., & Satpute, A. B. (2013). Large-scale brain networks in affective and social neuroscience: Toward an integrative functional architecture of the brain. *Current Opinion in Neurobiology, 23,* 361–372. http://dx.doi.org/10.1016/j.conb.2012.12.012

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience, 16,* 419–429. http://dx.doi.org/10.1038/nrn3950

Barrett, L. F., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). The conceptual act theory: A road map. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 83–110). New York, NY: Guilford Press.

Barsalou, L. W. (1993). Challenging assumptions about concepts. *Cognitive Development, 8,* 169–180. http://dx.doi.org/10.1016/0885-2014(93)90012-T

Barsalou, L. W. (2005). Situated conceptualization. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 619–650). St. Louis, MO: Elsevier. http://dx.doi.org/10.1016/B978-008044612-7/50083-4

Barsalou, L. W. (2012). The human conceptual system. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 239–258). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139029377.017

Baucom, L. B., Wedell, D. H., Wang, J., Blitzer, D. N., & Shinkareva, S. V. (2012). Decoding the neural representation of affective states. *NeuroImage, 59,* 718–727. http://dx.doi.org/10.1016/j.neuroimage.2011.07.037

Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1991). Autonomic determinism: The modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. *Psychological Review, 98,* 459–487. http://dx.doi.org/10.1037/0033-295X.98.4.459

Bloom, P. (2000). *How children learn the meanings of words: Learning, development and conceptual change*. Cambridge, MA: MIT Press.

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

Cacioppo, J. T., Berntson, G. G., Klein, D. J., & Poehlmann, K. M. (1997). Psychophysiology of emotion across the life span. In K. W. Schaie & M. P. Lawton (Eds.), *Annual review of gerontology and geriatrics* (Vol. 17, pp. 27–74). New York, NY: Springer.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (Vol. 2, pp. 173–191). New York, NY: Guilford Press.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology, 76,* 839–855. http://dx.doi.org/10.1037/0022-3514.76.5.839

Cai, J.-F., Candès, E. J., & Shen, Z. (2010). A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization, 20,* 1956–1982. http://dx.doi.org/10.1137/080738970

Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis? *Journal of the Association for Computing Machinery, 58,* 11. http://dx.doi.org/10.1145/1970392.1970395

Candès, E. J., & Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics, 9,* 717–772. http://dx.doi.org/10.1007/s10208-009-9045-5

Candès, E. J., & Tao, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory, 56,* 2053–2080. http://dx.doi.org/10.1109/TIT.2010.2044061

Ceulemans, E., Kuppens, P., & Van Mechelen, I. (2012). Capturing the structure of distinct types of individual differences in the situation-specific experience of emotions: The case of anger. *European Journal of Personality, 26,* 484–495. http://dx.doi.org/10.1002/per.847

Chanes, L., & Barrett, L. F. (2016). Redefining the role of limbic areas in cortical processing. *Trends in Cognitive Sciences, 20,* 96–106. http://dx.doi.org/10.1016/j.tics.2015.11.005

Clark-Polner, E., Johnson, T. D., & Barrett, L. F. (2016). Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cerebral Cortex, 27,* 1944–1948. http://dx.doi.org/10.1093/cercor/bhw028

Clark-Polner, E., Wager, T. D., Satpute, A. B., & Barrett, L. F. (2017). Neural fingerprinting: Meta-analysis, variation, and the search for brain-based essences in the science of emotion. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *The handbook of emotion* (4th ed.). New York, NY: Guilford.

Clore, G. L., & Ortony, A. (2008). Appraisal theories: How cognition shapes affect into emotion. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 628–642). New York, NY: Guilford Press.

