The psychophysiology of mixed emotional states: Internal and external replicability analysis of a direct replication study

Kreibig, Sylvia Dagmar; Samson, Andrea Christiane; Gross, James J.

How to cite

KREIBIG, Sylvia Dagmar, SAMSON, Andrea Christiane, GROSS, James J. The psychophysiology of mixed emotional states: Internal and external replicability analysis of a direct replication study. In: Psychophysiology, 2015, vol. 52, n° 7, p. 873–886. doi: 10.1111/psyp.12425

This publication URL: https://archive-ouverte.unige.ch//unige:97930

Publication DOI: 10.1111/psyp.12425

Psychophysiology, 52 (2015), 873–886. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/psyp.12425

The psychophysiology of mixed emotional states: Internal and external replicability analysis of a direct replication study

SYLVIA D. KREIBIG, ANDREA C. SAMSON, AND JAMES J. GROSS
Department of Psychology, Stanford University, Stanford, California, USA

Abstract

The replicability of emotion-related physiological changes constitutes a fundamental issue in affective science. We undertook a direct replication of the physiological differentiation of amusement, disgust, and a mixed emotional state as previously reported (Kreibig, Samson, & Gross, 2013). In the current study, 48 women watched 54 amusing, disgusting, and mixed emotional film clips while cardiovascular, electrodermal, and respiratory measures were obtained. Primary analyses indicated physiological differentiation of the mixed emotional state from amusement and disgust. We evaluated (a) the probability that future replications of the current study would yield similar results using bootstrapped confidence intervals of effect sizes, and (b) the stability of results of physiological reactivity between actual replications using correlation and regression analyses. Findings suggest replicable differentiation of amusement, disgust, and a mixed emotional state.

Descriptors: Direct replication, Mixed emotional states, Amusement, Disgust, Replicability, Cardiovascular, Electrodermal, Respiratory

One of the most difficult questions in statistical inference concerns how a statistic, calculated from a random sample of units, varies from sample to sample.

(Wasserman & Bockenholt, 1989, p. 208)

It is widely agreed that emotions are multicomponential responses that consist of coordinated changes in subjective feeling, motor expression, and physiology (Mauss, Levenson, McCater, Wilhelm, & Gross, 2005). But just how to conceptualize mixed emotional states remains an open question. Subjectively, mixed emotional states are characterized by the co-occurrence of two or more differing emotional feelings (Larsen & McGraw, 2011). How mixed emotional states are characterized in terms of the physiological response remains unclear.

We recently reported on a study that tested different theoretical accounts of mixed emotions (Kreibig, Samson, & Gross, 2013): A first account is based on the valence–arousal model (Russell, 1980), which, among other things, posits that positive and negative feelings are mutually exclusive, just like feelings of hot and cold (Schimmack, 2001). The small percentage of concurrent reports of positive and negative feelings is ascribed to measurement error. This *nondifferentiation account* predicts that responses associated with reported mixed emotional feelings do not differ from those of one of the pure constituent emotions. A second account is based on the appraisal tendency framework (Lerner & Keltner, 2000), which holds that appraisal tendencies from prior emotions influence subsequent appraisals, with incompatible appraisal tendencies cancelling each other out and compatible ones enhancing each other (cf. Pe & Kuppens, 2012). This *additive account* predicts intensity differences of the mixed emotional state from one of its pure constituent emotions. A third account is based on the component process model of emotion (Scherer, 1984), which views mixed emotions as combining appraisal outcomes typical for several different pure emotions to form an emotion that is qualitatively different from its pure constituent emotions. This *emergence account* predicts pattern differences of the mixed emotional state from both of its pure constituent emotions.

Based on an experiment in which 43 women watched film clips that elicited amusement, disgust, and an amusing–disgusting mixed emotional state, our previous study (Kreibig et al., 2013) was the first to report (a) univariate physiological differentiation of an amusing–disgusting mixed emotional state from its pure constituent emotions, and (b) intensity and pattern differences of the physiological response profile of the mixed emotional state from those of pure amusement and pure disgust. These findings suggest a distinct physiological response of the mixed emotional state, which is consistent with the emergence account of mixed emotions and inconsistent with both the nondifferentiation and additive accounts.

We thank research assistant Brittney Reyes for help with the data collection, Jeff Rector for help with scripting the stimulus presentation, and members of the Stanford Psychophysiology Laboratory for comments on drafts of this manuscript. This research was supported by the Swiss National Science Foundation (PBGEP1-125914, PA00P1_139593, PBFRP1-127896, PA00P1_136380).

Address correspondence to: Sylvia Kreibig, Department of Psychology, Stanford University, 450 Serra Mall, Bldg 420, Stanford, CA 94305, USA. E-mail: skreibig@stanford.edu

PSYCHOPHYSIOLOGY

Given their novelty and broad implications for emotion theory, these findings require replication. Direct replication involves repetition of an experimental procedure, whereas conceptual replication involves repetition of a test of a hypothesis with different methods (Schmidt, 2009). Because conceptual replication introduces variation of methods, failure to replicate is ambiguous with respect to whether the effect does not replicate or not generalize. We therefore focus here on direct replication to determine to what degree the psychophysiological differentiation of amusing, disgusting, and a mixed emotional state is replicable.

Because conventional statistical significance tests do not evaluate result replicability (Cumming, 2008; Cumming & Maillardet, 2006), it is useful to employ internal and external replicability analysis. Internal replicability addresses the probability that future replications of the current study would yield similar results. The bootstrap approach (Efron, 1979; Wasserman & Bockholt, 1989) represents a particularly powerful internal replication method: It generates a very large number of additional samples from the current sample and recalculates relevant parameters, for example, effect sizes, for each sample, based on which confidence intervals (CIs) are constructed to express the precision and uncertainty associated with such point estimates.

External replicability analysis uses an independent sample to determine whether actual replications yielded similar results (Thompson, 1996, 2006). External replications have traditionally been confined to presenting mirroring results from previous studies, with statistical significance of the effect of interest as the criterion for success (Cumming, 2014) and nonsignificant Experiment × Manipulation interactions to suggest no difference in response to the manipulation between experiments. However, not only can a null hypothesis of "no experimental differences" not be proven, but rejection of the null hypothesis also strongly depends on sample size (i.e., power)—interaction tests even more so than main effect tests (Wahlsten, 1991). Correlation and regression analyses of effect sizes, in contrast, quantify similarity of research results based on independent replication samples and thus allow statements about the degree of replicability of results between studies (Stemmler, Aue, & Wacker, 2007).

The current study examined internal and external replicability of the physiological differentiation of amusing, disgusting, and a mixed emotional state based on data from a direct replication and our previous study (Kreibig et al., 2013). As in our previous study, we elicited amusement, disgust, and a mixed emotional state by presenting film clips while assessing cardiovascular, electrodermal, and respiratory measures. While core features of the current study remained the same (laboratory, experimenter, emotional stimuli, physiological measures), it differed from our previous study in terms of data collection at a later time with different participants and within the larger context of a reappraisal task, allowing us to gauge the robustness of our initial findings in the face of these variations.

The aim of the current study was to evaluate result replicability, for which we employed a three-step process: We first calculated the same hypothesis tests as in our previous study (Kreibig et al., 2013) on the current data set to test competing predictions derived from nondifferentiation, additive, and emergence accounts of the psychophysiology of mixed emotions. Outcomes of these analyses not only describe the structure of results in this new sample, but also serve as input to internal and external replicability analyses. Second, for analysis of internal replicability, we constructed bootstrapped confidence intervals of effect sizes based on the current sample. This allowed us to evaluate the probability that future replications would yield similar results. Third, for analysis of external replicability, we computed correlation and regression analyses of effect sizes of physiological reactivity between the current and previous study. This allowed us to evaluate whether actual replications yielded similar results.

Method

Participants

Fifty-five women participated in a 120-min experiment for course credit or pay ($25). Seven participants were excluded (one terminated the study early, one was pregnant, two were noncompliant, three had computer problems). Of the remaining 48 participants (mean age 20.7 years, $SD = 3.7$), three were African American, five Asian American, 13 Caucasian, four Hispanic, 10 multiracial, and eight declined to state. Demographic data from five participants were missing because of experimenter error. The experiment was approved by the Institutional Review Board of Stanford University.

Materials

Stimuli were 54 amateur video clips, each 20–30 s long, that presented humorous lapses to induce pure amusement, painful accidents to induce pure disgust, and ambiguous bloopers to induce mixed amusement and disgust (18 film clips for each category). Only responses to the six film clips of each film category that were randomly paired with the natural–attend instructions for each subject are considered here (see Procedure for details). Stimuli had previously been validated (Kreibig et al., 2013; Samson, Kreibig, Soderstrom, Wade, & Gross, in press; see Table 1 for results of subjective feelings and facial expressions by film type in the current sample).

Apparatus

Stimuli were presented with a personal computer using Presentation software (Neurobehavioral Systems Inc., Albany, CA). They were displayed on a 19-inch computer monitor at a viewing distance of 55 cm under low ambient light. Responses were entered via keyboard. Participant compliance was monitored through a hidden camera. Physiological data were recorded and amplified with a multichannel BioNex 8-slot chassis (Mindware Technologies, Grahanna, OH) equipped with an impedance cardiograph and skin conductance amplifier (Model 50-371100-00), a 4-channel biopotential amplifier (Model 50-371102-00), a 4-channel transducer amplifier (Model 50-371106-00), and a 4-channel high-level interface module (Model 50-371103-00). Data were sampled at 1000 Hz, 16-bit digitized, and transmitted to a computer for viewing and storage using BioLab 2.4 (Mindware).

