12-2017

A Meta-Analysis of the Facial Feedback Literature: Effects of Facial Expressions on Emotional Experience are Small and Variable

Nicholas Alvaro Coles
University of Tennessee, colesn@vols.utk.edu

To the Graduate Council:

I am submitting herewith a thesis written by Nicholas Alvaro Coles entitled "A Meta-Analysis of the Facial Feedback Literature: Effects of Facial Expressions on Emotional Experience are Small and Variable." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Arts, with a major in Psychology.

Jeff T. Larsen, Major Professor

We have read this thesis and recommend its acceptance:

Lowen Gaertner, Michael Olson

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

A Meta-Analysis of the Facial Feedback Literature: Effects of Facial Expressions on Emotional Experience are Small and Variable

A Thesis Presented for the

Masters of Arts

Degree

The University of Tennessee, Knoxville

Nicholas Alvaro Coles

December 2017

"Columnated ruins domino."

- B. Wilson and V. Parks

Dedicated to ambitious SMiLErs.

Acknowledgments

I thank Jeff Larsen and Heather Lench for their mentorship throughout this project, Lowell Gaertner and Michael Olson for serving on the thesis committee, Joyce Kuribayashi and Nicholas Harp for assistance coding the studies, Ashley Kuelz for assistance in writing the R code, Jessica Van Westering, Sheryl Coles, and Deanna Jordan for general support, and the Psychological Methods Discussion Group for useful comments and discussions.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1452154 awarded to Nicholas Coles. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Abstract

The facial feedback hypothesis suggests that our facial expressions influence our emotional experience. In light of Wagenmakers et al.'s (2016) failure to replicate Strack, Martin, and Stepper's (1988) seminal demonstration of facial feedback effects, a meta-analysis was conducted on 286 effect sizes derived from 136 facial feedback studies. Results revealed that the overall effect of facial feedback on affective experience was significant, but small (d = .20, p < .000000005).

Approximately 70% of variation in facial feedback effect sizes is due to heterogeneity, which suggests that facial feedback effects are stronger in some circumstances than others. Eleven potential moderators were examined, and three were associated with differences in effect sizes: (1) Type of affective reaction: Facial feedback influenced emotional experience (e.g., reported amusement) and, to an even greater degree, perceptions of stimuli's affective quality (e.g., funniness of cartoons). However, after controlling for publication bias, there was little evidence that facial feedback influenced perceptions of affective quality. (2) Presence of emotional stimuli: Facial feedback effects on emotional experience were larger in the absence of emotionally evocative stimuli (e.g., cartoons). (3) Type of stimuli: When participants are presented with emotionally evocative stimuli, facial feedback effects were larger in the presence of some types of stimuli (e.g., imagined scenarios) than others (e.g., pictures).

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

“Sometimes your joy is the source of your smile, but sometimes your smile can be the source of your joy." - Thích Nhất Hạnh

Thích Nhất Hạnh makes two statements about facial expressions: one more obvious, the other a recent source of controversy in the social sciences. The first is the widely-accepted notion that our emotions often cause changes in our facial expressions (for a review, see Keltner, Ekman, Gonzaga, & Beer, 2003). When we are happy, we often smile. The second refers to the *facial feedback hypothesis,* which suggests that our facial expressions can influence our emotional experience. Is it possible that our smiles can be the source of our joy? Decades of research have assessed the facial feedback hypothesis and the circumstances under which facial expressions affect emotion. However, the last meta-analytic review occurred 30 years ago and a recent multi-lab failure-to-replicate has created uncertainty regarding the reliability of this phenomenon. Amid this uncertainty, a comprehensive meta-analysis of the facial feedback hypothesis literature is in order.

Uncertainty in the Facial Feedback Brickyard

Poincaré (1902) likened the task of a scientist to that of a bricklayer: scientists assemble bricks [empirical findings] in order to build edifices [theories]. But, as added by Forscher (1963), “If the bricks were faulty…the edifice would crumble, and this kind of disaster could be very dangerous to the innocent users of the edifice…” (p. 339).

The facial feedback edifice has been under construction for nearly 150 years. Discussions of the effect of facial expressions on emotional experience can be traced back to the mid-1800’s (Gratiolet, 1865; Piderit, 1858) and were further popularized by Charles Darwin (1872) and William James (1884). Contemporary researchers, however, are most familiar with Strack, Martin, and Stepper's (1988) classic pen-in-mouth paper. In two studies, participants held a pen in their mouth in a manner that either forced them to smile (pen held in teeth) or prevented them from doing so (pen held by lips). While maintaining these poses, participants viewed humorous cartoons and reported how amused they felt. Consistent with the facial feedback hypothesis, participants who posed smiles reported being more amused by the cartoons than those who were prevented from smiling.

Strack and colleagues' (1988) paper is one of the most well-known facial feedback findings. It has been cited over 1,600 times (Google Scholar, October 2017) and appears in many introductory psychology textbooks. It is not, however, the only brick in the edifice. Nearly 100 other studies have examined the facial feedback hypothesis, and these studies have used a variety of manipulations, stimuli, and emotional outcomes. Along the way, there were published reports that failed to find that facial expressions can influence emotional experience (Tourangeau & Ellsworth, 1979) and some criticism about methodological limitations (Buck, 1980). However, nearly all narrative reviews of the evidence have concluded that facial expressions can indeed influence emotional experience (Adelmann & Zajonc, 1989; Cappella, 1993; Izard, 1990; Laird, 1984; Laird & Lacasse, 2014; Martin, 1990; McIntosh, 1996; Price & Harmon-Jones, 2015; Soussignan, 2004; Whissell, 1985).

Just a year ago, there appeared to be little doubt about the validity of the facial feedback hypothesis. However, more recently, Wagenmakers and colleagues (2016) reported a pre-registered replication of one of Strack et al.'s (1988) two studies. Seventeen separate laboratories conducted replications. None found that the pen smiling manipulation made people feel significantly more amused while viewing cartoons. This large failure-to-replicate raises doubts about the reliability of a classic facial feedback finding and, by extension, the entire facial feedback hypothesis. But is the structure of this edifice compromised? Currently there is no consensus[1] and there is uncertainty in the facial feedback brickyard. In this thesis, meta-analysis is used to evaluate the facial feedback edifice. First, however, I review why this edifice is worth examining.

Why the facial feedback hypothesis matters.

Cognitive science historically viewed the central nervous system as the sole substrate of cognition. However, this classical view of cognition was challenged by the *embodied cognition* paradigm, which maintained that cognition is influenced by and partially dependent upon non-central physiological processes. By positing that facial expressions can influence emotional experience, the facial feedback hypothesis suggests that affective processes, like other cognitive processes, are embodied. This idea is featured in many theories of emotion (Gellhorn, 1964; Izard, 1971; Laird, 1984; Levenson, Ekman, & Friesen, 1990; Lindquist et al., 2012; Tomkins, 1962; Waynbaum, 1907; Zajonc, Murphy, & Inglehart, 1989). For example, Tomkins (1962, p. 204) argued that the face is the “primary site of the affects”. He suggested that emotional processes create bodily changes, but it is only when we perceive changes in our facial expressions that we become consciously aware of our emotional experience. Ekman (1972) and Levenson et al. (1990) suggested that facial expressions can trigger emotional processes, leading to changes in both autonomic processes and experience. More radically, Zajonc et al. (1989)

[1] Cochrane (2017) discusses the failure-to-replicate, but says most research supports it. Many others mention the failure-to-replicate in passing by adding “but see” citations (Baumeister et al., 2017; Berndt, Dudschig, & Kaup, 2016; Davis, Winkielman, & Coulson, 2017; Lewinski, Fransen, & Tan, 2016). Some say the evidence is controversial (Panasiti, Porciello, & Aglioti, 2017; Prochazkova & Kret, 2017) or inconclusive (Steinmetz & Posten, 2017). Nicholas and Ashton-James (2017) say the effects are, at best, small and difficult to replicate, and, at worst, the byproduct of methodological confounds. Smith and Apicella (2017) conclude that the finding has not survived replication.

suggested that facial actions produce changes in brain-blood temperature, which has hedonic effects through its impact on biochemical activity in the brain. Regardless of their proposed mechanisms, these theories converge on the idea that emotional experience is an embodied process that depends, in part, upon facial feedback.

The notion that facial feedback influences emotional experience also has practical applications. For example, some have suggested that smiling can help manage distress (Ansfield, 2007; Kraft & Pressman, 2012) and improve well-being (Schmitz, 2016). Many have even speculated that facial feedback interventions may reduce depression (Carruthers & Carruthers, 2017; Finzi & Rosenthal, 2014, 2016; Han et al., 2012; Hawlik, Freudenmann, Pinkhardt, Schoenfeldt-Lecuona, & Gahr, 2014; Kruger & Wollmer, 2015; Lin, Hu, & Gong, 2015; Magid et al., 2015; Wollmer et al., 2012; Zariffa, Hitzig, & Popovic, 2014). For instance, five randomized controlled trials have found that Botox injections that inhibit activity in frowning muscles reduce symptoms of major depressive disorder (Finzi, & Rosenthal, 2014; Magid et al., 2015; Wollmer et al., 2012). Given that many theories and interventions have been built on top of the facial feedback edifice, it is pressing to investigate the soundness of its structure.