Coan, J. A. (2010). Emergent ghosts of the emotion machine. *Emotion Review, 2,* 274–285.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics, 10,* 417–451. http://dx.doi.org/10.2307/3001616

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. San Diego, CA: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1,* 98–101. http://dx.doi.org/10.1111/1467-8721.ep10768783

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20,* 273–297. http://dx.doi.org/10.1007/BF00994018

Cunningham, W. A., Dunfield, K. A., & Stillman, P. E. (2013). Emotional states from affective dynamics. *Emotion Review, 5,* 344–355. http://dx.doi.org/10.1177/1754073913489749

Darwin, C. (1859/2001). *On the origin of species*. Cambridge, MA: Harvard University Press.

Darwin, C. (1872/2005). *The expression of the emotions in man and animals*: Boston, MA: Digireads.com Publishing. http://dx.doi.org/10.1037/10001-000

Davidson, R. J. (1992). Anterior cerebral asymmetry and the nature of emotion. *Brain and Cognition, 20,* 125–151. http://dx.doi.org/10.1016/0278-2626(92)90065-T

De Leersnyder, J., Boiger, M., & Mesquita, B. (2013). Cultural regulation of emotion: Individual, relational, and structural sources. *Frontiers in Psychology, 4,* 55. http://dx.doi.org/10.3389/fpsyg.2013.00055

Duffy, E. (1957). The psychological significance of the concept of arousal or activation. *Psychological Review, 64,* 265–275. http://dx.doi.org/10.1037/h0048837

Duval, S. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 127–144). Chichester, UK: Wiley.

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

Edelman, G. M. (1974). *Cellular selection and regulation in the immune response*. New York, NY: Raven.

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 13763–13768. http://dx.doi.org/10.1073/pnas.231499798

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629–634. http://dx.doi.org/10.1136/bmj.315.7109.629

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska symposium on motivation* (Vol. 19, pp. 207–283). Lincoln, Nebraska: University of Nebraska Press.

Ekman, P. (1984). Expression and the nature of emotion. In K. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 319–344). Hillsdale, NJ: Erlbaum.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*(3–4), 169–200.

Ekman, P. (2007). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life* (2nd ed.). New York, NY: Henry Holt and Company.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review, 3,* 364–370. http://dx.doi.org/10.1177/1754073911410740

Ekman, P., & Davidson, R. (1994). *The nature of emotion: Fundamental questions*. New York, NY: Oxford University Press.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science, 221,* 1208–1210. http://dx.doi.org/10.1126/science.6612338

Floyd Allport. (1924). *Social psychology*. Boston, MA: Houghton Mifflin Company.

Fridlund, A. J. (1992). Darwin's anti-darwinism and the expression of the emotions in man and animals. In K. T. Strongman (Ed.), *International review of emotion* (Vol. 2, pp. 117–137). New York, NY: Wiley.

Friedman, B. H. (2010). Feelings and the body: The Jamesian perspective on autonomic specificity of emotion. *Biological Psychology, 84,* 383–393. http://dx.doi.org/10.1016/j.biopsycho.2009.10.006

Frijda, N. H. (1986). *The emotions*. New York, NY: Cambridge University Press.

Gavaghan, D. J., Moore, R. A., & McQuay, H. J. (2000). An evaluation of homogeneity tests in meta-analyses in pain using simulations of individual patient data. *Pain, 85,* 415–424. http://dx.doi.org/10.1016/S0304-3959(99)00302-4

Gee, H. (2014). *The accidental species: Misunderstandings of human evolution*. Chicago, IL: University of Chicago Press.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195154061.001.0001

Gelman, S. A., & Rhodes, M. (2012). Two-thousand years of stasis: How psychological essentialism impedes evolutionary understanding. In K. R. Rosengren, S. Brem, E. M. Evans, & G. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 3–21). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199730421.003.0001

Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion, 12,* 314–325. http://dx.doi.org/10.1037/a0026007

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014a). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science, 25,* 911–920. http://dx.doi.org/10.1177/0956797613517239

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014b). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion, 14,* 251–262. http://dx.doi.org/10.1037/a0036052

Griffiths, P. E. (1997). *What emotions really are: The problem of psychological categories*. Chicago, IL: University of Chicago Press. http://dx.doi.org/10.7208/chicago/9780226308760.001.0001

Gross, J. J., & Barrett, L. F. (2011). Emotion generation and emotion regulation: One or two depends on your point of view. *Emotion Review, 3,* 8–16.