Procedure

Data collection took place individually. After obtaining informed consent, participants were seated in front of a computer screen and physiological sensors attached. Participants performed a verbal fluency task (not reported here). They then read definitions of the rating scales and instructions for the reappraisal task that was to be performed during the film viewing. They practiced the task with two example film clips for each film and reappraisal category prior to the film viewing. Next, participants completed a respiratory volume calibration using fixed volume bags (Morel, Forster, & Suter,

Table 1. *Subjective Feelings and Facial Expressions by Film Type in the Current Sample*

	Amusing films		Mixed films		Disgusting films		rm ANOVA			Post hoc HSD tests			Effect sizes		
	M	*SD*	*M*	*SD*	*M*	*SD*	*F*(2,94)	ε	η^2	A vs. D	M vs. A	M vs. D	$d_{A\text{-}D}$	$d_{M\text{-}A}$	$d_{M\text{-}D}$
Self-report															
Amusement	4.30	1.51	3.13	1.59	1.76	1.21	145.01	***	0.94	0.832	***	***	***	2.20	1.23	1.39
Disgust	1.48	0.98	2.53	1.56	4.04	1.57	169.98	***	0.86	0.838	***	***	***	−2.25	−1.29	−1.64
Mixed feelings	1.38	0.81	1.81	1.02	1.58	1.00	14.92	***	0.99	0.368	***	*	*	−0.37	−0.75	0.43
Valence	4.77	1.08	3.97	1.07	2.76	1.20	177.79	***	0.79	0.833	***	***	***	2.20	1.31	1.81
Arousal	3.19	1.62	2.87	1.49	2.94	1.73	2.15		0.71	0.136				0.17	0.36	−0.08
Facial expressions (%)																
CS	96.8	38.8	117.2	42.9	154.2	64.8	84.89	***	0.75	0.750	***	***	***	−1.63	−1.01	−1.06
ZM	179.3	116.5	110.2	41.2	106.1	35.0	59.78	***	0.58	0.582	***	***	–	1.17	1.12	0.20

Note. Electromyographic activity was quantified as percent change from the immediately preceding baseline. rm ANOVA = repeated measures analysis of variance; post hoc HSD tests = Tukey honestly significant difference post hoc tests; d = Cohen's d effect size; CS = corrugator supercilii; ZM = zygomaticus major.
*$p < .05$. ***$p < .001$.

1983). The experimenter then left the room, and participants started the film viewing. Film clips of each film category were randomly partitioned into three thirds and paired with instructions to either focus on the positive elements of the clip, focus on its negative elements, or attend naturally to it (only natural–attend conditions are considered here). Film clip–instruction pairs were presented in random sequences under the condition that there be no more than two repetitions of the same film and regulation type. Before each film, a 4-s slide instructed participants to "focus on the funny, amusing, or positive elements of the clip," "focus on the repulsive, painful, or negative elements of the clip," or "attend naturally" to it, after which participants started the film clip with a button press. Each film clip–instruction pair was preceded by a 20–30 s rest period (varying independently of subsequent film clip length), during which participants were instructed to sit quietly, clear their mind, and avoid moving or speaking. Participants rated their current emotional feelings immediately after each film clip. Upon completing the film viewing, participants performed a second respiratory volume calibration and a paced breathing task for vagal assessment at 8, 10.5, 13, and 18 cycles per minute (Ritz, Thöns, & Dahme, 2001). Participants then were unhooked from physiological recording equipment, completed demographic and personality questionnaires (not reported here), and were debriefed. Procedures were identical to our previous study (Kreibig et al., 2013) except for the addition of the verbal fluency and reappraisal tasks.

Measures

Subjective feelings. After each film clip, participants rated on a 6-point Likert scale their current feelings regarding amusement, disgust, valence, arousal, compassion, and pain, as well as effectiveness and difficulty of following instructions (the last four items are not considered here). Scale definitions were identical to our previous study (Kreibig et al., 2013).

Facial expressions. Surface electromyography (EMG, in μV) over zygomaticus major and corrugator supercilii on the left side of the face was recorded with 4-mm miniature Beckman Ag/AgCl electrode pairs filled with Teca electrode gel (Oxford Instruments, Hawthorne, NY). Skin sites were cleaned and abraded to achieve interelectrode impedance of less than 10 kW. Signal conditioning included 500 Hz antialiasing hardware, 60 Hz notch, and 20–500 Hz digital band-pass filters, rectification, and smoothing using a 10-ms running average.

Physiology. The identical array of 15 physiological measures reported in our previous study (Kreibig et al., 2013) was obtained according to the same procedures.

The electrocardiogram (ECG) was recorded using Ag/AgCl spot electrodes (TraceRite LT430S, Forth-Rite Technologies, Austin, TX) positioned in a three-lead unipolar modified chest configuration. The signal was amplified and band-pass filtered at 10–40 Hz. Heart rate (HR, in beats per minute) was calculated from interbeat intervals between automatically detected R spikes (Kreibig et al., 2013). Respiratory sinus arrhythmia (uncorrected RSA_{uc}, in ms) was scored using the peak–valley method (Eckberg, 1983). RSA was corrected for within-individual effects of respiration rate and tidal volume (RSA_c, in ms/l) based on measurements from the paced-breathing task (Schulz, Ayala, Dahme, & Ritz, 2009).

Impedance cardiography (ICG) was recorded using a four-spot electrode configuration over neck and thorax with Ag/AgCl electrodes (TraceRite LT430S). The ICG was ensemble-averaged over

20–30 s task intervals in synchrony with the ECG R wave. Characteristic points were identified to calculate pre-ejection period (PEP, in ms), defined as the interval from the ECG R peak onset to the ICG B point (Lozano et al., 2007); left ventricular ejection time (LVET, in ms), defined as the interval from B to X point in the ICG; and stroke volume (SV, in ml) according to the Bernstein formula.

Mean arterial blood pressure (MAP, in mm of mercury) was calculated beat to beat from the continuous arterial pressure waveform (Finapres 2300, Ohmeda, Madison, WI) recorded at the first finger of the nondominant hand. Total peripheral resistance (TPR, in dyne $\cdot$ s $\cdot$ cm^{-5}) was calculated as MAP divided by cardiac output, which was calculated as the product of ICG-derived SV and HR.

Blood volume waveform was measured with an infrared pulse plethysmograph (1020 FC, UFI, Morro Bay, CA) clipped to the thumb of the nondominant hand. Pulse volume amplitude (PVA, in volts) was scored as the difference between the peak (maximal value) and foot (25% of maximal slope) of the pulse waveform.

A thermistor (409B YSI, Yellow Springs, OH) attached at the distal phalange of the nondominant hand's fifth finger measured surface finger temperature (FT, in degrees Fahrenheit).

Skin conductance was recorded from two 1 cm-diameter Ag/AgCl electrodes (EL507, Biopac, Goleta, CA) attached to the palmar surface of the middle phalanges of the second and third fingers of the nondominant hand. Data were low-pass filtered and downsampled to 10 Hz to calculate skin conductance level (SCL, in μSiemens).

Respiration was measured using piezoelectric respiration transducers (model 1310, Ambu Sleepmate, Glen Burnie, MD) attached around the upper chest for thoracic respiration and at the height of the umbilicus for abdominal respiration. Raw signals were converted to calibrated lung volume change using data from the fixed volume bag calibration (Morel et al., 1983). Respiratory rate (RR, in cycles per minute) and tidal volume (V_i, in ml) were calculated breath by breath. Duty cycle (T_i/T_{tot}) was calculated as the ratio of inspiratory to total breath time and inspiratory flow rate (V_i/T_i, in ml/s) as the ratio of V_i to inspiratory time.

Data Reduction

For each subject and each film clip, we quantified mixed feelings as $I[MF] = minimum(I[AMU], I[DIS])$, with $I[MF]$ being the intensity of mixed feelings, $I[AMU]$ the intensity of experienced amusement, and $I[DIS]$ the intensity of experienced disgust. Physiological data were processed in biosignal analysis software to derive period averages for each rest and film period. Reactivity was calculated between the average of each film period and the immediately preceding rest period (percent change for facial expressions; delta scores for physiological measures).

Data Analysis

Primary analyses calculated the same hypothesis tests as in our previous study (Kreibig et al., 2013) on the current data set and generated the input data for replication analyses that focused on effect sizes. Internal replicability analysis examined result stability by computing bootstrapped confidence intervals on the current study's effect sizes of physiological reactivity. External replicability analysis examined result stability by computing correlation and regression analyses on the current and previous study's effect sizes of physiological reactivity.

Primary analyses. Primary analyses tested reactivity scores for (a) significant deviation from baseline using two-sided t tests; (b) between-film effects using repeated measures analysis of variance (ANOVA) with Greenhouse-Geisser correction and eta-squared effect sizes (η^2), which, if significant, were followed up with pairwise Tukey HSD post hoc tests and Cohen's d effect sizes for mean differences; and (c) whether univariate physiological differences between conditions represented differences in profile elevation (intensity) or profile nonparallelism (pattern) using multivariate analyses of variance (MANOVAs). The condition effect tests whether profile levels (i.e., intensity) are equal between conditions, whereas the Condition $\times$ Variable effect tests whether profiles are parallel (i.e., will be significant if differences in pattern—profile scatter and shape—exist; cf. Tabachnick & Fidell, 2007). For MANOVA, physiological reactivity scores were C-transformed ($M = 100$, $SD = 10$) using within-subject standardization and reverse scored such that higher scores relate to higher physiological activation. Separate MANOVAs were calculated for contrasting amusing versus disgusting, mixed versus amusing, and mixed versus disgusting films on physiological variables that (a) had a significant univariate omnibus effect in the current study, or (b) in the previous study (Kreibig et al., 2013), or (c) were significant in both the current and previous study.