Picking up the pieces after a failed replication.

If scientists are bricklayers who assemble bricks to build edifices, then meta-analysts who review effect sizes across multiple studies, populations, and methods are building inspectors. Failures-to-replicate in science are increasingly prominent (e.g., Begley & Ellis, 2012; Open Science Collaboration, 2015; Wagenmakers et al., 2016), and when they occur it is easy to become fixated on the individual bricks that have crumbled. However, we should not mistake a brick for an edifice. As Wagenmakers et al. (2016) was careful to point out, the registered replication results “do not invalidate the more general facial feedback hypothesis” (p. 924). This is the case for two reasons. First, single studies often test a narrow operationalization of a phenomenon, but the literature typically includes a wider array of operational definitions. Second, single studies often cannot examine the effects of multiple moderators whose effects may be apparent when the entire literature is examined. Meta-analyses permit a broader perspective on the reliability of effects and moderators of those effects by examining the cumulative evidence of a large, diverse pool of studies.

Chapter 2: Present Meta-Analysis

By combining multiple effects in a literature, meta-analysis can: (1) Provide a more precise estimate of an effect size than individual studies, (2) Determine how much an effect varies and examine moderators that cause this variability, and (3) Diagnose the extent and impact of publication bias. The last meta-analysis on the facial feedback hypothesis was performed 30 years ago and revealed a medium effect size (r = .34) among 16 studies that included 532 participants (Matsumoto, 1987). Two more recent meta-analyses have included facial feedback effects, but did not address the effects of facial feedback separately from other types of behavioral manipulations (e.g., changing breathing rate; Lench, Flores, & Bench, 2011) or included a very small group of studies (s = 8; Westermann, Stahl, & Hesse, 1996). Given the large number of studies that have been published since the last meta-analysis specifically reviewing the facial feedback hypothesis, as well as recent controversies over the reliability of some of the effects, an up-to-date meta-analysis is in order.

To summarize, the present meta-analysis is the first to comprehensively examine the facial feedback hypothesis literature in 30 years. To address inconsistencies and limitations from previous meta-analyses, this meta-analysis will provide a more precise and up-to-date effect size estimate and tests of the impact of publication bias and moderators of facial feedback effects.

Moderators of Interest

Eleven potential moderators of facial feedback effects were identified through a review of the facial feedback literature. Moderator coding was completed by three coders (myself and two trained research assistants) who discussed and resolved discrepancies throughout the coding process. Each moderator is described below, and details regarding their coding criteria are available in Table 1.

Type of affective reaction.

The central tenet of the facial feedback hypothesis is that facial expressions influence emotional experience. However, many researchers have also investigated whether facial expression can influence *perceptions of affective quality*, which are more similar to an evaluative judgment of a stimulus (Hunter, Schellenberg, & Schimmack, 2010; Itkes, Kimchi, Haj-Ali, Shapiro, & Kron, 2017; Russell, 2003). As an example of this distinction, Russell (2003) describes a hypothetical depressed patient who is not cheered up by a beautiful sunset (emotional experience) despite finding it pleasant (perception of affective quality). Interestingly, this rare distinction was made in the original Strack et al. (1988) paper, where participants reported both their emotional experience (i.e., how amused they were by the cartoons) and perception of affective quality (i.e., how funny they thought the cartoons were). They found evidence that facial feedback influenced emotional experience, but no evidence that it influenced perceptions of affective quality.

The facial feedback hypothesis makes no direct predictions about the effects of facial feedback on perceptions of affective quality. However, given the large amount of studies that examined this effect, both types of affective reactions (emotional experience and perception of affective quality) were included in the meta-analysis. Subgroup analyses tested whether facial feedback influences emotional experience and perceptions of affective quality, and moderator analyses illuminate whether the magnitude of these effects differ.

Modulation vs. initiation of emotional experience.

Researchers have produced two hypotheses about the effects of facial feedback on emotional experience. The *modulation hypothesis* posits that facial feedback can only modify emotional experiences that are already ongoing. The *initiation hypothesis* suggests that facial feedback can instigate emotional experiences (for reviews, see Adelmann & Zajonc, 1989;

McIntosh, 1996; Soussignan, 2004). Different studies often test different versions of the facial feedback hypothesis. For example, Larsen, Kasimatis, & Frey (1992) tested whether a furrowed brow modulates how sad people feel after examining aversive photographs. Other studies have tested the initiation hypothesis by manipulating facial activity when no other emotionally evocative stimuli were present (e.g., Flack, 2006). To distinguish between these two types of studies, we coded whether participants were presented with emotionally evocative stimuli. Subgroup analyses tested whether facial feedback modulates and initiates emotional experiences, and moderator analyses illuminated whether the magnitude of these effects differ.

Discrete vs. dimensional measures of emotional experience.

There is ongoing debate about whether emotions are best conceptualized as discrete categories, such as happiness, anger, and sadness (Ekman, 1999; Izard, 2007; Tomkins, 1962), or as experiences that fall along dimensions, such as valence (i.e., degree of positivity vs negativity) and arousal (Russell, 1980; Watson, Clark, & Tellegen, 1988). A similar discrete vs. dimensional distinction exists in the facial feedback literature (Winton, 1986). Some studies focus on the impact of facial feedback on discrete emotions, such as the effects of facially expressing sadness on reports of sadness (e.g., Whissell, 1985). Other studies focus on the effects of facial feedback on positivity or negativity (e.g., Davis, Senghas, Brandt, & Ochsner, 2010). By this account, expressing negative emotions (such as sadness) may not increase feelings of sadness, but will increase negative affect in general. Previous reviews have generally concluded that facial feedback can influence dimensional reports of emotion, but that the evidence for its impact on discrete reports of emotion is preliminary (Adelmann & Zajonc, 1989), mixed (McIntosh, 1996), or controversial (Soussignan, 2004).

In addition to identifying whether effect size was based on a discrete or dimensional emotion measurement, the specific type of emotion measured was coded. For discrete measures, categories included studies that assessed Anger (e.g., Davey et al., 2013), Disgust (e.g., Yartz, 2003), Fear (e.g., Tourangeau & Ellsworth, 1979), Happiness (e.g., Flack, 2006), Sadness (e.g., Larsen et al., 1992), and Surprise (e.g., Reisenzein & Studtmann, 2007). For dimensional measures, categories consisted of studies that assessed Positivity (e.g., Soussignan, 2002) and Negativity (e.g., Duncan & Laird, 1980). Three sets of subgroup/moderator analyses examined the effects of facial feedback on different types of emotion. The first examined whether facial feedback influences discrete and dimensional reports of emotion, and whether the magnitude of these two effects differ. The second more specifically examined whether facial feedback influences reports of Disgust, Fear, Happiness, Sadness, and Surprise, and whether the magnitude of these five effects differ. The third examined whether facial feedback influences reports of Positivity and Negativity, and whether the magnitude of these effects differ.

Between vs. within-subjects design.

An early criticism of the facial feedback literature was that it focused almost exclusively on within-subject designs. In fact, Buck (1980) noted that all studies that found evidence for the facial feedback hypothesis to that point had employed within-subject designs. Since then, researchers have used more between-subject than within-subject designs. Subgroup analyses tested whether facial feedback effects are observed in between- and within-subject comparisons, and moderator analyses illuminated whether the magnitude of these effects differ.

Facial feedback manipulation procedure.

Facial feedback has been manipulated in a variety of ways, including: tasks that incidentally produce facial postures (such as Strack et al.'s 1988 pen-in-mouth technique), experimenter-instructed facial posing (e.g., Tourangeau & Ellsworth, 1979), expression suppression (e.g., Gross, 1998), expression exaggeration (e.g., Demaree et al., 2006) and Botox treatments (e.g., Davis et al., 2010). Methodological differences are a common source of variation in effect sizes, and Izard (1990) speculated that some facial feedback methodologies may produce larger effect sizes than others.

Effect sizes in this meta-analysis represent the magnitude of the difference in a comparison between two groups. Therefore, codes for manipulation procedure had to convey the procedure used in both groups. Consequently, the moderator variable captured both the procedure used in the experimental group and the procedure used in the comparison group (for a similar approach, see Webb, Miles, & Sheeran, 2012). For example, if a study compared the effects of posing a smile to the effects of suppressing a smile, it was coded as “posing-suppression.” This led to a total of 10 different manipulation procedures. Subgroup analyses tested the effect of each individual procedure, and moderator analyses determined whether the magnitude of these effects differed.

Type of stimuli.

Facial feedback experiments that include emotionally evocative stimuli have used a variety of stimuli, including emotional sounds (e.g., Vieillard, Harm, Bigand, 2015), images (e.g., Strack et al., 1988), films (e.g., Soussignan, 2002), imagined scenarios (e.g., McCanne & Anderson, 1987), sentences (e.g., Lewis, 2012), stories (e.g., Paredes, Stavraki, Briñol, & Petty, 2013), and emotional social contexts (e.g., Butler et al, 2003). Subgroup analyses tested the effects associated with each stimulus, and moderator analyses determined whether the magnitude of these effects differed.

Proportion of women.