Gross, C. T., & Canteras, N. S. (2012). The many paths to fear. *Nature Reviews Neuroscience, 13,* 651–658. http://dx.doi.org/10.1038/nrn3301

Guillory, S. A., & Bujarski, K. A. (2014). Exploring emotions using invasive methods: Review of 60 years of human intracranial electrophysiology. *Social Cognitive and Affective Neuroscience, 9,* 1880–1889. http://dx.doi.org/10.1093/scan/nsu002

Halko, N., Martinsson, P.-G., & Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review, 53,* 217–288. http://dx.doi.org/10.1137/090771806

Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: Controversies and consensus. *Trends in Cognitive Sciences, 16,* 458–466. http://dx.doi.org/10.1016/j.tics.2012.07.006

Harrison, N. A., Kreibig, S. D., & Critchley, H. D. (2013). A two-way road: Efferent and afferent pathways of autonomic activity in emotion. In J. Armony & P. Vuilleumier (Eds.), *The Cambridge handbook of human affective neuroscience* (pp. 82–106). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511843716.006

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin, 92,* 490–499. http://dx.doi.org/10.1037/0033-2909.92.2.490

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

Hedges, L. V., & Vevea, J. L. (1996). Estimating effect size under publication bias: Small sample properties and robustness of a random effects

selection model. *Journal of Educational and Behavioral Statistics, 21,* 299–332. http://dx.doi.org/10.3102/10769986021004299

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539–1558. http://dx.doi.org/10.1002/sim.1186

Hortensius, R., Schutter, D. J. L. G., & Harmon-Jones, E. (2012). When anger leads to aggression: Induction of relative left frontal cortical activity with transcranial DC stimulation increases the anger-aggression relationship. *Social Cognitive and Affective Neuroscience, 7,* 342–347. http://dx.doi.org/10.1093/scan/nsr012

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). *The relation between short-term emotion dynamics and psychological well-being: A meta-analysis*. Washington, DC: American Psychological Association.

Hunt, W. A. (1941). Recent developments in the field of emotion. *Psychological Bulletin, 38,* 249–276. http://dx.doi.org/10.1037/h0054615

Izard, C. E. (1977). *Human emotions*. New York, NY: Plenum Press. http://dx.doi.org/10.1007/978-1-4899-2209-0

James, W. (1884). What is an emotion? *Mind, 9,* 188–205. http://dx.doi.org/10.1093/mind/os-IX.34.188

James, W. (1994). The physical bases of emotion. *Psychological Review, 101,* 205–210. http://dx.doi.org/10.1037/0033-295X.101.2.205 (Original work published 1894)

Kleckner, I. R., & Quigley, K. S. (2015). An approach to mapping the neurophysiological state of the body to affective experience. In L. F. Barrett & J. Russell (Eds.), *The psychological construction of emotion* (pp. 265–301). New York, NY: Guilford Press.

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical-subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *NeuroImage, 42,* 998–1031. http://dx.doi.org/10.1016/j.neuroimage.2008.03.059

Kolodyazhniy, V., Kreibig, S. D., Gross, J. J., Roth, W. T., & Wilhelm, F. H. (2011). An affective computing approach to physiological emotion specificity: Toward subject-independent and stimulus-independent classification of film-induced emotions. *Psychophysiology, 48,* 908–922. http://dx.doi.org/10.1111/j.1469-8986.2010.01170.x

Kornblith, H. (1993). *Inductive inference and its natural ground: An essay in naturalistic epistemology*. Cambridge, MA: The MIT Press.