Internal replicability analysis. Stability of results from the current study was tested based on CIs of effect sizes derived from nonparametric bootstrap analysis (Efron, 1979; Wasserman & Bockenholt, 1989). For constructing bootstrap samples, subjects were used as units for resampling with replacement given the within-subject design of the current study. One thousand samples of the original sample size ($N = 48$) were created. For each bootstrap sample, the series of primary analyses was computed and resulting effect sizes were saved. Bootstrap distributions were generated by compiling respective effect sizes across bootstrap samples, from which CIs were constructed (95% adjusted bootstrap percentile [BCa] intervals for t and Cohen's d; 90% for η^2). Conventional definitions of small effect sizes (Cohen, 1988, 1992) and relative likelihood of values (Cumming, 2014) were used as lower bounds for evaluating replicability of similar-sized effects in future studies.

External replicability analysis. To identify individual physiological variables that replicated with small to large effect sizes, we inspected scatter plots of effect sizes d of physiological variables for emotional reactivity from baseline ($\Delta_{Amusement}$, $\Delta_{MixedEmotions}$, $\Delta_{Disgust}$) and for specific emotional responses ($\Delta_{Amusement}$ vs. $\Delta_{Disgust}$, $\Delta_{MixedEmotions}$ vs. $\Delta_{Amusement}$, $\Delta_{MixedEmotions}$ vs. $\Delta_{Disgust}$) between the current and previous study (cf. Stemmler et al., 2007). To quantify replicability of multivariate physiological response profiles and compare magnitude of effect sizes between studies, we calculated correlation and regression analyses over the set of 15 physiological variables on effect sizes d for baseline and specific-emotion contrasts.

To control for familywise inflation of Type I error rate, α level was set to .01 for all tests.

Results

Primary Analyses

Univariate effects of physiological reactivity. Tests of significant deviation from baseline of physiological reactivity for

amusing, mixed, and disgusting film conditions are summarized in Table 2 (left) and illustrated in Figure 1. Amusing film clips led to decreased HR, RSA_{uc}, FT, SCL, and V_i and increased RSA_c, TPR, PVA, and RR; mixed film clips led to decreased HR, RSA_{uc}, MAP, FT, and V_i and increased RSA_c, SV, and RR; and disgusting film clips led to decreased HR, MAP, FT, V_i, and V_i/T_i and increased RSA_c, SV, and RR.

ANOVAs on physiological reactivity among amusing, mixed, and disgusting film conditions (Table 2, right) identified effects on eight (out of 15) variables: RSA_{uc}, RSA_c, SV, MAP, TPR, SCL, V_i, and V_i/T_i (HR, $p = .023$). Mean differences between amusing and disgusting film conditions were present on RSA_c, MAP, SCL, V_i, and V_i/T_i (HR, RSA_{uc}, and TPR, $p = .020$, .048, .011, respectively); between mixed and amusing film conditions on SCL (SV and V_i, $p = .016$ and .040, respectively); and between mixed and disgusting film conditions on RSA_{uc} (RSA_c, $p = .023$).

Multivariate effects of physiological reactivity. Profile analysis on physiological variables that had a significant univariate omnibus effect in the current study showed condition effects for contrasts of mixed versus amusing film clips and mixed versus disgusting film clips. Condition ×Variable effects were present for all three contrasts (Table 3 top and Figure 2a). This indicates that physiological profiles in response to mixed film clips differed from those of amusing and disgusting film clips in profile elevation (intensity); physiological profiles in response to all three film conditions differed from each other in profile parallelism (pattern).

Profile analyses on physiological variables that had a significant univariate omnibus effect in the previous study (Kreibig et al., 2013) and on those that were significant in both the current and previous study showed significant condition effects for amusing versus disgusting and mixed versus amusing film clips. We again found significant Condition × Variable effects for all three contrasts (Table 3 middle and bottom and Figure 2b and c). These results indicate that physiological profiles in response to amusing film clips differed from those of mixed and disgusting film clips in profile elevation (intensity) and that all three film conditions differed from each other in profile parallelism (pattern).

Internal Replicability Analysis

Univariate effects of physiological reactivity. For tests of significant physiological deviation from baseline, Figure 3 illustrates that 95% CIs of Cohen's *d* for amusement spanned medium-to-large effects for HR, FT, and SCL, small-to-large effects for TPR and RR, and small-to-medium effects for RSA_c and V_i; CIs for mixed emotions spanned medium-to-large effects for HR, FT, RR, and V_i and small-to-medium effects for RSA_{uc} and SV; and CIs for disgust spanned medium-to-large effects for HR, RSA_c, FT, RR, and V_i, small-to-large effects for MAP, and small-to-medium effects for V_i/T_i.

Figure 4a shows 90% CIs of effect sizes (η^2) from ANOVAs on physiological reactivity among amusing, mixed, and disgusting film conditions. Results indicated that the CI for SCL covered large effects (explaining potentially 32–69% of variance); for RSA_c medium-to-large effects (9–56%); and for HR, RSA_{uc}, SV, MAP, TPR, V_i, and V_i/T_i small-to-large effects (1–42%).

For follow-up tests of pairwise mean differences between amusement and disgust, 95% CIs of Cohen's *d* enclosed medium-to-large effects for SCL and small-to-medium effects for RSA_c, MAP, TPR, V_i, and V_i/T_i; between mixed emotions and amusement, CIs enclosed medium-to-large effects for SCL and small-to-medium effects for SV; and between mixed emotions and disgust, CIs enclosed small-to-medium effects for RSA_{uc} and RSA_c (Figure 4b).

Multivariate effects of physiological reactivity. To evaluate replicability of multivariate effects, 90% CIs of effect sizes (η^2) from MANOVAs for main and interaction effects were analyzed (Table 3). Condition effects were less stable across variable sets, with CIs covering small-to-large effects for amusement versus disgust (potentially explaining 0–54% of variance in level of physiological response profiles), large effects for mixed versus amusing (14–49%), and small-to-medium effects for mixed emotion versus disgust (0–29%). Condition × Variable effects were more consistent across variable sets: CIs covered large effects throughout, potentially explaining 44–78% of variance in patterning of physiological response profiles for amusement versus disgust; 32–64% for mixed emotions versus amusement; and 17–54% for mixed emotion versus disgust. Observed effect sizes generally fell onto the lower half of a bell-shaped bootstrap distribution, suggesting that results did not capitalize on outliers.

External Replicability Analysis

Univariate effects of physiological reactivity. Inspection of scatter plots of effect sizes *d* of physiological variables for baseline contrasts (Figure 5a–c) identified HR, SCL (large effect sizes), TPR, FT, RR (medium effect sizes), RSA_{uc}, PVA, V_i, and T_i/T_{tot} (small effect sizes) for amusing film clips; HR, FT, RR (large effect sizes), V_i (medium effect size), RSA_{uc}, and SV (small effect sizes) for mixed film clips; and HR, FT, RR, V_i (large effect sizes), SV, and MAP (small effect sizes) as replicating between studies. For specific-emotion contrasts, Figure 5d–f indicated SCL (large effect size), HR, SV, MAP, TPR, RR, V_i, T_i/T_{tot}, and V_i/T_i (small effect sizes) for amusing versus disgusting film clips; SCL (large effect size), SV, MAP, TPR, PVA, RR, T_i/T_{tot}, and V_i/T_i (small effect sizes) for mixed versus amusing film clips; and HR, RSA_{uc}, TPR, SCL, and V_i/T_i (small effect sizes) for mixed versus disgusting film clips as replicating between studies.

Multivariate effects of physiological reactivity. To test similarity of results of multivariate physiological emotion profiles between the current and previous study (Kreibig et al., 2013), we calculated correlations on effect sizes *d* over the set of 15 physiological variables between the two studies. For replication of emotional reactivity from baseline, Figure 5a–c shows a tight scatter for physiological responses to amusing film clips, $r(13) = .88$, CI: .68–.96, $p < .001$, $R^2 = .77$; to mixed film clips, $r(13) = .87$, CI: .65–.96, $p < .001$, $R^2 = .76$; and to disgusting film clips, $r(13) = .81$, CI: .52–.94, $p < .001$, $R^2 = .66$. Regressions of the current (C) on the previous (P) study indicated that effect sizes were of similar magnitude between studies: for amusing film clips, $b = 1.00$ ($SE = 0.15$), $t(13) = 6.80$, $p < .001$; for mixed film clips, $b = 0.86$ ($SE = 0.13$), $t(13) = 6.48$, $p < .001$; and for disgusting film clips, $b = 0.79$ ($SE = 0.16$), $t(13) = 5.07$, $p < .001$.