There are many well-documented gender effects in the emotion literature. For example, researchers have reported gender differences in emotion regulation (Gross & John, 2003; McRae, Ochsner, Mauss, Gabrieli, & Gross, 2008; Nolen-Hoeksema & Aldao, 2011), emotional expressivity (Kring, Smith, & Neale, 1994), and smiling behavior (LaFrance, Hecht & Paluck, 2003). Some researchers have suggested that there may also be gender differences in bodily feedback effects, like facial feedback. For example, Pennebaker and Roberts (1992) suggested that proprioceptive signals may have larger effects on men versus women's emotional experience. If so, women should show smaller facial feedback effects than men. Examining the proportion of women in the sample as a moderator highlights whether there are gender differences in facial feedback effects.

Timing of measurement.

Studies differ in terms of whether the dependent variable (i.e., perception of affective quality, emotional experience) is measured during or after the facial feedback manipulation. For example, Reisenzein and Studtmann (2007) had participants maintain a facial expression until they had completed a measure of emotional experience. In contrast, Duncan and Laird (1980) had participants complete a measure of emotional experience after having completed a posing procedure. Emotions can be fleeting (i.e., Ekman, 1994; Verduyn, Delaveau, Rotgé, Fossati, & Van Mechelen, 2015), so facial feedback effects may be stronger when the dependent measure is measured during the facial feedback manipulation. To test this hypothesis, timing of measurement was included as a moderator.

Awareness of video recording.

In a comment on the failure-to-replicate, Strack (2016) suggested that one reason the results of the original experiment may not have replicated is that cameras were directed at participants in the replication studies. Strack reasoned that awareness of video recording may induce a self-focus that disrupts the flow of experience and suppresses emotional responses. To test this possibility, studies were coded according to whether participants were aware vs. unaware of video recording. Subgroup analyses test whether facial feedback influences affective experience when there is and is not a camera present, and moderator analyses tests whether the magnitude of these effects differ.

Publication year.

The *decline effect* refers to the observation that effect sizes sometimes get smaller over time (Lehrer, 2010). Publication year was included as a moderator to assess whether there is a decline effect in the facial feedback literature.

Publication status.

Publication bias is a well-documented phenomenon in science (e.g., Rothstein, Sutton, & Borenstein, 2006). Publication bias poses a risk to meta-analyses if the unpublished literature differs systematically from the published literature. If published studies have larger effect sizes and are more likely to have significant findings than studies that are not published, then a meta-analysis of only the published studies will yield inflated effect size estimates. Fortunately, a large proportion of the effect sizes in this meta-analysis come from unpublished sources (reviewed later in *Selection of studies*). This moderator was included to test whether published studies had larger effects than unpublished studies in this meta-analysis.

Chapter 3: Materials and Method

All materials for this meta-analysis are available on the Open Science Framework (https://osf.io/v8kxb/?view_only=3beb309cd96e4414a6356fe2dfe97473), including: a) pre-registered analysis plan, b) detailed outline of search strategy, c) list of all screened articles and other reports (e.g., dissertations, unpublished manuscripts) with explanations of exclusions, d) quotes and rationale behind all moderator and effect size coding decisions, e) materials and instructions for an open-source plot extraction tool that was used to extract relevant statistics (e.g., means) that were not reported but were displayed in figures (Rohatgi, 2011), and f) R code to replicate all analyses. After public discussion of a pre-print of this thesis, some minor modifications were made to the pre-registration plan. Materials detailing these modifications are also available on the Open Science Framework.

Scope

For the purposes of this meta-analysis, only effect sizes where the dependent variables matched the facial feedback manipulation were coded. For example, if a researcher manipulated whether participants smiled and collected measures of both happiness and sadness, only the happiness ratings were coded. Although the effects of facial feedback manipulations on non-target emotions would be theoretically interesting to debates about whether specific facial poses have emotion-specific effects (e.g., whether posing sadness can produce sadness, but not other discrete negative emotions), this question fell beyond the scope of this meta-analysis.

Selection of Studies

The literature search strategy was developed in consultation with an experienced librarian at the University of Tennessee. Figure 1 is a PRISMA flowchart that outlines the process for selecting studies for inclusion in the meta-analysis (Moher, Liberati, Tetzlaff, Altman, & Prisma Group, 2009). To gather reports, I first searched:

- PsycInfo: SU.EXACT.EXPLODE("Feedback") AND SU.EXACT("Facial Expressions")
- PsycInfo: expressive suppression AND "emotion regulation"
- Pubmed: feedback[All Fields] AND "facial expressions"[All Fields] OR "facial feedback" OR "facial feedback hypothesis"
- Web of Science: ("feedback" AND "facial expression*" AND emotion) OR ("facial feedback" AND emotion) OR "facial feedback hypothesis"
- References of 17 reviews on the facial feedback hypothesis (Adelmann & Zajonc, 1989, Buck, 1980, Gerrards-Hesse et al., 1994, Izard, 1990, Laird, 1984, Lench et al., 2011, Martin, 1990, Matsumoto, 1987, McIntosh, 1996, Price & Harmon-Jones, 2015, Price, Peterson, & Harmon-Jones, 2012, Soussignan, 2004, Strack, 2016, Wagenmakers et al., 2016, Webb et al., 2012, Westerman et al., 1996, Whissell, 1985)

To capture the unpublished literature, I searched ProQuest Dissertations and Theses Global ("facial feedback hypothesis" AND "emotion"), made three calls for unpublished data (SPSP Open Forum, ResearchGate, Facebook Psychological Methods Discussion Group), and directly requested unpublished data from 81 facial feedback researchers who were identified throughout the screening process.

After removing duplicate records, there were 1,496 records to screen. The titles and abstracts of these records were screened for studies that manipulated facial expressions, measured affective experience, or seemed relevant to the facial feedback hypothesis for any other reason. If there was any doubt about an article's eligibility, it was retained for further review. During this screening, 1,063 full-text reports were excluded, leaving 433 reports to assess for eligibility.

To assess full-text reports for eligibility, the following criteria were used:

a) Facial expressions were manipulated. To provide a clear assessment of the facial feedback hypothesis, studies that simultaneously manipulated facial expressions and other bodily postures were excluded.
b) Measures of perceptions of affective quality or emotional experience were collected. Studies that measured pain were excluded because previous facial feedback and emotion researchers have argued that pain is not a clearly emotional outcome (e.g., Lumley et al., 2011; McIntosh, 1996).
c) Data from non-clinical samples were reported. If a study examining a clinical sample also included data from a non-clinical sample, only the data from the non-clinical sample was included.
d) Information necessary to compute effect sizes was included (reviewed in *Variable Coding*).
e) Article was in English.
f) Article was a primary study that reported data not reported elsewhere.

Based on these criteria we included 98 reports that contained a total of 136 studies. From these 136 studies, 286 effect sizes were extracted.

Variable Coding

Moderator coding was completed by three coders (myself & two trained research assistants) who discussed and resolved discrepancies throughout the coding process (see Table 1 for coding criteria). I extracted all information related to effect size (sample size, means and standard deviation, *t*-value, *F*-value, or *p*-value). If relevant statistics were not included in the report, but informative graphs were included, an open-source program was used to extract data from the graphs (WebPlotDigitzer; Rohatgi, 2011). If a report did not include additional information or graphs but it did indicate there was or was not a significant facial feedback effect, conservative *p*-values of .05 or .50, respectively, were assumed in the effect size calculations. If the sample size for each condition was not reported in a study with between-subject comparisons, sample size was estimated by dividing the total sample by the number of conditions.

Meta-Analytic Approach

Effect size index.

Cohen's standardized *d* was used as the effect size index, which represents the difference between two group means divided by their pooled standard deviation (Cohen, 1987). Effect sizes were calculated in R 3.4.0 (R Core Team, 2017) using the formulas provided by Borenstein (2009). Effect sizes were calculated so that positive values always indicated an effect consistent with the facial feedback hypothesis. For example, the facial feedback hypothesis predicts that facilitating facial expressions leads to increased emotional intensity, whereas suppressing facial expressions leads to decreased emotional intensity. Therefore, increased emotional intensity in a facilitative condition (e.g., Flack, 2006) and decreased emotional intensity in a suppression condition (e.g., Gross & Levenson, 1997) both represent predicted facial feedback effects and were coded in the positive direction.

For within-subject designs, the correlation between the pre- and post- measures is necessary for calculating Cohen's *d*. Unfortunately, this correlation is rarely reported, so it is recommended to assume a correlation and perform a sensitivity analysis on the assumed value (Borenstein, 2009). A default correlation of .50 was pre-registered, but sensitivity analyses were performed to determine the impact of the assumed correlation on the overall effect size estimate (testing r = .10, .30, .50, .70, 90). These sensitivity analyses did not affect the inferences made about the overall effect size, so only the analyses that used the default r = .50 value are reported here.

Meta-analysis with robust variance estimates.