Kragel, P. A., Knodt, A. R., Hariri, A. R., & LaBar, K. S. (2016). Decoding spontaneous emotional states in the human brain. *PLoS Biology, 14,* e2000106. http://dx.doi.org/10.1371/journal.pbio.2000106

Kragel, P. A., & Labar, K. S. (2013). Multivariate pattern classification reveals autonomic and experiential representations of discrete emotions. *Emotion, 13,* 681–690. http://dx.doi.org/10.1037/a0031820

Kragel, P. A., & LaBar, K. S. (2015). Multivariate neural biomarkers of emotional states are categorically distinct. *Social Cognitive and Affective Neuroscience, 10,* 1437–1448. http://dx.doi.org/10.1093/scan/nsv032

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84,* 394–421. http://dx.doi.org/10.1016/j.biopsycho.2010.03.010

Kreibig, S. D., Wilhelm, F. H., Roth, W. T., & Gross, J. J. (2007). Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films. *Psychophysiology, 44,* 787–806. http://dx.doi.org/10.1111/j.1469-8986.2007.00550.x

Kuppens, P., Van Mechelen, I., & Rijmen, F. (2008). Toward disentangling sources of individual differences in appraisal and anger. *Journal of Personality, 76,* 969–1000. http://dx.doi.org/10.1111/j.1467-6494.2008.00511.x

Kuppens, P., Van Mechelen, I., Smits, D. J. M., De Boeck, P., & Ceulemans, E. (2007). Individual differences in patterns of appraisal and anger experience. *Cognition and Emotion, 21,* 689–713. http://dx.doi.org/10.1080/02699930600859219

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97,* 377–395. http://dx.doi.org/10.1037/0033-295X.97.3.377

Lang, P. J., & Davis, M. (2006). Emotion, motivation, and the brain: Reflex foundations in animal and human research. *Progress in Brain Research, 156,* 3–29. http://dx.doi.org/10.1016/S0079-6123(06)56001-7

Lange, C. G. (1922). *The emotions* (Vol. 1). Baltimore, MD: Williams & Wilkins.

Larsen, J. T., Berntson, G. G., Poehlmann, K. M., Ito, T. A., & Cacioppo, J. T. (2008). *The psychophysiology of emotion handbook of emotions* (Vol. 3, pp. 180–195). New York, NY: Guilford.

LeDoux, J. E., Iwata, J., Cicchetti, P., & Reis, D. J. (1988). Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear. *The Journal of Neuroscience, 8,* 2517–2529.

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

Levenson, R. W. (1992). Autonomic nervous system differences among emotions. *Psychological Science, 3,* 23–27. http://dx.doi.org/10.1111/j.1467-9280.1992.tb00251.x

Levenson, R. W. (2011). Basic emotion questions. *Emotion Review, 3,* 379–386. http://dx.doi.org/10.1177/1754073911410743

Levenson, R. W. (2014). The autonomic nervous system and emotion. *Emotion Review, 6,* 100–112. http://dx.doi.org/10.1177/1754073913512003

Lewontin, R. C. (2000). *The triple helix: Gene, organism, and environment*. Cambridge, MA: Harvard University Press.

Light, R. J., Pillemer, & D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

Lin, Z., Chen, M., & Ma, Y. (2010). The augmented LaGrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprintarXiv:1009.5055.*

Lindquist, K. A., & Barrett, L. F. (2012). A functional architecture of the human brain: Emerging insights from the science of emotion. *Trends in Cognitive Sciences, 16,* 533–540. http://dx.doi.org/10.1016/j.tics.2012.09.005

Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2016). The brain basis of positive and negative affect: Evidence from a meta-analysis of the human neuroimaging literature. *Cerebral Cortex, 26,* 1910–1922. http://dx.doi.org/10.1093/cercor/bhv001

Lindquist, K. A., Siegel, E. H., Quigley, K. S., & Barrett, L. F. (2013). The hundred-year emotion war: Are emotions natural kinds or psychological constructions? Comment on Lench, Flores, and Bench (2011). *Psychological Bulletin, 139,* 255–263. http://dx.doi.org/10.1037/a0029038

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences, 35,* 121–143.