For replication of specific emotional responses, Figure 5d–f shows a moderately tight scatter for physiological responses to amusing versus disgusting film clips, $r(13) = .69$, CI: .27–.89, $p < .01$, $R^2 = .48$; to mixed versus amusing film clips, $r(13) = .77$, CI: .42–.92, $p < .001$, $R^2 = .59$; and to mixed versus disgusting film clips, $r(13) = .53$, CI: .02–82, $p < .05$, $R^2 = .28$. Regression analysis indicated that effect sizes were smaller in the current than in the previous study: for amusing versus disgusting films, $b = 0.48$

Table 2. *Physiological Reactivity in Response to Film Conditions*

	Deviation from baseline			rm ANOVA				Post hoc HSD tests			Effect sizes		
Response variable	t_A	t_M	t_D	F	p	ε	η^2	A vs. D	M vs. A	M vs. D	$d_{A\text{-}D}$	$d_{M\text{-}A}$	$d_{M\text{-}D}$
Cardiac													
HR	−10.70 ***	−7.47 ***	−11.61 ***	3.99[b]	*	0.97	0.178	*	–		0.41	0.05	0.34
	(−14.00, −7.27)	(−9.54, −5.19)	(−14.87, −8.04)				(0.012, 0.338)				(0.09, 0.71)	(−0.26, 0.36)	(0.05, 0.65)
RSA_{uc}	−2.58 *	−3.02 **	0.42	6.37[b]	**	0.96	0.226	*	–	**	−0.37	0.10	−0.53
	(−4.58, −0.46)	(−5.37, −0.76)	(−1.58, 2.53)				(0.033, 0.423)				(−0.70, −0.03)	(−0.23, 0.39)	(−0.85, −0.19)
RSA_c	2.84 **	2.67 *	6.55 ***	10.35[b]	***	0.94	0.386	***	–	**	−0.72	−0.13	−0.51
	(0.82, 4.81)	(0.61, 5.00)	(4.02, 9.26)				(0.086, 0.561)				(−1.09, −0.34)	(−0.49, 0.17)	(−0.79, −0.15)
PEP	−0.40	0.46	−1.10	0.77[b]		0.98	0.040				0.09	−0.11	0.20
	(−2.55, 1.54)	(−1.48, 2.44)	(−3.08, 0.96)				(0.000, 0.150)				(−0.24, 0.39)	(−0.47, 0.18)	(−0.17, 0.50)
LVET	1.78	0.19	1.37	0.81[b]		0.96	0.030				0.06	0.17	−0.13
	(−0.65, 4.34)	(−1.88, 2.27)	(−0.65, 3.26)				(0.000, 0.126)				(−0.26, 0.37)	(−0.13, 0.49)	(−0.44, 0.17)
SV	−0.92	2.79 **	2.38 *	5.25[b]	**	0.89	0.168		*	–	−0.35	−0.44	0.07
	(−2.96, 0.97)	(0.73, 5.28)	(0.37, 4.42)				(0.023, 0.343)				(−0.64, −0.08)	(−0.73, −0.16)	(−0.25, 0.37)
Vascular													
MAP	−1.27	−2.66 *	−4.79 ***	5.91[b]	**	0.96	0.190	**	–		0.48	0.21	0.32
	(−3.24, 0.69)	(−4.56, −0.32)	(−7.10, −2.65)				(0.016, 0.365)				(0.13, 0.77)	(−0.11, 0.49)	(0.01, 0.67)
TPR	4.02 ***	1.42	0.08	5.78[b]	**	0.94	0.191	*		–	0.48	0.33	0.21
	(1.82, 6.08)	(−0.60, 3.12)	(−2.01, 2.08)				(0.018, 0.375)				(0.15, 0.79)	(0.00, 0.68)	(−0.11, 0.50)
PVA	2.55 *	0.64	1.38	2.40[b]		0.86	0.102				0.15	0.31	−0.19
	(0.42, 4.23)	(−1.42, 2.40)	(−0.59, 3.15)				(0.005, 0.267)				(−0.14, 0.44)	(0.01, 0.60)	(−0.50, 0.10)
FT	−8.13 ***	−6.75 ***	−7.20 ***	1.04[a]		0.93	0.054				−0.17	−0.22	0.01
	(−10.58, −5.92)	(−8.83, −4.60)	(−9.31, −4.82)				(0.000, 0.167)				(−0.47, 0.13)	(−0.47, 0.06)	(−0.25, 0.34)
Electrodermal													
SCL	−7.91 ***	−0.68	1.14	21.76[b]	***	0.92	0.546	***	***	–	−0.94	−0.83	−0.27
	(−10.21, −5.60)	(−2.66, 1.50)	(−0.87, 3.36)				(0.322, 0.688)				(−1.29, −0.61)	(−1.20, −0.50)	(−0.60, 0.02)
Respiratory													
RR	4.27 ***	6.61 ***	6.52 ***	1.69[b]		0.88	0.052				−0.21	−0.22	−0.04
	(2.18, 6.15)	(4.40, 8.63)	(4.30, 8.21)				(0.001, 0.174)				(−0.51, 0.08)	(−0.52, 0.10)	(−0.34, 0.27)
V_i	−2.82 **	−6.53 ***	−7.04 ***	6.33[b]	**	0.89	0.18	**	*	–	0.46	0.37	0.08
	(−4.72, −0.79)	(−8.73, −4.24)	(−9.30, −4.78)				(0.016, 0.342)				(0.12, 0.73)	(0.02, 0.69)	(−0.20, 0.38)
T_i/T_{tot}	−1.86	−0.36	0.16	2.72[b]		0.93	0.086				−0.3	−0.24	−0.08
	(−3.93, 0.16)	(−2.38, 1.72)	(−1.93, 1.97)				(0.003, 0.242)				(−0.62, −0.01)	(−0.53, 0.05)	(−0.38, 0.20)
V_i/T_i	0.04	−1.66	−3.08 **	5.89[b]	**	0.94	0.182	**	–		0.45	0.21	0.33
	(−1.94, 1.86)	(−3.57, 0.51)	(−5.14, −0.94)				(0.018, 0.332)				(0.15, 0.74)	(−0.13, 0.48)	(0.00, 0.60)

Note. See Figure 1 for abbreviations of physiological variables. CIs for effect sizes (95% for t and d values, 90% for η^2) are given in parentheses based on 1,000 bootstrap samples. A = amusing film clips; M = mixed film clips; D = disgusting film clips; rm ANOVA = repeated measures analysis of variance; post hoc HSD tests = Tukey honestly significant difference post hoc tests; d = Cohen's d effect size.
[a]$df = 2,94$; [b]$df = 2,92$; [c]$df = 2,90$; [d]$df = 2,88$; [e]$df = 2,86$; [f]$df = 2,84$; [g]$df = 2,82$; [h]$df = 2,80$; [i]$df = 2,76$.
*$p < .05$. **$p < .01$. ***$p < .001$.

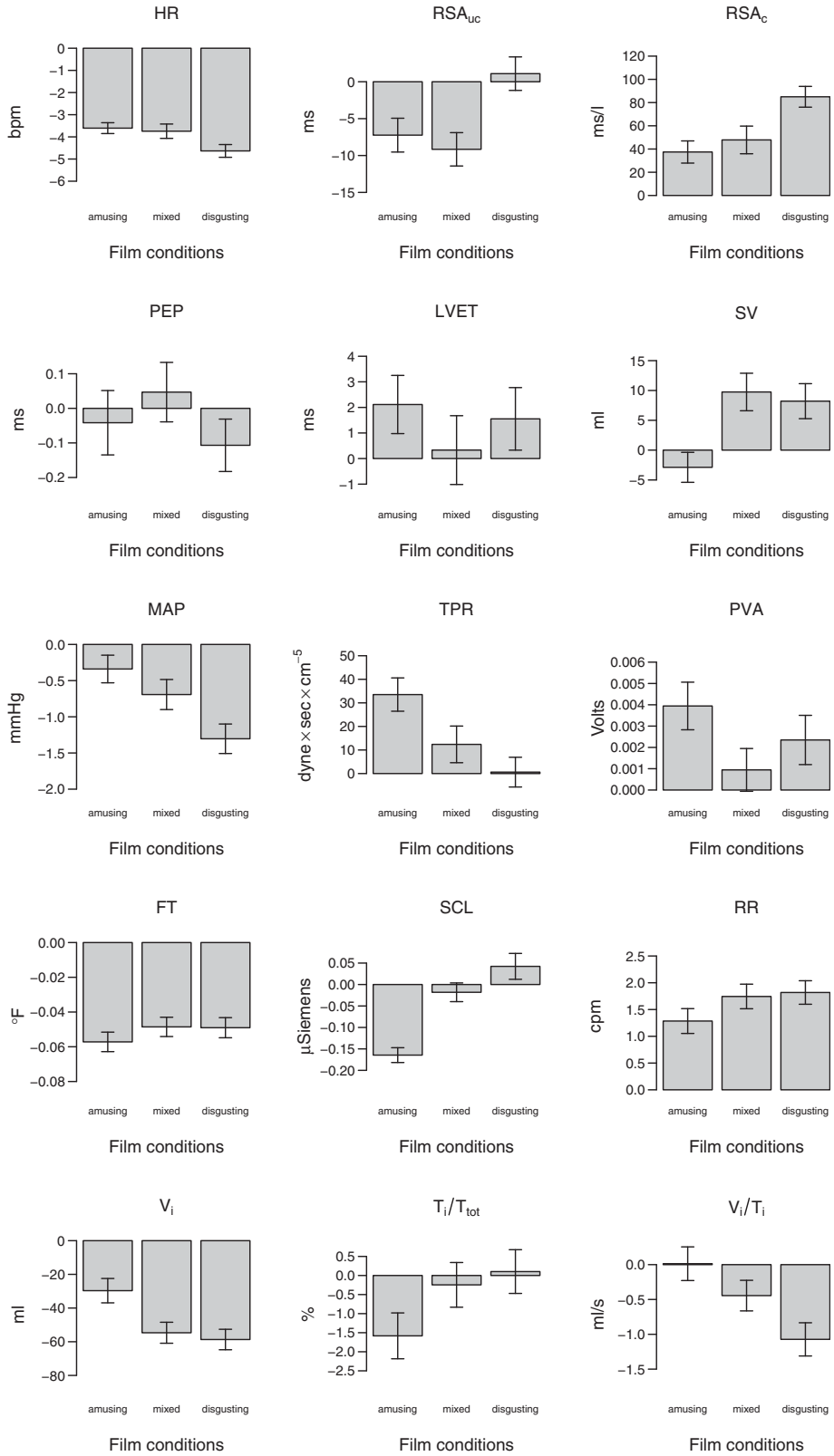

Figure 1. Condition means of amusing, mixed, and disgusting film clips for cardiovascular, electrodermal, and respiratory variables. Illustrated values depict change in physiological activation during film clips from prefilm baselines. Error bars indicate ± 1 standard error of the mean (*SE*). HR = heart rate; RSA_{uc} = uncorrected respiratory sinus arrhythmia; RSA_c = breathing-corrected RSA; PEP = pre-ejection period; LVET = left ventricular ejection time; SV = stroke volume; MAP = mean arterial pressure; TPR = total peripheral resistance; PVA = pulse volume amplitude; FT = finger temperature; SCL = skin conductance level; RR = respiration rate; V_i = inspiratory volume; T_i/T_{tot} = duty cycle; V_i/T_i = inspiratory flow rate.