Fifty-three percent of studies provided multiple effect sizes of interest. For example, Flack, Laird, and Cavallaro (1999b) examined the impact of angry, sad, fearful, and happy facial expressions on emotional experience. When a study provides multiple effect size estimates, it is best to record all effect sizes in order to be comprehensive. However, one drawback of this approach is that it violates the statistical assumption that effect sizes are independent. There are several ways to deal with dependency in meta-analysis. The simplest approach is to aggregate effect sizes drawn from the same study (Borenstein et al., 2009; Rosenthal & Rubin, 1986). Although this removes dependency, it results in a loss of information regarding comparisons among multiple levels of a moderator in a single study. A second approach is to use multivariate meta-regression (Raudenbush, Becker, & Kalaian, 1988). However, this approach requires knowledge of the underlying covariation structure among effect sizes, which is almost always unknown. A third, more recent approach is to use meta-analysis with robust variance estimates (RVE) (Hedges, Tipton, & Johnson, 2010). Similar to its application in general linear models, RVE can be used in meta-analysis to adjust for dependencies among effect sizes. This approach does not result in the loss of any information, does not require knowledge of the underlying correlation structure, and can accommodate multiple sources of dependencies. This RVE approach was used to estimate the overall effect size, and conduct subgroup and moderator analyses.[2]

Meta-analysis with RVE weighting scheme.

When averaging the results of multiple studies, meta-analyses typically give more weight to effect sizes with more precision (via a procedure termed inverse-variance weighting). Meta-analysis with robust variance estimates uses similar weighting schemes that provide adjustments for the types of dependency among effect sizes (Tanner-Smith & Tipton, 2014). If dependency primarily arises from studies providing multiple effect sizes for the same outcome of interest, the correlated effects weighting scheme is recommended. On the other hand, if dependency primarily arises from authors reporting multiple studies, the hierarchical effects weighting scheme is recommended (Hedges et al., 2010). In practice, both types of dependencies often exist in a meta-analysis, and it is recommended to choose weighting based on the predominate type of dependency (Tanner-Smith & Tipton, 2014). Twenty-one percent of the reports in the present meta-analysis included multiple studies and 53% of the reports included studies that provided multiple effect sizes for the outcome of interest. Therefore, the correlated effect weighting scheme was used.

When calculating weights, meta-analysis with RVE requires an estimate of the within-study effect-size correlation (i.e., the average correlation among the dependent effect sizes). The default assumed value is r = .80. This default value was pre-registered, but additional sensitivity

[2] Although I suggest that meta-analysis with RVE was the best data analysis approach, the overall effect size was also calculated using the Borenstein et al. (2009) aggregation method for correcting for dependencies as well as a more-traditional random-effects meta-regression model that does not adjust for dependencies. The overall effect sizes in these two models were nearly precisely identical to those obtained using the RVE approach. Therefore, only the results of the RVE approach are reported.

analyses were performed to determine the impact of this assumed value on the overall effect size estimate (testing $r = 0, .20, .40, .60, .80, 1.00$). These sensitivity analyses produced nearly identical overall effect size estimates, so all further analyses used the default value of $r = .80$.

Overall effects, subgroup, and moderator analyses.

To test the overall effect size, an intercept-only meta-regression model with RVE was fit using the R package, robumeta (Fisher & Tipton, 2015). The intercept of this model can be interpreted as the precision-weighted overall effect size, adjusted for correlated-effect dependencies. The same approach was used to calculate mean effect sizes for each level of each moderator (i.e., subgroup analyses). For a few cases with too few observations for the RVE approach, their mean effect size was calculated using a traditional random-effects meta-regression model. These exceptions are noted in Table 2.

The RVE approach was also used to perform separate hypothesis tests for the effects of each moderator [3]. Continuous moderators were entered into a meta-regression equation without transformation, except publication year, which was centered at 2017 to ease interpretation of the regression intercept. Categorical moderators with two levels (i.e., type of experience) were dummy coded and entered into separate meta-regression equations. The significance test corresponding to the regression coefficient for the predictor variable in these models can be interpreted as a test of whether the variable is a significant moderator.

Examining categorical moderators with more than two levels required an additional step. Like the former process, they were first dummy coded and entered into meta-regression

[3] The pre-registration plan notes that important theoretical moderators would be re-examined with significant methodological moderators included as covariates. These analyses did not change any conclusions, so they are not reported here.

equations. However, the regression coefficients only test whether there is a difference between a single level of a moderator and a single comparison level. To perform an omnibus test of moderators with more than two levels, I followed the recommendations of Tanner-Smith, Tipton, and Polanin (2016) and conduced Approximate Hotelling-Zhang with small sample correction tests using the clubSandwhich R package (Pustejovsky, 2016). This test produces an *F*-value that indicates whether there is a difference among all levels of the moderator[4].

Outlier detection.

Methods for identifying outliers for meta-regression models with RVE are not yet available, so outliers were instead identified using a random-effects intercept-only meta-regression model. This was done using the base R function, influence.measures. After fitting an intercept-only meta-regression model, this function calculates a variety of influential outlier diagnostics (e.g., covariance ratios, Cook's distances, and diagonal elements of the hat matrix), and identifies cases that are influential on any one of the tests.

Examining publication bias.

Many methods for testing the extent and impact of publication bias in meta-analysis have been developed. Unfortunately, most of these methods were developed and tested under the assumption that the effect sizes are independent, which is unrealistic in meta-analyses of the psychology literature. Below are the two approaches used to examine publication bias with dependent effect sizes in the present meta-analysis.

[4] The Approximate Hotelling-Zhang produces atypical degrees of freedom. See to Tanner-Smith, Tipton, & Polanin (2016) for an assessable explanation.

Publication bias analyses on aggregated dependent effect sizes.

The most common way to assess publication bias with dependent structures is to aggregate the dependent effect sizes and perform standard publication bias tests on the aggregated effect size estimates. To aggregate dependent effect sizes, the R package, MAd, was used (Del Re & Hoyt, 2014). Using the Borenstein et al. (2009) aggregation method, this function calculates aggregated estimates of effect size and effect size variance by taking into account a pre-specified correlation among the clusters of dependent effect sizes (set, by default, at $r = .50$[5]).

These aggregated effect size estimates were used to examine the funnel plot distribution of effect sizes, calculate the fail-safe N, and perform three statistical tests of publication bias: trim-and fill, weight-function modeling, and PET-PEESE (Duval and Tweedie, 2000; Rosenberg, 2005; Stanley & Doucouliagos, 2014; Vevea & Hedges, 1995).

PET-PEESE with robust variance estimates.

Although researchers typically aggregate dependent effect sizes before examining publication bias, the PET-PEESE approach can be conducted using RVE. Because PET-PEESE is essentially a meta-regression equation with standard error or variance as a predictor, RVE can easily be used when fitting the meta-regression model. Compared to the aggregation method, the benefit of this approach is that it does not require an assumed correlation among the clusters of dependent effect sizes.

---

[5] When the correlation among clusters of dependent effect sizes is unknown, it is recommended that meta-analysts assume a correlation and perform additional sensitivity analyses on this assumed value (Borenstein, 2009). In line with this recommendation, we assumed a default correlation of $r = .5$ and performed sensitivity analyses to determine impact of the assumed correlation on our tests of publication bias (testing $r = .10, .30, .50, .70, 90$). We indicate in the manuscript the one instance where this affected our conclusions.

Publication bias sensitivity analyses.

Heterogeneity, which represents how much variation is observed beyond what would be expected from sampling error alone, can pose problems for of publication bias tests (Sterne et al., 2011; Terrin, Schmid, Lau, & Olkin, 2003; Stanley, 2017). Therefore, pre-planned sensitivity publication bias analyses were performed by splitting the dataset by significant moderators.

In instances where no significant evidence of publication bias was uncovered, additional pre-planned sensitivity analyses were conducted by re-running the analyses: 1) excluding suppression studies; 2) excluding Wagenmakers et al. (2016), and 3) excluding Wagenmakers et al. (2016) and all unpublished data. The purpose of these sensitivity analyses was to ensure that publication bias was not masked by subsets of studies that might skew the distribution of effect sizes. For example, the emotion regulation literature suggests that suppression is a relatively ineffective way of managing emotional experience (e.g., Gross, 1998). Therefore, it is feasible that publication bias could be masked by the inclusion of relatively small effect sizes from suppression studies. By this same logic, replication and unpublished studies could have similar effects on publication bias analyses. These sensitivity analyses never affected our conclusions, but they are reported to demonstrate the robustness of the publication bias results.

Chapter 4: Results

The meta-analysis included 98 articles, 136 studies, and 286 effect sizes. Notably, 26% of these came from unpublished sources.

Overall Effect

Using meta-regression with RVE, the overall size of the effect of facial feedback on self-reported affective experience (i.e., emotional experience and perceptions of affective quality) was $d = 0.20$, 95% CI [0.14, 0.26], $t(135) = 6.43$, $p = .000000002$. This indicates that, overall, facial feedback had a small effect on affective reactions in the studies included in the meta-analysis.

Outlier detection.

The base R function, influence.measures, was used to identify influential outliers. This method detected eight influential outliers[6], two of which were in the negative direction. Removing the eight outliers did not affect the overall effect size estimate (adjusted $d = 0.19$, 95% CI [0.13, 0.24], $t(135) = 6.23$, $p = .000000006$) or any of the overall publication bias results reported below. Therefore, all effect size estimates were retained in all further analyses.

Moderator Analyses

There was a large amount of heterogeneity in the effect sizes ($T^2 = 0.11$, $I^2 = 75.6$). Such heterogeneity suggests that there are meaningful differences among studies that can be further explored through moderator analyses. Table 2 contains effect size estimates for each subgroup of each moderator, and the relevant moderator analyses are reported below.