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

Lipsitz, L. A. (2004). Physiological complexity, aging, and the path to frailty. *Science of Aging Knowledge Environment, 2004,* 16. http://dx.doi.org/10.1126/sageke.2004.16.pe16

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS for mixed models*. Cary, NC: SAS Institute, Inc.

Mandler, G. (1984). *Mind and body: Psychology of emotion and stress*. New York, NY: Norton.

Marder, E., & Taylor, A. L. (2011). Multiple models to capture the variability in biological neurons and networks. *Nature Neuroscience, 14,* 133–138. http://dx.doi.org/10.1038/nn.2735

Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motivation and Emotion, 14,* 195–214. http://dx.doi.org/10.1007/BF00995569

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

Mayr, E. (1988). The why and how of species. *Biology & Philosophy, 3,* 431–441. http://dx.doi.org/10.1007/BF00647963

Mayr, E. (2004). *What makes biology unique?: Considerations on the autonomy of a scientific discipline*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511617188

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science, 11,* 730–749. http://dx.doi.org/10.1177/1745691616662243

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511529863.009

Mendes, W. B. (2010). Weakened links between mind and body in older age: The case for maturational dualism in the experience of emotion. *Emotion Review, 2,* 240–244. http://dx.doi.org/10.1177/1754073910364149

Mosteller, F., & Tukey, J. W. (1968). Data analysis, including statistics. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (2nd ed., pp. 80–203). Reading, UK: Addison Wesley.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT press.

Nezlek, J. B., Vansteelandt, K., Van Mechelen, I., & Kuppens, P. (2008). Appraisal-emotion relationships in daily life. *Emotion, 8,* 145–150. http://dx.doi.org/10.1037/1528-3542.8.1.145

Nyklíček, I., Thayer, J. F., & Van Doornen, L. J. (1997). Cardiorespiratory differentiation of musically induced emotions. *Journal of Psychophysiology, 11,* 304–321.

Obrist, P. A. (1981). *Cardiovascular psychophysiology*. New York, NY: Plenum Press. http://dx.doi.org/10.1007/978-1-4684-8491-5

Obrist, P. A., Webb, R. A., Sutterer, J. R., & Howard, J. L. (1970). The cardiac-somatic relationship: Some reformulations. *Psychophysiology, 6,* 569–587. http://dx.doi.org/10.1111/j.1469-8986.1970.tb02246.x

Oosterwijk, S., Touroutoglou, A., & Lindquist, K. A. (2015). The neuroscience of construction: What neuroimaging approaches can tell us about how the brain creates the mind. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion*. New York, NY: Guilford Press.

Ortony, A., & Clore, G. (2015). Can an appraisal model be compatible with psychological constructionism? In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 305–333). New York, NY: Guilford Press.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review, 97,* 315–331. http://dx.doi.org/10.1037/0033-295X.97.3.315

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York, NY: Oxford University Press.

Panksepp, J., & Watt, D. (2011). What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emotion Review, 3,* 387–396. http://dx.doi.org/10.1177/1754073911410741

Park, B.-J., Jang, E.-H., Chung, M.-A., & Kim, S.-H. (2013). Design of prototype-based emotion recognizer using physiological signals. *ETRI Journal, 35,* 869–879. http://dx.doi.org/10.4218/etrij.13.0112.0751

Petrantonakis, P. C., & Hadjileontiadis, L. J. (2010). Emotion recognition from EEG using higher order crossings. *IEEE Transactions on Information Technology in Biomedicine, 14,* 186–197. http://dx.doi.org/10.1109/TITB.2009.2034649

Pfeifer, M. A., Weinberg, C. R., Cook, D., Best, J. D., Reenan, A., & Halter, J. B. (1983). Differential changes of autonomic nervous system function with age in man. *The American Journal of Medicine, 75,* 249–258. http://dx.doi.org/10.1016/0002-9343(83)91201-9

Pinker, S. (1997). *How the mind works*. New York, NY: Norton.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363. http://dx.doi.org/10.1037/h0025953

Price, C. J., & Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends in Cognitive Sciences, 6,* 416–421. http://dx.doi.org/10.1016/S1364-6613(02)01976-9