Table 3. *Profile Analysis of Physiological Responses Between Film Conditions*

		Condition				Condition x Variable				
(a) Condition	λ	F(1,47)		p	η^2	CI	λ	F(7,41)		p	η^2	CI
Amusing vs. disgusting	0.98	0.93		0.34	0.019	0.000–0.100	0.26	16.64	***	3.61E-10	0.740	0.613–0.784
Mixed vs. amusing	0.70	19.76	***	5.33E-05	0.296	0.139–0.431	0.41	8.37	***	2.54E-06	0.588	0.439–0.639
Mixed vs. disgusting	0.85	8.01	**	6.84E-03	0.146	0.032–0.286	0.54	5.06	***	3.38E-04	0.463	0.244–0.542
		Condition						Condition x Variable				
(b) Condition	λ	F(1,47)		p	η^2	CI	λ	F(10,38)		p	η^2	CI
Amusing vs. disgusting	0.60	31.22	***	1.13E-06	0.399	0.210–0.541	0.38	6.29	***	1.32E-05	0.623	0.441–0.665
Mixed vs. amusing	0.68	22.35	***	2.10E-05	0.322	0.135–0.486	0.44	4.82	***	1.77E-04	0.559	0.406–0.587
Mixed vs. disgusting	0.96	1.81		0.18	0.037	0.001–0.185	0.61	2.46	*	0.022	0.393	0.166–0.423
		Condition						Condition x Variable				
(c) Condition	λ	F(1,47)		p	η^2	CI	λ	F(6,42)		p	η^2	CI
Amusing vs. disgusting	0.75	15.88	***	2.34E-04	0.253	0.092–0.402	0.38	11.65	***	1.20E-07	0.625	0.462–0.699
Mixed vs. amusing	0.69	20.64	***	3.86E-05	0.305	0.143–0.450	0.45	8.71	***	3.56E-06	0.554	0.323–0.611
Mixed vs. disgusting	0.99	0.28		0.60	0.006	0.000–0.048	0.62	4.30	**	0.002	0.380	0.177–0.501

Note. Results are based on multivariate analysis of variance using Wilks' lambda (λ). (a) Variables that had a significant univariate omnibus effect in the current study; (b) variables that had a significant univariate omnibus effect in the previous study (Kreibig et al., 2013); and (c) set of seven variables that reproduced across the current study and previous study (Kreibig et al., 2013). 90% confidence intervals (CIs) for effect sizes (η^2) were derived based on 1,000 bootstrap samples.

($SE = 0.14$), $t(13) = 3.40$, $p < .01$; for mixed versus amusing films, $b = 0.53$ ($SE = 0.12$), $t(13) = 4.30$, $p < .001$; and for mixed versus disgusting films, $b = 0.39$ ($SE = 0.18$), $t(13) = 2.23$, $p < .05$.

Discussion

The current study examined internal and external replicability of the physiological differentiation of amusing, disgusting, and a mixed emotional state based on data from a direct replication and our previous study (Kreibig et al., 2013). Three levels of analysis produced converging evidence for replicability: Primary analyses produced similar findings to our previous study, demonstrating (a) univariate differentiation of the mixed emotional state from its pure constituent emotions on physiological response variables, and (b) intensity and pattern differences of the multivariate physiological response profile of the mixed emotional state from those of pure amusement and pure disgust, supporting the emergence account of mixed emotions; internal replicability analysis indicated (c) likely replication of a considerable number of univariate effects in future studies, and (d) strong likelihood of replication of the physiological differentiation according to multivariate response profiles; and external replicability analysis suggested (e) replicable univariate differentiation according to a subset of individual physiological variables, graded according to magnitude of effect size, and (f) replicable multivariate physiological response profiles for baseline and specific-emotion contrasts between studies. This study also gives useful indications for selection of physiological measures that allow replicable differentiation of amusement, disgust, and a mixed emotional state. We discuss implications of core findings by emotion in the following.

Replicability of the Psychophysiology of Amusement and Disgust

Amusement. Consistent with our previous study (Kreibig et al., 2013), amusement led to decreased cardiac activity, peripheral vasoconstriction, decreased electrodermal activity, and faster and shallower breathing; the current study additionally showed increased vagal reactivity and PVA. Both internal and external replicability analyses identified HR, TPR, FT, SCL, RR, and V_i as measures with replicable results, suggesting involvement of cardiovascular, electrodermal, and respiratory systems in amusement.

Results of external replicability analysis demonstrated excellent stability of the multivariate physiological response profile of amusement and commensurate magnitude of effect sizes between studies. These results are considerably stronger than those that we calculated based on a set of studies by Levenson and colleagues (Levenson, Ekman, & Friesen, 1990; Levenson, Ekman, Heider, & Friesen, 1992): $r = .17$, $t(5) = 0.39$, $p = .71$, for replication of responses on seven physiological measures to amusement induction in Minangkabau men and a mixed-sex American comparison sample. Our results are comparable to those from a study by Tsai, Levenson, and Carstensen (2000) between experimental groups of young and old European and Chinese Americans: mean $r = .89$, $t(3) = 3.47$, $p < .05$, for replication of responses on five physiological measures to amusing films.

Disgust. As in our previous study, disgust led to decreased cardiac activity, vasoconstriction, and rapid and shallow breathing. While the current study showed effects neither on PEP nor on SCL—two primary sympathetic indicators—it additionally demonstrated vasodilation and decreased central inspiratory drive. Both internal and

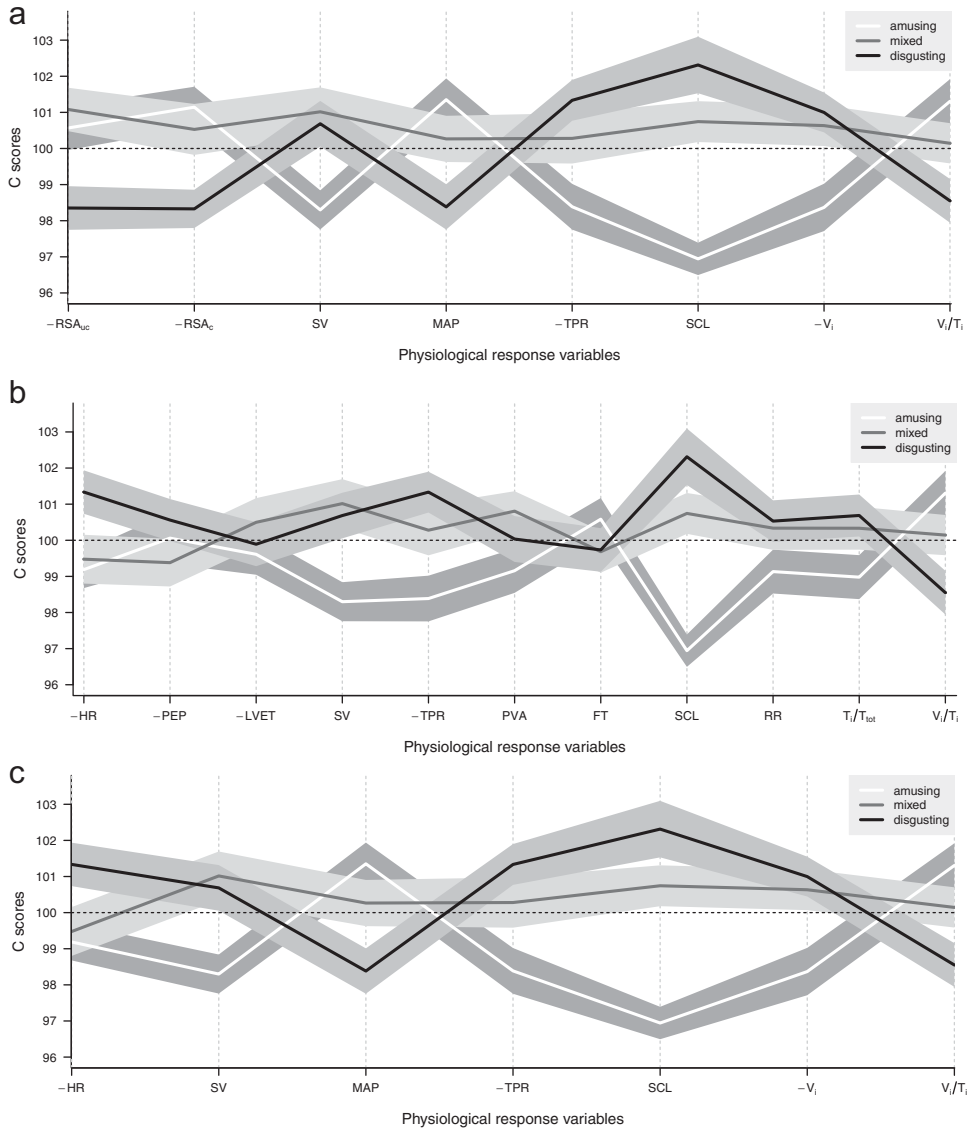

Figure 2. Physiological response profiles for amusing, mixed, and disgusting film conditions based on (a) variables that had a significant univariate omnibus effect in the current study, (b) variables that had a significant univariate omnibus effect in the previous study (Kreibig et al., 2013), and (c) set of seven variables that reproduced across the current and previous study (Kreibig et al., 2013). Change scores were transformed onto a *C* scale ($M = 100$, $SD = 10$), and variables were reverse scored such that higher scores indicate higher physiological activation. Gray bands illustrate ± 1 standard error of the mean (*SE*). FT = finger temperature; HR = heart rate; LVET = left ventricular ejection time; MAP = mean arterial pressure; PEP = pre-ejection period; PVA = pulse volume amplitude; RR = respiration rate; RSA_c = breathing-corrected RSA; RSA_{uc} = uncorrected respiratory sinus arrhythmia; SCL = skin conductance level; SV = stroke volume; T_i/T_{tot} = duty cycle; TPR = total peripheral resistance; V_i = inspiratory volume; V_i/T_i = inspiratory flow rate.

external replicability analyses identified HR, MAP, FT, RR, and V_i as measures with replicable results, suggesting involvement of cardiovascular and respiratory systems in disgust.