[6] Single influential outliers were detected in Flack, Laird, & Cavallaro (1999), Kalokerinos, Greenaway, & Denson (2015), and Kircher et al., (2012). Five influential outliers were detected in McCanne & Anderson (1987).

Type of affective reaction.

Although the facial feedback hypothesis is primarily concerned with the effects of facial expressions on emotional experience, many researchers have extended this phenomenon to examine facial feedback effects on perceptions of affective quality. Subgroup analyses suggested that facial feedback has a significant effect on both emotional experience (d = 0.17, 95% CI [0.11, 0.23], p = .0000004) and perceptions of affective quality (d = 0.38, 95% CI [0.19, 0.57], p = .0004) (Table 2), and moderator analysis suggested that the facial feedback effects were larger for perceptions of affective quality than emotional experience, β_1 = 0.23, 95% CI [0.03, 0.43], p = .02.

Modulation vs. initiation of emotional experience.

Researchers have long debated whether facial feedback can only modulate emotional experiences that are already unfolding, versus initiate emotional experiences (for reviews see Adelmann & Zajonc, 1989; McIntosh, 1996; Soussignan, 2004). Our results suggested that can both initiate (d = 0.32, 95% CI [0.15, 0.49], p = .0005) and modulate (d = 0.13, 95% CI [0.07, 0.19], p = .00007) emotional experience, and that the effect size for initiating emotion was larger than for modulating emotion, β_1 = 0.19, 95% CI [0.01, 0.37], p = .04.

Discrete vs. dimensional measures of emotional experience.

Studies have assessed the impact of facial feedback on discrete emotions (Whissell, 1985) and on general positivity/negativity dimensions (Winton, 1986). Our results uncovered no evidence of differences in the magnitude of the effects of facial feedback on specific emotions (d = 0.19, 95% CI [0.09, 0.29], p = .0003) versus general positivity/negativity (d = 0.14, 95% CI [0.06, 0.21], p = .0005), β_1 = 0.05, 95% CI [-0.08, 0.18], p = .43.

For studies in which discrete emotions were measured, follow-up analyses tested whether different emotions yielded different effect sizes. This analysis uncovered no evidence that specific discrete emotion was a moderator of facial feedback effects, $F(6.42) = 0.77$, $p = .60$[7]. As shown in Table 2, facial feedback had small-to-medium effects on self-reports of happiness (d = 0.23, 95% CI [0.08, 0.37], p = .004), sadness (d = 0.30, 95% CI [0.06, 0.55], p = .02), anger (d = 0.53, 95% CI [0.19, 0.87], p = .006), and disgust (d = 0.29, 95% CI [0.03, 0.56], p = .03). However, the effect sizes for fear (d = 0.13, 95% CI [-0.05, 0.30], p = .13) and surprise (d = -0.31, 95% CI [-5.57, 4.95], p = .59) did not statistically differ from zero, although these estimates are based on relatively few effect sizes ($k_{\text{fear}} = 15$; $k_{\text{surprise}} = 9$).

For studies that included dimensional measures of emotion, follow-up analyses tested whether effect sizes differed based on whether the target dimension was positivity or negativity. The effect sizes were small for both positivity (d = 0.18, 95% CI [0.07, 0.29], p = .002) and negativity (d = -0.12, 95% CI [0.01, 0.22], p = .03), and the magnitude of these effects did not differ, $\beta_1 = 0.04$, 95% CI [-0.12, 0.19], p = .63.

Between vs. within-subjects design.

There were early concerns that facial feedback effects may not emerge in between-subject comparisons (Buck, 1980). There was little evidence that between-subject designs (d = 0.15, 95% CI [0.07, 0.23], p = .0002) yielded effect sizes smaller than within-subject designs (d = 0.25, 95% CI [0.16, 0.35], p = .000001), $\beta_1 = 0.10$, 95% CI [-0.02, 0.22], p = .12.

[7] Reminder: F-values are based on an Approximate Hotelling-Zhang test with small sample correction (Pustejovsky, 2016). Even though this analysis has 285 effect sizes, the degrees of freedom are low because some of the levels of this moderator had a small number of effect sizes in it (see Tanner-Smith, Tipton, & Polanin, 2016 for more information on degrees of freedom).

Facial feedback manipulation procedure.

Determining whether some facial feedback manipulations have stronger effects on affective reactions than others is complicated by the fact that some researchers use one manipulation compared to a control group (e.g., Stel, van den Huevel, & Smeets, 2008 had participants hold a marker in their teeth or in their hand), the same type of manipulation in both conditions (e.g., Dimberg & Söderkvist, 2011 had participants smile or frown using similar manipulations), or two different manipulations in the two different conditions (e.g., Larsen et al., 1992 had participants frown or suppress frowning). To provide the clearest test of whether there are differences in effect sizes among facial feedback manipulations, we limited our analyses to studies featuring a true control condition (e.g., Stel et al., 2008), although we found similar results when we did limit these analyses[8]. Though effect sizes for different manipulation procedures varied from d = -.04 (exaggeration) to d = .71 (Botox), we did not find evidence that manipulation procedure was a significant moderator of facial feedback effects, $F(10.4) = 0.62$, $p = .66$. Furthermore, although all methodologies except expression exaggeration had a small to large effect, only suppression had a statistically significant effect (Table 2).

Type of stimuli.

Facial feedback experiments that include the presentation of emotional stimuli have used a variety of different stimuli. Results indicated that there were differences in the magnitude of facial feedback effects based on the type of stimulus used, $F(2.77) = 92.1$, $p < .005$. When examined in isolation, the only stimuli that produced statistically significant effects were films,

[8] Although we believe that comparing cases where the experiment had a control group that received no facial feedback manipulation provides the clearest test of whether procedure is a significant moderator, we obtained similar results when we ran the analyses including studies that did not include a control group ($F(14.38) = 1.78$, $p = .16$).

pictures, and sentences (emotional audio [d = 0.72, 95% CI [-0.82, 2.27], p = .18]; pictures [d = 0.16, 95% CI [0.08, 0.23], p = .0002]; films [d = 0.13, 95% CI [0.03, 0.22], p = .009]; imagined scenarios [d = 1.28, 95% CI [-0.98, 3.53], p = .27]; sentences [d = 0.70, 95% CI [0.43, 0.96], p = .0000003]; stories [d = 0.41, 95% CI [-0.29, 1.10], p = .25]; social contexts [d = -0.14, 95% CI [-0.74, 0.46], p = .61]) (see Table 2).

Proportion of women.

Given gender differences in other emotion effects (Gross & John, 2003; Kring et al., 1994; LaFrance et al., 2003; McRae et al., 2008; Nolen-Hoeksema & Aldao, 2011) and proposed gender differences in embodied effects (Pennebaker & Roberts, 1992), proportion of women was examined as a potential moderator of facial feedback effects. Contrary to the proposition that proprioceptive signals may influence women's emotional experience less so than men's, the results indicated that larger proportions of women tended to have *larger* effect sizes, but that this difference was not significant, β_1 = 0.16, 95% CI [-0.09, 0.42], p = .21.

Timing of measurement.

There are inconsistencies regarding whether experimenters collect self-reports of emotional experience during (d = 0.18, 95% CI [0.09, 0.26], p = .0001) or after the facial feedback manipulation (d = 0.22, 95% CI [0.12, 0.33], p = .00008). Results provided no evidence that this methodological difference influences the magnitude of facial feedback effects, β_1 = -0.03, 95% CI [-0.17, 0.11], p = .65.

Awareness of video recording.

In reply to the failed replication attempt, Strack (2016) suggested that one reason the results of the original experiment may not have replicated is that there was a camera directed at participants in the replication study. Across all studies included in this review, there was little evidence that this methodological difference was associated with different facial feedback effects, $\beta_1 = -0.06$, 95% CI [-0.20, 0.07], $p = .36$. Facial feedback effects were small both when participants were aware ($d = 0.17$, 95% CI [0.06, 0.28], $p = .003$) and unaware of video recording ($d = 0.23$, 95% CI [0.15, 0.32], $p = .0000007$).

Publication year.

Results indicated that effect sizes in the facial feedback literature tend to become smaller over time of publication (i.e., that effect sizes increased with distance from 2017), $\beta_1 = -0.006$, 95% CI [-0.01, -0.0006], $p = .03$. When controlling for publication year, the overall effect of facial feedback on affective reactions is smaller, but still significant, $d = 0.14$, 95% CI [0.05, 0.22], $t(131) = 3.23$, $p < .002$. However, exploratory follow-up analyses suggest that the relationship between publication year and observed effect sizes may be driven by the 17 studies included in Wagenmakers et al.'s (2016) registered replication. When these studies are removed, the relationship between publication year and observed effect sizes is no longer significant, $\beta_1 = -0.003$, 95% CI [-0.008, 0.003], $p = .37$.

Publication status.

A common concern is that effect sizes in the published literature are larger than those in the unpublished literature. Twenty-six percent of effect size estimates in this meta-analysis came from unpublished sources, but the magnitude of effect sizes were not significantly smaller for unpublished studies ($d = 0.15$, 95% CI [0.04, 0.26], $p = .01$) than it was for published studies ($d = 0.21$, 95% CI [0.14, 0.28], $p = .00000003$), $\beta_1 = -0.05$, 95% CI [-0.18, 0.08], $p = .43$. This analysis cannot rule out the possibility that there is a large unpublished literature that is not represented in the meta-analysis, but it does not support the proposition that a file-drawer would change the reported effect.