Quigley, K. S., & Barrett, L. F. (2014). Is there consistency and specificity of autonomic changes during emotional episodes? Guidance from the conceptual act theory and psychophysiology. *Biological Psychology, 98,* 82–94. http://dx.doi.org/10.1016/j.biopsycho.2013.12.013

Rainville, P., Bechara, A., Naqvi, N., & Damasio, A. R. (2006). Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology, 61,* 5–18. http://dx.doi.org/10.1016/j.ijpsycho.2005.10.024

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raz, G., Touroutoglou, A., Wilson-Mendenhall, C., Gilam, G., Lin, T., Gonen, T., . . . Barrett, L. F. (2016). Functional connectivity dynamics during film viewing reveal common networks for different emotional experiences. *Cognitive, Affective & Behavioral Neuroscience, 16,* 709–723. http://dx.doi.org/10.3758/s13415-016-0425-4

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

Roseman, I. J. (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition and Emotion, 10,* 241–278. http://dx.doi.org/10.1080/026999396380240

Roseman, I. J. (2001). A model of appraisal in the emotion system: Integrating theory, research, and applications. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 249–267). Oxford, UK: Oxford University Press.

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

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86,* 638–641. http://dx.doi.org/10.1037/0033-2909.86.3.638

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (2nd ed.). Newbury Park, CA: Sage. http://dx.doi.org/10.4135/9781412984997

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York, NY: Russell Sage Foundation.

Rosenthal, R., & Rubin, D. B. (1982). Comparing effect sizes of independent studies. *Psychological Bulletin, 92,* 500–504. http://dx.doi.org/10.1037/0033-2909.92.2.500

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. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110,* 145–172. http://dx.doi.org/10.1037/0033-295X.110.1.145

Russell, J. A., Bachorowski, J.-A., & Fernández-Dols, J.-M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54,* 329–349. http://dx.doi.org/10.1146/annurev.psych.54.101601.145102

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology, 76,* 805–819. http://dx.doi.org/10.1037/0022-3514.76.5.805

Rychlowska, M., Miyamoto, Y., Matsumoto, D., Hess, U., Gilboa-Schechtman, E., Kamble, S., . . . Niedenthal, P. M. (2015). Heterogeneity of long-history migration explains cultural differences in reports of emotional expressivity and the functions of smiles. *Proceedings of the National Academy of Sciences of the United States of America, 112*(19), E2429–E2436. http://dx.doi.org/10.1073/pnas.1413661112

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., . . . Nummenmaa, L. (2016). Discrete neural signatures of basic emotions. *Cerebral Cortex, 26,* 2563–2573. http://dx.doi.org/10.1093/cercor/bhv086

Scarantino, A. (2015). Basic emotions, psychological construction and the problem of variability. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 334–376). New York, NY: Guilford Press.

Scarantino, A., & Griffiths, P. (2011). Don't give up on basic emotions. *Emotion Review, 3,* 444–454.

Schachter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review, 69,* 379–399. http://dx.doi.org/10.1037/h0046234

Sinha, R. (1996). Multivariate response patterning of fear and anger. *Cognition and Emotion, 10,* 173–198. http://dx.doi.org/10.1080/026999396380321

Sitaram, R., Lee, S., Ruiz, S., Rana, M., Veit, R., & Birbaumer, N. (2011). Real-time support vector classification and feedback of multiple emotional brain states. *NeuroImage, 56,* 753–765. http://dx.doi.org/10.1016/j.neuroimage.2010.08.007

Stemmler, G. (2004). Physiological processes during emotion. In P. Philippot & R. S. Feldman (Eds.), *The regulation of emotion* (pp. 33–70). Mahwah, NJ: Erlbaum & Associates.