Results of external replicability analysis showed excellent stability of the multivariate physiological response profile of disgust and similar magnitude of effect sizes between studies. These results are comparable to those of prior studies in the literature, for which we calculated $r = .61$, $t(5) = 1.74$, $p = .14$, for replication of responses on seven physiological measures to disgust induction in Minangkabau men and a mixed-sex American sample (Levenson et al., 1990, 1992); $r = .64$, $t(3) = 1.44$, $p = .24$, for replication of responses on five physiological measures to disgust films in undergraduates from two West coast universities (Gross, 1998; Gross & Levenson, 1993); and $r = .75$, $t(7) = 3.03$, $p < .05$, for replication of responses on the same five measures to disgust films in a male and female sample assessed 18 months apart.

Amusement versus disgust. In both studies, amusement differed from disgust by (marginally) less pronounced HR deceleration, higher TPR, and lower SCL. Whereas fast and shallow breathing characterized both emotions, decreased central inspiratory drive in disgust suggested different underlying mechanisms.

Divergent involvement of physiological response systems between studies was implied by differentiation of amusement and disgust on indicators of sympathetic activity (PEP, SV, and FT) and respiratory rhythm (T_i/T_{tot}) in the previous study in contrast to differentiation on indicators of parasympathetic activity (RSA_{uc} and RSA_c) and peripheral resistance (MAP) in the current study.

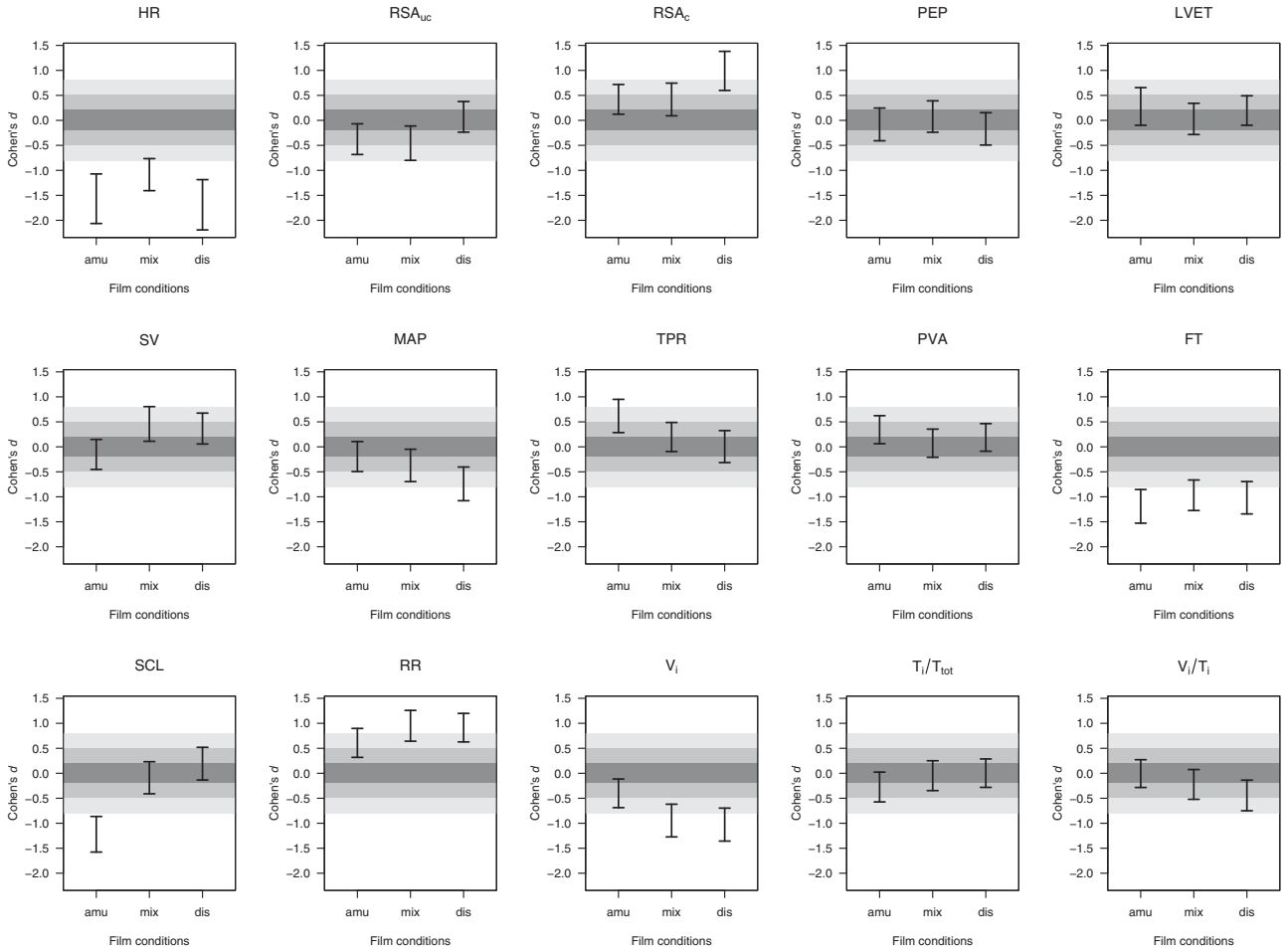

Figure 3. 95% CIs of effect size estimates (Cohen's *d*) for tests of significant deviation from baseline of physiological reactivity for amusing (amu), mixed (mix), and disgusting (dis) film conditions. Gray bands indicate boundaries for small, medium, and large effect sizes ($d = 0.20$, 0.50, and 0.80; Cohen, 1992). See Figure 1 for abbreviations of physiological measures.

Uninstructed versus regulated film viewing tasks may sensitize different response systems. Accordingly, prior research has found increased sympathetic activation in the presence of stressors, but increased parasympathetic activation upon additional engagement of emotion regulation efforts (Butler, Wilhelm, & Gross, 2006; Ingjaldsson, Laberg, & Thayer, 2003; Sloan & Epstein, 2005; Smith et al., 2011). Both internal and external replicability analyses identified MAP, TPR, SCL, V_i, and V_i/T_i as measures with replicable results, pointing to involvement of vascular, electrodermal, and respiratory systems in differentiating amusement and disgust.

Results of multivariate profile analysis support the view that the physiology of amusement and disgust is robustly distinct across different definitions of variable sets. Taken together, these results suggest robust multivariate differentiation of physiological response profiles, whereas robustness of individual physiological measures was more graded.

Replicability of the Psychophysiology of Mixed Emotions

Mixed emotions. As in our previous study, cardiac deactivation, peripheral vasoconstriction, and fast and shallow breathing characterized the mixed emotional state. We additionally observed increased vagal reactivity and decreased vascular resistance in the current study. Internal and external replicability analyses identified HR, RSA_{uc}, SV, FT, RR, and V_i as measures with replicable results, suggesting involvement of cardiac sympathetic and parasympathetic, vascular, and respiratory systems in the mixed emotional state. External replicability analysis further indicated excellent stability of the multivariate physiological response profile, with effect sizes of similar magnitude between studies.

Mixed emotions versus amusement. Consistent with our previous study, the mixed emotional state differed from amusement by higher preload (which is the main mechanism at lower work rates contributing to [marginally] higher SV) and higher electrodermal activity. Whereas the current study also showed differentiation through marginally lower respiratory volume, the previous study additionally found lower systemic resistance, higher peripheral vasoconstriction, and more rapid breathing. Both internal and external replicability analyses identified SV and SCL as measures with replicable results, suggesting involvement of sympathetic α-adrenergic venous and cholinergic electrodermal systems in differentiating the mixed emotional state and amusement.