Publication Bias

Two methods were used to address the problem of publication bias more directly.

Overall publication bias analyses with aggregated dependent effect sizes.

The first approach was to aggregate our dependent effect sizes and use these aggregated dependent effect sizes to examine the funnel plot distribution of effect sizes, calculate fail-safe N, and perform three statistical tests of publication bias: trim-and-fill, PET-PEESE, and weight-function modeling.

To visually assess the possibility of publication bias, the aggregated effect size estimates were plotted against their standard errors in a funnel plot. In the absence of publication bias, this pattern should resemble a funnel, where effect size estimates with smaller standard errors cluster around the mean effect size, and effect size estimates with larger standard errors fan out in both directions. A typical pattern suggestive of publication bias is asymmetry in the bottom of the distribution. As can be seen in Figure 2, there was no asymmetry in the overall funnel plot of the aggregated effect sizes.

To further assess the possibility of publication bias, three statistical tests of publication bias were conducted. First, Duval and Tweedie's (2000) trim-and-fill technique was used. This method trims the values of extreme observations that lead to asymmetry in the funnel plot distribution and imputes values to even out the distribution. This technique was not able to impute any missing studies in the data (i.e., did not provide significant evidence of publication bias). Second, PET-PEESE models were fit (Stanley & Doucouliagos, 2014). PET-PEESE models estimate publication bias by calculating the relationship between effect size and variability, and controlling for this relationship in a meta-regression model. Both the PET and PEESE models failed to uncover significant evidence of publication bias, PET $\beta_1 = 0.61$, $p = .18$;

PEESE $\beta_1 = 1.51$, $p = .16$[9]. Last, Vevea and Hedges' (1995) weight-function modeling was used. This method creates a meta-analytic model that is adjusted for publication bias at predefined intervals (in our case, using default intervals that distinguish between significance and non-significance), and compares its fit to an unadjusted model. If an increase in fit is observed, publication bias is a concern. Results indicated that the model adjusted for publication bias did not increase model fit, which provides no evidence of publication bias, $\chi^2(1) = 0.05$, $p = .84$.

Last, the Rosenberg Fail-safe N metric was calculated, which provides an estimate of how much undetected publication bias would have to exist in order to fail to find a significant effect (i.e., make $p > .05$). Although these metrics rely on unrealistic assumptions in their calculations (Ioannidis, 2008), the value is reported because of its popularity and simplicity: Rosenberg's Fail-Safe N = 5,437 studies.

Overall publication bias analyses with robust variance estimates.

The second approach for examining publication bias was to re-examine PET-PEESE using robust variance estimates to adjust for dependency instead of aggregation. Compared to the aggregation method, the benefit of this approach is that it does not require an assumed correlation among the clusters of dependent effect sizes. Contrary to the results produced by the aggregation method, the results of both the PET and PEESE models with robust variance estimates uncovered evidence of publication bias, PETrve $\beta_1 = 1.07$, $p = .03$; PEESErve $\beta_1 = 2.25$, $p = .02$. Furthermore, after controlling for publication bias, the estimate of the overall effect

---

[9] $r = .50$ was pre-registered as the assumed correlation among effect sizes in the aggregation of dependent effect sizes. Similar results were obtained in sensitivity analyses where $r = .10$, .30, and .70. At $r = .90$, marginal evidence of publication bias was uncovered in the PET ($\beta_1 = 0.84$, $p = .05$) and PEESE ($\beta_1 = 1.77$, $p = .06$) models, but not in the trim-and-fill (no imputed missing studies) and weight-function modeling ($\chi^2(1) = 0.39$, $p = .53$) approaches.

size did not significantly differ from zero, PETrve $d = -0.03$, $p = .79$; PEESErve $d = 0.09$, $p = .09$.

Summary of overall publication bias analyses.

Different approaches for assessing publication bias in the overall facial feedback literature led to different conclusions. When the dependent effect sizes were aggregated, no significant evidence of publication bias was uncovered. However, when PET-PEESE analyses were re-run using robust variance estimates instead of aggregation to adjust for dependency, significant evidence of publication bias was uncovered. Future research will shed light on which approach is superior. In the meantime, it is possible that publication bias exists in the overall facial feedback literature.

Publication bias sensitivity analyses.

As noted above, there is a large degree of heterogeneity in the overall size of facial feedback effects, $T^2 = 0.11$, $I^2 = 75.6$. This heterogeneity can pose problems for many tests of publication bias (Sterne et al., 2011; Terrin et al., 2003; Stanley, 2017), and suggests that it may be more fruitful to examine publication bias on individual levels of significant moderators. The present meta-analysis uncovered three significant moderators: (1) type of affective reaction (emotional experience or perception of affective quality), (2) whether facial feedback initiates or modulates emotional experience, and (3) the type of stimuli used in the experiment. In line with the pre-registration plan, all publication bias analyses were re-run on individual levels of these significant moderators. No evidence of publication bias was uncovered when analyses were split by the initiation vs. modulation or stimulus type moderators, but evidence of publication bias was uncovered when the analyses were split by type of affective reaction.

There was evidence of publication bias in studies that examined the effects of facial feedback on perceptions of affective quality. As shown in the top panel of Figure 3, the funnel plot is largely asymmetrical. The trim-and-fill method imputed 5 missing observations, but suggested that the adjusted overall effect was still significant (adjusted $d = 0.25$, 95% CI [0.06, 0.44], $p = .01$). The PET and PEESE models from both the aggregation and RVE approach provided evidence of publication bias (PET $\beta_1 = 2.65$, $p = .03$; PEESE $\beta_1 = 5.05$, $p = .048$; PETrve $\beta_1 = 2.28$, $p = .01$; PEESErve $\beta_1 = 3.41$, $p = .04$), and suggested that the adjusted overall effect was not significant (PET $d = -0.21$, $p = .36$; PEESE $d = 0.08$, $p = .52$; PETrve $d = -0.17$, $p = .49$; PEESErve $d = 0.16$, $p = .28$). The weight-function model also provided marginal evidence for publication bias, $\chi^2(1) = 3.17$, $p = .07$.

No evidence of publication bias was uncovered in the studies examining the effects of facial feedback on emotional experience. As shown in the bottom panel of Figure 3, the funnel plot of effect sizes appeared symmetrical. Furthermore, the trim-and-fill method imputed no missing studies, PET and PEESE models from both the aggregation and RVE approach provided non-significant estimates of publication bias (PET $\beta_1 = 0.15$, $p = .30$; PEESE $\beta_1 = 0.47$, $p = .40$; PETrve $\beta_1 = 0.70$, $p = .20$; PEESErve $\beta_1 = 1.75$, $p = .13$), and the Vevea and Hedges' (1995) weight-function modeling found that the meta-analytic model that is adjusted for publication bias did not provide better fit than a non-adjusted model ($\chi^2(1) = 1.02$, $p = .31$). Since significant evidence of publication bias was not uncovered in the studies, additional pre-planned sensitivity analyses were conducted. More specifically, publication bias tests were re-run: 1) excluding suppression studies, 2) excluding Wagenmakers et al. (2016), and 3) excluding Wagenmakers et al. (2016) and all unpublished data. None of these sensitivity analyses suggested the presence of publication bias in studies that examined the effects of facial feedback on emotional experience.

Chapter 5: Discussion

Was Thích Nhất Hạnh correct when he suggested that our smiles can be the source of our joy? Can our facial expressions influence our emotional reactions? To answer these questions, I performed a comprehensive meta-analysis of the facial feedback hypothesis literature, combining nearly 300 effect sizes generated from 136 studies conducted over the past 50 years.

Results support the facial feedback hypothesis' central claim that facial expression influence emotional experience. However, the results do not support the extended claim that facial expressions influence perceptions of affective quality. Although the analyses initially indicated that facial expressions had a medium effect on perceptions of affective quality, publication bias tests uncovered clear evidence of bias in this subset of studies. When adjusting for publication bias, there is not clear evidence that facial feedback influences perceptions of affective quality.

The effects of facial feedback on emotional experience tend to be quite small (d = .17), but this is not to say that there are not instances in which the effects are larger. In fact, approximately 70% of variation in facial feedback effect sizes is due to heterogeneity (i.e., variation beyond what would expect from sampling error alone), which suggests that the effects are larger under some conditions than others. Below, I review the significant moderators that were uncovered, which help illuminate the conditions in which facial feedback effects may be larger.

Facial Feedback Effects Are Larger for Initiating than Modulating Emotional Experience

Researchers have long debated whether facial feedback can only *modulate* emotional experiences that are already unfolding, or if facial feedback can also *initiate* emotional experiences in otherwise non-emotional circumstances (for reviews see Adelmann & Zajonc, 1989; McIntosh, 1996; Soussignan, 2004). The cumulative evidence suggest that it can do both. Furthermore, effect sizes were larger in studies that focused on initiating emotional responses compared to studies that focused on modulating emotional responses. The tendency for people to have different appraisals of the same evocative stimulus may help explain this pattern (Ellsworth & Scherer, 2003). In the context of an evocative stimulus, variance in how the stimulus is appraised may produce enough variance in emotional experience to make facial feedback effects indiscernible. In the absence of evocative stimuli, however, facial feedback effects may be more discernible.