Stemmler, G., Aue, T., & Wacker, J. (2007). Anger and fear: Separable effects of emotion and motivational direction on somatovisceral responses. *International Journal of Psychophysiology, 66,* 141–153. http://dx.doi.org/10.1016/j.ijpsycho.2007.03.019

Stephens, C. L., Christie, I. C., & Friedman, B. H. (2010). Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis. *Biological Psychology, 84,* 463–473. http://dx.doi.org/10.1016/j.biopsycho.2010.03.014

Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior, 106,* 5–15. http://dx.doi.org/10.1016/j.physbeh.2011.06.004

Straub, R. H., Miller, L. E., Schölmerich, J., & Zietz, B. (2000). Cytokines and hormones as possible links between endocrinosenescence and immunosenescence. *Journal of Neuroimmunology, 109,* 10–15. http://dx.doi.org/10.1016/S0165-5728(00)00296-4

Takahashi, K. (2004, December 13–15). *Remarks on svm-based emotion recognition from multimodal bio-potential signals*. Paper presented at the 13th IEEE International Workshop on Robot and Human Interactive Communication, Palmerston North, New Zealand.

Tassinary, L. G., Cacioppo, J. T., & Vanman, E. J. (2007). The skeletomotor system: Surface. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 267–302). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511546396.012

Thayer, R. E. (1989). *The biopsychology of mood and arousal*. New York, NY: Oxford University Press.

Tomkins, S. (1962). *Affect imagery consciousness: The positive affects*. New York, NY: Springer Publishing.

Tomkins, S. (1963). *Affect imagery consciousness: The negative affects*. New York, NY: Springer Publishing.

Tononi, G., Edelman, G. M., & Sporns, O. (1998). Complexity and coherency: Integrating information in the brain. *Trends in Cognitive Sciences, 2,* 474–484. http://dx.doi.org/10.1016/S1364-6613(98)01259-5

Tononi, G., Sporns, O., & Edelman, G. M. (1999). Measures of degeneracy and redundancy in biological networks. *Proceedings of the National Academy of Sciences of the United States of America, 96,* 3257–3262. http://dx.doi.org/10.1073/pnas.96.6.3257

Tooby, J., & Cosmides, L. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology, 11,* 375–424. http://dx.doi.org/10.1016/0162-3095(90)90017-Z

Touroutoglou, A., Lindquist, K. A., Dickerson, B. C., & Barrett, L. F. (2015). Intrinsic connectivity in the human brain does not reveal networks for 'basic' emotions. *Social Cognitive and Affective Neuroscience, 10,* 1257–1265. http://dx.doi.org/10.1093/scan/nsv013

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

Turner, J. R., & Carroll, D. (1985). Heart rate and oxygen consumption during mental arithmetic, a video game, and graded exercise: Further evidence of metabolically exaggerated cardiac adjustments? *Psychophysiology, 22,* 261–267. http://dx.doi.org/10.1111/j.1469-8986.1985.tb01597.x

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods, 45,* 576–594. http://dx.doi.org/10.3758/s13428-012-0261-6

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods, 47,* 1274–1294. http://dx.doi.org/10.3758/s13428-014-0527-2

Van den Noortgate, W., & Onghena, P. (2003). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, & Computers, 35,* 1–10. http://dx.doi.org/10.3758/BF03195492

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika, 60,* 419–435.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods, 10,* 428–443. http://dx.doi.org/10.1037/1082-989X.10.4.428

Voorspoels, W., Vanpaemel, W., & Storms, G. (2011). A formal ideal-based account of typicality. *Psychonomic Bulletin & Review, 18,* 1006–1014. http://dx.doi.org/10.3758/s13423-011-0122-9

Vrana, S. R. (1993). The psychophysiology of disgust: Differentiating negative emotional contexts with facial EMG. *Psychophysiology, 30,* 279–286. http://dx.doi.org/10.1111/j.1469-8986.1993.tb03354.x

Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: A voxel-based meta-analysis. *Journal of Cognitive Neuroscience, 22,* 2864–2885. http://dx.doi.org/10.1162/jocn.2009.21366

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., & Barrett, L. F. (2015). A Bayesian model of category-specific emotional brain responses. *PLoS Computational Biology, 11,* e1004066. http://dx.doi.org/10.1371/journal.pcbi.1004066

Wang, Z., Lai, M.-J., Lu, Z., Fan, W., Davulcu, H., & Ye, J. (2014, June). *Rank-one matrix pursuit for matrix completion*. Paper presented at the International Conference on Machine Learning.