Mixed emotions versus disgust. The mixed emotional state differed from disgust by less vagal reactivity in the current study, in contrast to less α-adrenergic sympathetic influence and a shift to expiration in the previous study. Notably, both internal and external replicability analyses identified RSA_{uc} as a measure with replicable

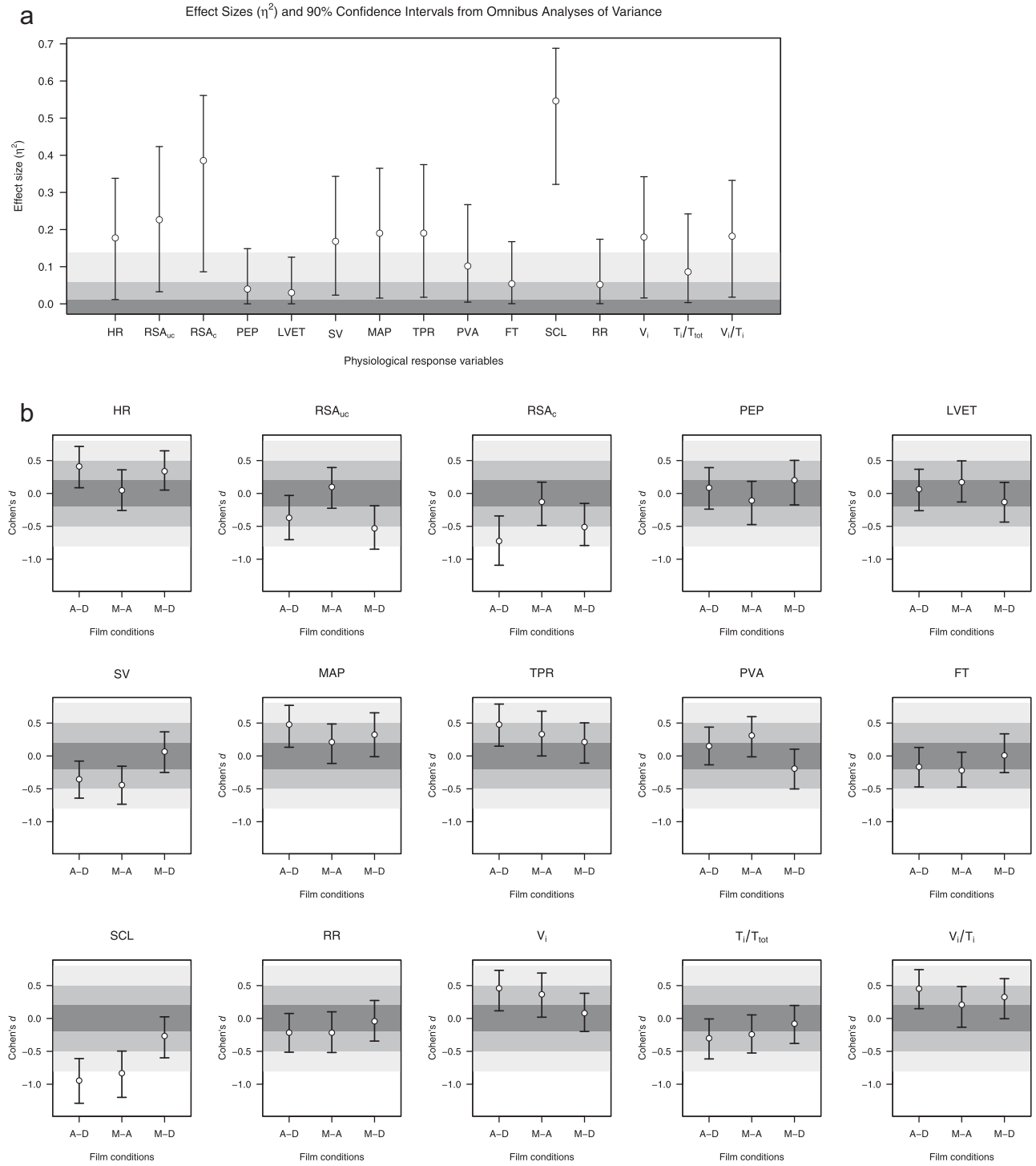

Figure 4. (a) 90% CIs of effect sizes (η^2) from ANOVAs on univariate differences of physiological reactivity among amusing, mixed, and disgusting film conditions. Gray bands indicate boundaries for small, medium, and large effect sizes ($\eta^2 = .010$, .059, and .138; Cohen, 1988). (b) 95% CIs of effect size estimates (Cohen's *d*) for follow-up tests of pairwise mean differences on physiological reactivity between amusing and disgusting (A–D), mixed and amusing (M–A), and mixed and disgusting (M–D) film conditions. Gray bands indicate boundaries for small, medium, and large effect sizes ($d = 0.20$, 0.50, and 0.80; Cohen, 1992). See Figure 1 for abbreviations of physiological measures.

results, suggesting involvement of the vagal system in differentiating the mixed emotional state from disgust—an effect that was not identified by statistical significance testing.

Results of multivariate profile analysis support the view that the physiology of the mixed emotional state is robustly distinct from its pure constituent emotions across different definitions of variable sets, with excellent stability of the physiological differentiation from amusement and fair stability of that from disgust, albeit with smaller effect sizes in the current than the previous study (less than half the magnitude for disgust).

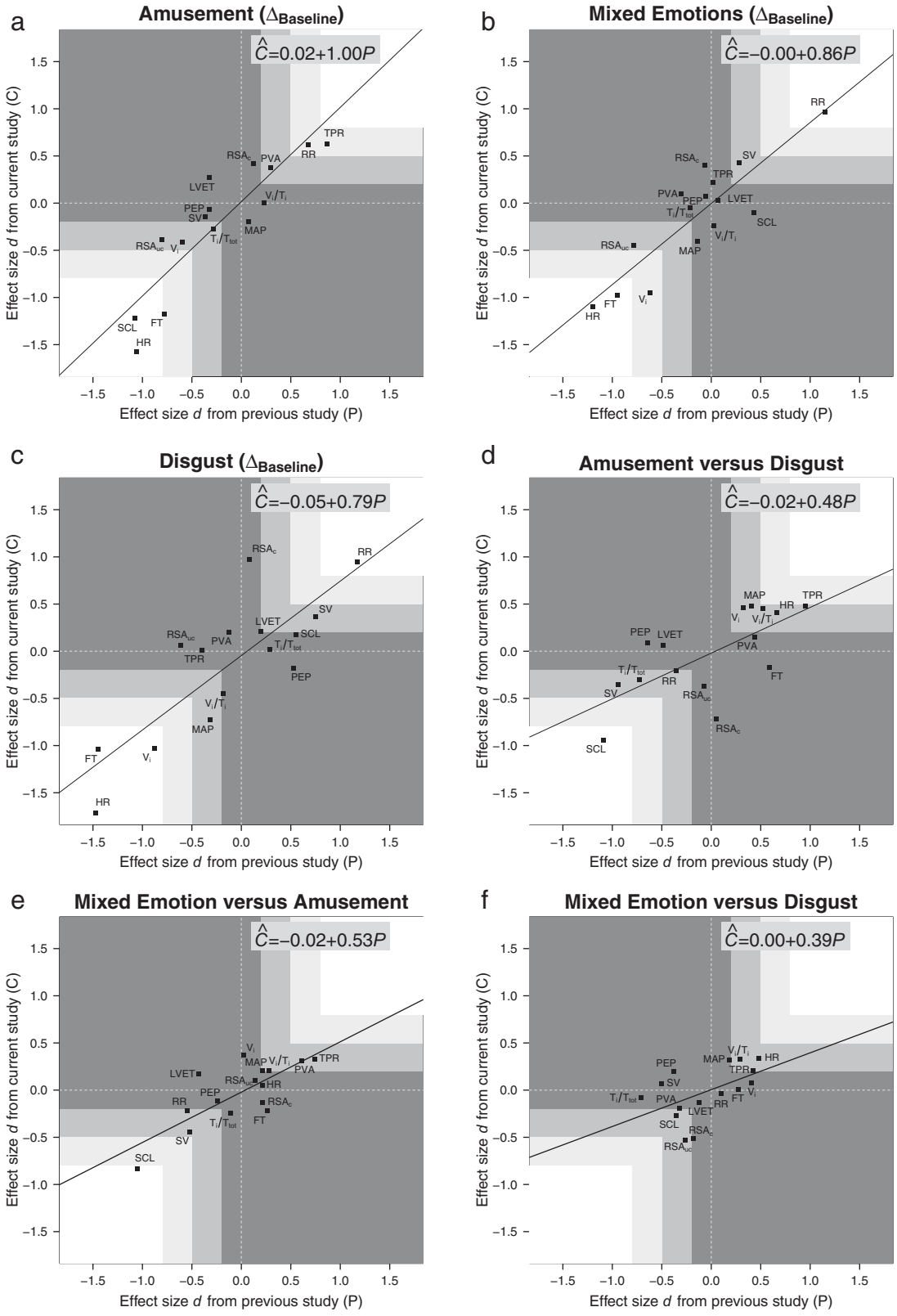

Figure 5. Effect size (Cohen's *d*) comparisons between the current and previous study (Kreibig et al., 2013) of physiological reactivity from baseline for (a) amusement, (b) mixed emotions, and (c) disgust and for specific emotional responses contrasting (d) amusement versus disgust, (e) mixed emotions versus amusement, and (f) mixed emotions versus disgust. Gray bands indicate boundaries for small, medium, and large effect sizes ($d = 0.20$, 0.50, and 0.80; Cohen, 1992). Equations give the regression of the current (C) on the previous (P) study.

Empirically, this speaks to the robustness of the finding of multivariate physiological profile differentiation of these three emotion conditions (whereas individual physiological measures were not as uniformly robust). Theoretically, this lends additional support to the emergence account of mixed emotions, which predicted that physiological responses of mixed emotional states should differ from both of their constituent emotions in pattern. In contrast, this finding is inconsistent with both the nondifferentiation account, which predicted that physiological responses of mixed emotional states should not differ from one of the pure constituent emotions but should differ in pattern from the other, and the additive account, which predicted that physiological responses of mixed emotional states should differ from one of their constituent emotions in intensity and from the other in pattern. This finding may be taken to suggest that the mixed emotional state constitutes a unique state with distinct characteristics, including a differential physiological response, from those of its pure constituent emotions (cf. Scherer, 1984). Still, a separate profile is also expected to occur if physiological variables have different transfer functions with growing intensity, such as when the same amount of arousal increase leads to an unrestricted increase in one variable but a flattened increase in another variable.[1]

Limitations and Directions for Future Research

Results of the present study should be interpreted in the context of three types of limitations—procedural limitations of the study of the psychophysiology of mixed emotional states, procedural limitations of the direct replication, and statistical limitations of internal and external replicability analysis.