Facial Feedback Effects Are Larger for Certain Stimuli.

Facial feedback experiments that included evocative stimuli have used a variety of stimuli. Results indicated that stimulus type is actually the largest moderator of facial feedback effects observed in the present meta-analysis. Furthermore, there are also large amounts of heterogeneity within different stimulus types, suggesting that even among a group of studies using similar types of stimuli (e.g., images), methodological differences (e.g., different images; different presentation modes) may create differences in the magnitude of facial feedback effects.

Facial Feedback May Not Influence All Types of Emotions

Facial feedback influenced reported emotion regardless of whether dimensional or discrete measures of emotion were used. Furthermore, type of emotion measured was not a significant moderator. However, while an examination of the associated effect sizes revealed a small-to-medium effect for all emotions, these effects did not differ from zero for fear and surprise. These findings suggest that facial feedback may not influence people’s reports of experienced fear and surprise (see Reisenzein, Horstmann, & Schutzwohl, 2017 for a similar conclusion). Appraisal models of emotion provide one explanation for these findings. Some appraisal models suggest that one unique characteristic of fear and surprise is that an event is appraised as unexpected and sudden (Smith & Ellsworth, 1985). One possibility is that facial feedback has larger effects on some types of appraisals, such as valence, than appraisals of unexpectedness or suddenness. Since only a few studies have examined the effects of facial feedback on fear or surprise, reliable conclusions about the size of the effect cannot be reached. However, these preliminary results are worth examining further.

Other Potential Sources of Heterogeneity

Several other moderators proposed by previous facial feedback researchers were examined, including whether effect sizes came from between- or within-subject comparisons (Buck, 1980), the procedure used to manipulate facial poses (Izard, 1990), the proportion of women in the sample (Pennebaker & Roberts, 1992), and whether participants were aware of video recording (Strack, 2016). In addition, I tested moderators I hypothesized would influence facial feedback effects, such as the timing of self-reported affective experience. However, there was no significant evidence that these factors were associated with differences in the magnitude of facial feedback effects.

Although moderators that figured prominently in the literature were examined, given the large degree of heterogeneity in facial feedback effects there are likely to be moderators of facial feedback effects that were not initially considered. A few of these possibilities below are highlighted below.

Internal vs. external focus.

Facial feedback effects may be moderated by the degree to which participants focus on internal vs. external emotional cues. So far, only a few studies have investigated this moderator, and these studies have solely focused on individual differences. Laird and Crosby (1974) suggested that people who rely on internal cues (e.g., their smiles) exhibit stronger facial feedback effects than those who rely on external cues (e.g., the funniness of a cartoon). Similarly, Dzokoto et al. (2014) suggested that individuals who focus more on their cognitive evaluation of a situation (i.e., external cues) may exhibit weaker facial feedback effects than those who focus on these evaluations less. However, both lines of research used problematic operationalizations of the construct. Laird and colleagues used a circular line of reasoning in that they classified participants who demonstrated smaller facial feedback effects as individuals who rely more on situational cues, and participants who demonstrated larger facial feedback effects as individuals who rely more on self-produced cues. On the other hand, Dzokoto et al. (2014) used the Attention to Emotion subscale of the Trait-Meta-Mood Scale (Salovey et al., 1995) to index the degree to which participants focus on external emotional cues. However, this subscale includes items such as "*Feelings give directions to life*" and "*I often think about my feelings,*" which may also index the degree to which participants are sensitive to internal emotional cues. Therefore, their results, although interesting in their own right, do not shed light on the internal vs. external focus moderator. Future researchers interested in this individual difference should consider using measures that more directly assess the construct, such as the Rational-Experiential Inventory (Pacini & Epstein, 1999).

Although the internal vs. external focus has been primarily investigated as an individual difference, this potential moderator could also vary situationally. Future research can experimentally manipulate the degree to which participants focus on internal vs. external cues by instructing some participants to focus on their intuition and others to focus on the situation. Furthermore, researchers can use emotion measures that limit the degree to which participants reflect on the emotional nature of the situation, such as the as the Affect Misattribution Procedure (Payne, Cheng, Govorun, & Stewart, 2005) or the Implicit Positive and Negative

Affect Test (Qurin, Kazén, & Kuhl, 2009). Alternatively, researchers can limit the influence of external cues by manipulate facial feedback in scenarios without emotional cues (i.e., test the initiating function of facial feedback).

Exclusion criteria.

Exclusion criteria may be especially important source of heterogeneity in facial feedback research because researchers use diverse sets of exclusion criteria to sometimes exclude large proportions of participants. Approximately half of the studies in the meta-analysis did not use any exclusion criteria, and studies that did exclude participants used a wide range of exclusion criteria. For example, researchers excluded participants who were aware of the purpose of the experiment (e.g., Baumeister, Papa, & Foroni, 2016 Duncan & Laird, 1971; Laird, 1974), failed an attention-check (e.g., Kalokerinos, Greenaway, & Denson, 2014), experienced equipment errors (e.g., Pedder et al., 2016), produced unreadable or missing data (e.g., Demaree et al., 2006; Dzokoto et al., 2014; Zajonc, Murphy, & Inglehart, 1989), or were outliers (e.g., Korb et al., 2012; Marmolejo-Ramos & Dunn, 2013; Zhu et al., 2015). Exclusion criteria choice might be especially important in the facial feedback literature given the large proportions of participants that are sometimes excluded. For example, Soussignan (2002) excluded approximately 30% of participants because the facial feedback manipulation did not activate the intended action units in the face. Wagenmakers et al. (2016) used a combination of several exclusion criteria and, on average, excluded 25% of their participants. These various exclusion criteria have the potential to both deflate (e.g., excluding participants who were aware of hypothesis) and inflate (e.g., excluding participants who failed manipulation checks) observed effect sizes, which further contributes to heterogeneity in the facial feedback literature.

Intensity of posed expressions.

If smiling can make us happy, it would follow that intense smiles make us happier than mild smiles. Consistent with this reasoning, Soussignan (2002) found that participants manipulated to pose more intense Duchenne smiles exhibited larger facial feedback effects than participants manipulated to pose less intense, non-Duchenne smiles. Similarly, Kraft & Pressman (2012) found that participants manipulated to pose Duchenne smiles had quicker cardiovascular recovery to a stressful event than participants manipulated to pose non-Duchenne smiles. The available evidence suggests that larger effects could be obtained with more intense facial expression manipulations. However, no researchers have investigated whether the intensity of other types of facial expressions (e.g., sad or angry poses) moderates facial feedback effects.

In addition to having implications for facial feedback theory, this potential moderator deserves more attention because researchers may inadvertently use different intensity manipulations, which can lead to inconsistencies in effect size estimates. Future research can examine this issue by measuring the intensity of posed facial expressions. Research to date (Soussignan, 2002; Kraft & Pressman, 2012) has used coders trained in the Facial Action Coding System (Ekman & Friesen, 1978), but researchers can also consider naïve coders, less extensive training programs (e.g., Levenson, 2005), or even commercially available automated facial coding software.

Weighing in on Strack et al. (1988) and Wagenmakers et al. (2016)

Although Wagenmakers et al. (2016) was unable to replicate the pen-in-mouth effect published by Strack et al. (1988) we should not to mistake a brick for an edifice. In this section, I first review the pen-in-mouth brick and then evaluate its position in the facial feedback edifice.

Strack et al.'s (1988) seminal studies provided evidence for three main findings: 1) Facial feedback influences emotional experience, 2) There is no evidence that facial feedback influences perceptions of affective quality, and 3) Incidentally manipulating facial expressions can produce facial feedback effects. The results of this meta-analysis are somewhat consistent with their conclusions. Consistent with Strack et al. (1988), results indicated that facial feedback can influence emotional experience, and that, when controlling for publication bias, facial feedback may not influence perceptions of affective quality. However, clear evidence of the effectiveness of incidental facial feedback manipulations was not uncovered (which includes, but is not limited to the pen-in-mouth manipulation). Studies that compared incidental posing to control groups that received no manipulation (s = 14; k = 31) had a small overall effect size that was not significantly greater than zero (d = 0.13, 95% CI [-0.05, 0.31]). The 30 studies (k = 47) that, like Strack et al. (1988), compared incidental posing to an expression suppression manipulation (the intended effect of the pen-in-lips manipulation; see Strack et al., 1988), also had a small non-significant overall effect size (d = 0.07; 95% CI [-0.02, 0.16]). Last, the 10 studies (k = 14) that compared two incidental posing groups (i.e., one group that incidentally poses a smile compared to another that incidentally poses a frown) did have a small-to-medium overall effect size (d = 0.43; 95% CI [0.22, 0.63]), but none of these studies specifically used the pen-in-mouth manipulation featured in Strack et al. (1988). While the results do support the facial feedback hypothesis, they do not provide clear evidence of the efficacy of the type of incidental manipulation used in the seminal Strack et al., (1988) studies. Given the results of this research synthesis, as well as similar preliminary results from a narrower in-prep research synthesis of the pen-in-mouth paradigm (Schimmack & Chen, 2017), the pen-in-mouth brick may be faulty.