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

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76,* 820–838. http://dx.doi.org/10.1037/0022-3514.76.5.820

Whitacre, J. M. (2010). Degeneracy: A link between evolvability, robustness and complexity in biological systems. *Theoretical Biology & Medical Modeling, 7,* 6. http://dx.doi.org/10.1186/1742-4682-7-6

Whitacre, J., & Bender, A. (2010). Degeneracy: A design principle for achieving robustness and evolvability. *Journal of Theoretical Biology, 263,* 143–153. http://dx.doi.org/10.1016/j.jtbi.2009.11.008

Whitacre, J. M., Rohlfshagen, P., Bender, A., & Yao, X. (2010). The role of degenerate robustness in the evolvability of multiagent systems in dynamic environments. In R. Schaefer, C. Cotta, J. Kołlodziej & G. Rudolph (Eds.), *Parallel problem solving from nature* (pp. 284–293). Berlin, Germany: Springer.

Whitton, A. E., Henry, J. D., Rendell, P. G., & Grisham, J. R. (2014). Disgust, but not anger provocation, enhances levator labii superioris activity during exposure to moral transgressions. *Biological Psychology, 96,* 48–56. http://dx.doi.org/10.1016/j.biopsycho.2013.11.012

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2013). Neural evidence that human emotions share core affective properties. *Psychological Science, 24,* 947–956. http://dx.doi.org/10.1177/0956797612464242

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2015). Variety in emotional life: Within-category typicality of emotional experiences is associated with neural activity in large-scale brain networks. *Social Cognitive and Affective Neuroscience, 10,* 62–71. http://dx.doi.org/10.1093/scan/nsu037

Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., & Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia, 49,* 1105–1127. http://dx.doi.org/10.1016/j.neuropsychologia.2010.12.032

Wright, J., Ganesh, A., Rao, S., Peng, Y., & Ma, Y. (2009, December). *Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization*. Paper presented at the Advances in neural information processing systems.

Received May 22, 2015
Revision received September 6, 2017
Accepted September 13, 2017 ■

Correction to Verhage et al. (2016)

In the article "Narrowing the Transmission Gap: A Synthesis of Three Decades of Research on Intergenerational Transmission of Attachment," by Marije L. Verhage, Carlo Schuengel, Sheri Madigan, R. M. Pasco Fearon, Mirjam Oosterman, Rosalinda Cassibba, Marian J. Bakermans-Kranenburg, and Marinus H. van IJzendoorn (*Psychological Bulletin,* 2016, Vol. 142, No. 4, pp. 337–366. http://dx.doi.org/10.1037/bul0000038), there are errors in Table 7. The percentages of the attachment classifications do not add up to 100%.

Please find the corrected version of Table 7 below.

Table 7
Cross Tabulation of Four-Way Classifications of Caregiver Attachment Representations and Child-Caregiver Attachment

	Child attachment					
Adult attachment	Secure	Avoidant	Resistant	Disorganized	Total	%
Autonomous	**912 (17.1)**	122 (−7.6)	88 (−5.5)	196 (−9.9)	1,318	47.5
Dismissing	235 (−9.0)	**192 (12.5)**	68 (0.6)	146 (−0.3)	641	23.1
Preoccupied	77 (−6.5)	39 (0.7)	**66 (9.5)**	58 (0.4)	240	8.7
Unresolved	223 (−7.2)	53 (−4.1)	55 (−0.4)	**244 (12.3)**	575	20.7
Total	1,447	406	277	644	2,774	
%	52.2	14.6	10.0	23.2		

Note. Predicted transmission patterns are in bold font. Adjusted standardized residuals within brackets.

http://dx.doi.org/10.1037/bul0000149