We discussed procedural limitations of the study of the psychophysiology of mixed emotional states in detail in our prior report (Kreibig et al., 2013). These included (a) inclusion of only college-aged female participants of multiracial American ethnicity, (b) use of film clips for emotion elicitation, (c) focus on peripheral physiological measures, (d) highlighting of amusement, disgust, and a mixed emotional state of both, (e) absence of control for contextual effects (e.g., behavioral differences in spontaneous responses to the film viewing task; Stemmler, Heldmann, Pauls, & Scherer, 2001), and (f) choice of our primary statistical tests. With eye to the latter, profile differences might have been enlarged by selecting only those variables for profile analysis, which had a significant omnibus effect, and contrasts might have been maximized by performing three separate profile analyses for each of the comparisons. Accordingly, a three-level multivariate analysis of variance would be calculated and contrasts performed in that space, although others discourage calculating contrasts with a pooled error term in within-designs (e.g., Kirk, 1995).

Procedural limitations of the direct replication included data collection (a) at a later time and different season; (b) with different participants who closely resembled our previous sample in age, but may not have for other characteristics; and (c) within the larger context of a reappraisal task. Particularly the context of the reappraisal task might have introduced variability in the observed physiological responses: For experimental randomization and management of time constraints, each participant saw a different subset of only six film clips for each emotion category in the current study instead of the same set of 15 film clips in the previous study; additionally, the activation of emotion regulation goals might have led to more parasympathetic than sympathetic autonomic activation, as discussed above. Still, if anything, these deviations from our original protocol would have made it less likely to replicate emotion-related physiological changes for amusement, disgust, and a mixed emotional state.

Statistical limitations of internal and external replicability analysis included (a) presumption of representativeness of the current sample for the population it is intended to represent (else biases in data collection will also be reflected in bootstrapped results), (b) inflated estimates of true replicability by internal replicability analysis, (c) actual correspondence of 95% CIs to 83% prediction intervals for effect sizes that would be given by future replications (Cumming & Maillardet, 2006), and (d) choice of effect size measures. With respect to the latter, first, standardized differences effect sizes (e.g., Cohen's *d*) are expressed in units of standard deviation. Their units stretch out or in with smaller or larger standard deviations found upon repetition of the experiment (Cumming, 2014). The smaller number of film stimuli and resulting larger variability might have caused smaller effect sizes in the current than in the previous study. Second, although variance-accounted-for effect sizes (e.g., η^2) allow statements about the percentage of explained variance within the sample, they tend to overestimate effect sizes in the population and future samples (Snyder & Lawson, 1993) because their calculation capitalizes on sample-specific variance (Vacha-Haase & Thompson, 2004). While bias-corrected effect sizes exist and should be used in future research, we chose to employ these effect sizes for comparability with our previous study.

In conclusion, the present study demonstrated support for the direct replicability of emotion-related physiological changes in general and the nature of the psychophysiological response of a mixed emotional state in particular. Result replicability "is almost universally accepted as the most important criterion of genuine scientific knowledge" (Rosenthal & Rosnow, 1984, p. 9). As a next step, conceptual replication should explore alternative accounts for the nature of the psychophysiology of mixed emotional states, drawing on alternative induction methods, such as the paired picture paradigm (Schimmack & Colombe, 2007), and alternative analysis approaches, such as time course analysis supplemented by continuous feeling self-report of mixed emotional states (Mauss et al., 2005). Only if we verify that the current findings are robust can future research further elucidate the psychophysiology of mixed emotional states.

References

Butler, E. A., Wilhelm, F. H., & Gross, J. J. (2006). Respiratory sinus arrhythmia, emotion, and emotion regulation during social interaction. *Psychophysiology*, *43*, 612–622. doi: 10.1111/j.1469-8986.2006.00467.x

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

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. doi: 10.1037/0033-2909.112.1.155

Cumming, G. (2008). Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science*, *3*, 286–300. doi: 10.1111/j.1745-6924.2008.00079.x

1. We thank an anonymous reviewer for pointing out this alternative theoretical account supported by our data.

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

Cumming, G., & Maillardet, R. (2006). Confidence intervals and replication: Where will the next mean fall? *Psychological Methods*, *11*, 217–227. doi: 10.1037/1082-989X.11.3.217

Eckberg, D. L. (1983). Human sinus arrhythmia as an index of vagal cardiac outflow. *Journal of Applied Physiology*, *54*, 961–966.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, *7*, 1–26. doi: 10.1214/aos/1176344552

Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology*, *74*, 224–237. doi: 10.1037/0022-3514.74.1.224

Gross, J. J., & Levenson, R. W. (1993). Emotional suppression: Physiology, self-report, and expressive behavior. *Journal of Personality and Social Psychology*, *64*, 970–986. doi: 10.1037/0022-3514.64.6.970

Ingjaldsson, J. T., Laberg, J. C., & Thayer, J. F. (2003). Reduced heart rate variability in chronic alcohol abuse: Relationship with negative mood, chronic thought suppression, and compulsive drinking. *Biological Psychiatry*, *54*, 1427–1436. doi: 10.1016/S0006-3223(02)01926-1

Kirk, R. E. (1995). *Experimental design: Procedures for behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Kreibig, S. D., Samson, A. C., & Gross, J. J. (2013). The psychophysiology of mixed emotional states. *Psychophysiology*, *50*, 799–811. doi: 10.1111/psyp.12064

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

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

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, *27*, 363–384. doi: 10.1111/j.1469-8986.1990.tb02330.x

Levenson, R. W., Ekman, P., Heider, K., & Friesen, W. V. (1992). Emotion and autonomic nervous system activity in the Minangkabau of West Sumatra. *Journal of Personality and Social Psychology*, *62*, 972–988. doi: 10.1037/002-3514.62.6.972

Lozano, D. L., Norman, G., Knox, D., Wood, B. L., Miller, B. D., Emery, C. F., & Bernston, G. G. (2007). Where to B in dZ/dt. *Psychophysiology*, *44*, 113–119. doi: 10.1111/j.1469-8986.2006.00468.x

Mauss, I. B., Levenson, R. W., McCater, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion*, *5*, 175–190. doi: 10.1037/1528-3542.5.2.175

Morel, D. R., Forster, A., & Suter, P. M. (1983). Noninvasive ventilatory monitoring with bellows pneumographs in supine subjects. *Journal of Applied Physiology*, *55*, 598–606.

Pe, M. L., & Kuppens, P. (2012). The dynamic interplay between emotions in daily life: Augmentation, blunting, and the role of appraisal overlap. Emotion, *12*, 1320–1328. doi: 10.1037/a0028262

Ritz, T., Thöns, M., & Dahme, B. (2001). Modulation of respiratory sinus arrhythmia by respiration rate and volume: Stability across posture and volume variations. *Psychophysiology*, *38*, 858–862. doi: 10.1111/1469-8986.3850858

Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York, NY: McGraw-Hill.

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

Samson, A. C., Kreibig, S. D., Soderstrom, B., Wade, A. A., & Gross, J. J. (in press). *Eliciting positive, negative, and mixed emotional states: A film library for affective scientists*. Cognition and Emotion. doi: 10.1080/02699931.2015.1031089.

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

Schimmack, U. (2001). Pleasure, displeasure, and mixed feelings: Are semantic opposites mutually exclusive? Cognition and *Emotion*, *15*, 81–97.

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

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*, 90–100. doi: 10.1037/a0015108

Schulz, S. M., Ayala, E., Dahme, B., & Ritz, T. (2009). A MATLAB toolbox for correcting within-individual effects of respiration rate and tidal volume on respiratory sinus arrhythmia during variable breathing. *Behavior Research Methods*, *41*, 1121–1126. doi: 10.3758/BRM.41.4.1121

Sloan, D. M., & Epstein, E. M. (2005). Respiratory sinus arrhythmia predicts written disclosure outcome. *Psychophysiology*, *42*, 611–615. doi: 10.1111/j.1469-8986.2005.347.x

Smith, T. W., Cribbet, M. R., Nealey-Moore, J. B., Uchino, B. N., Williams, P. G., MacKenzie, J., & Thayer, J. F. (2011). Matters of the variable heart: Respiratory sinus arrhythmia response to marital interaction and associations with marital quality. *Journal of Personality and Social Psychology*, *100*, 103–119. doi: 10.1037/a0021136

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education*, *61*, 334–349. doi: 10.1080/00220973.1993.10806594

Stemmler, G., Heldmann, M., Pauls, C. A., & Scherer, T. (2001). Constraints for emotion specificity in fear and anger: The context counts. *Psychophysiology*, *38*, 275–291. doi: 10.1111/1469-8986.3820275

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. doi: 10.1016/j.ijpsycho.2007.03.019

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn and Bacon.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher*, *25*, 26–30. doi: 10.3102/0013189X025002026

Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. New York, NY: Guilford.

Tsai, J. L., Levenson, R. W., & Carstensen, L. L. (2000). Autonomic, subjective and expressive responses to emotional films in older and younger Chinese Americans and European Americans. *Cultural Diversity and Ethnic Minority Psychology*, *15*, 684–693. doi: 10.1037/0882-7974.15.4.684

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, *51*, 473–481. doi: 10.1037/0022-0167.51.4.473

Wahlsten, D. (1991). Sample size to detect a planned contrast and a one degree-of-freedom interaction effect. *Psychological Bulletin*, *110*, 587–595. doi: 10.1037/0033-2909.110.3.587

Wasserman, S., & Bockenholt, U. (1989). Bootstrapping: Applications to psychophysiology. *Psychophysiology*, *26*, 208–221. doi: 10.1111/j.1469-8986.1989.tb03159.x

(Received March 21, 2014; Revised January 21, 2015; Accepted January 27, 2015)