Although some of its bricks may be faulty, this does not mean that the facial feedback edifice as a whole is compromised. In fact, Wagenmakers et al. (2016) were careful to point out their failure-to-replicate raises questions about the reliability of the pen-in-mouth manipulation, but do not necessarily refute the facial feedback hypothesis. Consistent with this reasoning, the meta-analysis suggests that the edifice remains standing. Yet, in the wake of the failure-to-replicate, researchers have described the entire facial feedback literature as controversial (Panasiti et al., 2017), inconclusive (Steinmetz & Posten, 2017), a potential byproduct of methodological confounds (Nicholas & Ashton-James, 2017), or even a phenomonon that has not survived replication (Smith & Apicella, 2017).[1] All of these narrative reviewers apparently gave greater weight to the Wagenmakers et al. studies that focused on one specific facial feedback manipulation rather than to the dozens of studies that came before and used multiple operational definitions. In other words, they mistook a brick for an edifice.

It does make sense to overweight Wagenmakers et al.'s studies because they were pre-registered and we can be more confident that they avoided questionable research practices (QRPs). Although effect sizes are typically weighted by precision (studies with lower variance receive more weight), it is also possible to weight by the quality of the study (Doi & Thalib, 2008). Quantifying the quality of studies in psychology can be challenging, but the data allows us to examine how much weight we would need to give to Wagenmakers et al.'s studies in order to observe a non-significant effect in our comprehensive meta-analysis.

Using our aggregated dependent effect sizes, the weighting of the Wagenmakers et al. (2016) findings were increased until a non-significant overall effect was observed. Using the standard precision-weighting approach, the 17 registered replications provided 6% of the effect size estimates, and received 13.16% of the overall weight in the original meta-analysis. By gradually increasing the weighting of the replication studies, the results indicated that the facial feedback effect remains statistically significant until the failed replication studies are given 59%

of the weight. In other words, in order to fail to accept the facial feedback hypothesis, researchers must give 59% of the weight to only 6% of the evidence (i.e., the registered replication studies). This illustrates that when researchers dismiss well-established phenomenon in light of single replication efforts, they are giving an inordinate amount of weight to the replication results.

Limitations of the Meta-Analytic Approach

Meta-analytic conclusions can be compromised when the effect size estimates are biased. Historically, meta-analysts have been interested in the effects of publication bias, and have developed many tests of the extent and impact of this bias. However, it is important to emphasize that these methods are fallible (Carter et al., 2017; Stanley, 2017), in part because they were developed and validated under the assumption that effect sizes are independent. Over half (53%) of our studies provided multiple effect sizes, and different approaches for dealing with such dependencies led to slightly different conclusions regarding publication bias in the overall facial feedback literature. Fortunately, more clear patterns emerged in the sensitivity analyses, where both approaches produced evidence of publication bias in studies examining perceptions of affective quality, and a lack of evidence of publication bias in studies examining emotional experience. Nevertheless, future research should continue to develop and validate methods for detecting publication bias, and evaluate the effectiveness of these approaches when dependent data structures exist.

Other QRPs, such as optimal stopping, *p*-hacking, and infrequent cases of outright fraud, also threaten the validity of meta-analytic conclusions. John et al. (2012) found that a high proportion of researchers admitted to performing these practices, including: deciding whether to exclude data after looking at the impact of doing so on the results (43%), deciding whether to continue data collection after looking to see whether the results were significant (58%) and stopping data collection early once significant results have been found (23%). These practices inflate meta-analytic estimates, which can create misleading conclusions. Unfortunately, little is known about the effects of QRPs on meta-analytic conclusions and methods that may control for their impact. For example, Head, Holman, Lanfear, Kahn, and Jennions (2015) suggested *p*-hacking has a weak influence on effect size estimates. But Bierman, Spottiswoode and Bijl (2016) suggested that the effects of these QRPs may add up. They found that QRP's can explain 60% of the effect size reported in a parapsychology meta-analysis. Currently, the field does not don't know much about the effects of QRP's and how to control for their effects, so meta-analyses are limited when they are a concern.

Future Directions

Researchers should strive to understand the mechanism(s) that produce facial feedback effect not just because that understanding is a hallmark of a good theory, but also because it can produce novel predictions. Early facial feedback researchers were very interested in mechanism, and proposed several competing hypotheses that were never resolved (for a review, see Soussignan, 2004). For example, some researchers suggested that facial expressions provide sensory feedback that influences our emotional experience without cognitive mediation (e.g., Gellhorn, 1964; Izard, 1971; Tomkins, 1962). Zajonc et al., (1989) suggested that facial expressions influence brain blood temperature, which subsequently has hedonic effects. Ekman (1972) and Levenson et al., (1990) suggested that facial expressions trigger autonomic and experiential changes. Last, Laird & Lacasse (2014) suggested that facial feedback effects are driven by a self-attributional process (Laird & Lacasse, 2014). Without understanding the mechanism that underlies facial feedback effects, it will be difficult to predict when facial feedback effects will occur. For example, if facial feedback effects are driven by a self-attributional process, self-perception theory would predict that smiling will only lead to more positive emotions in scenarios where it produces cognitive dissonance.

Chapter 6: Conclusion

A meta-analysis of nearly 300 effect sizes from over 130 studies support the long-standing claim that facial feedback influences emotional experience. However, these effects tend to be weak and extremely heterogeneous. Given that these effects tend to be small, it seems unlikely that facial feedback has a substantial impact on people's emotional reactions. Nevertheless, the effect is consistent with the many emotion theories that conceptualize emotion as an embodied process that depends, in part, upon facial feedback (i.e., Gellhorn, 1964; Izard, 1971; Laird, 1984; Levenson, Ekman, & Friesen, 1990; Lindquist et al., 2012; Tomkins, 1962; Waynbaum, 1907; Zajonc, Murphy, & Inglehart, 1989).

Extending Poincaré's (1902) brick-layer analogy, if meta-analysts are the inspectors of the edifices built by primary researchers, the provisional determination is the edifice as a whole is fairly sound: facial feedback influences emotional experience. However, this is not to say that every brick in that edifice is sound. Indeed, some of the bricks have been shown to be suspect, as would be expected from any edifice as large as the facial feedback literature. However, although the edifice is standing, it is still unfinished. Finishing it will require a great deal of construction, and, with such small effects, will also require a great deal of statistical power. Along the way, smiling won't likely be a substantial source of our joy, but perhaps the journey to improve upon and expand the facial feedback edifice will.

References

References with asterisks were included in the meta-analysis synthesis.

Adelmann, P. K., & Zajonc, R. B. (1989). Facial efference and the experience of emotion. *Annual Review of Psychology*, *40*(1), 249-280.

*Andréasson, P. (2010). *Emotional empathy, facial reactions, and facial feedback.* (Doctoral Dissertation). Uppsala University, Sweden.

*Andréasson, P., & Dimberg, U. (2008). Emotional empathy and facial feedback. *Journal of Nonverbal Behavior*, *32*(4), 215-224.

Ansfield, M. E. (2007). Smiling when distressed: When a smile is a frown turned upside down. *Personality and Social Psychology Bulletin*, *33*(6), 763-775.

*Baumeister, J. C., Papa, G., & Foroni, F. (2016). Deeper than skin deep–The effect of botulinum toxin-A on emotion processing. *Toxicon*, *118*, 86-90.

Baumeister, J. C., Foroni, F., Conrad, M., Rumiati, R. I., & Winkielman, P. (2017). Embodiment and Emotional Memory in First vs. Second Language. *Frontiers in Psychology*, *8*, 394. http://doi.org/10.3389/fpsyg.2017.00394

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531-533.

Berndt, E., Dudschig, C., & Kaup, B. (2016). Activating concepts by activating experiential traces: investigations with a series of anagram solution tasks. *The Quarterly Journal of Experimental Psychology*. http://dx.doi.org/10.1080/17470218.2016.1261913

Bierman DJ, Spottiswoode JP, Bijl A (2016) Testing for Questionable Research Practices in a Meta-Analysis: An Example from Experimental Parapsychology. *PLOS ONE*, *11*(5), e0153049. https://doi.org/10.1371/journal.pone.0153049

*Bodenhausen, G. V., Kramer, G. P., & Süsser, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of Personality and Social Psychology*, *66*(4), 621-632.

Borenstein, M. (2009). Effect sizes for continuous data. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 221-235). New York, NY: Russell Sage Foundation.

Borenstein M., Hedges, L. V., Higgins, J. P. L., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. West Sussex, UK: John Wiley & Sons Ltd.

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science*, *9*(3), 333-342.

Buck, R. (1980). Nonverbal behavior and the theory of emotion: The facial feedback hypothesis. *Journal of Personality and Social Psychology*, *38*(5), 811-824.

*Bush, L. K., Barr, C. L., McHugo, G. J., & Lanzetta, J. T. (1989). The effects of facial control and facial mimicry on subjective reactions to comedy routines. *Motivation and Emotion*, *13*(1), 31-52.

*Butler, E. A., Wilhelm, F. H., & Gross, J. J. (2006). Respiratory sinus arrhythmia, emotion, and emotion regulation during social interaction. *Psychophysiology*, *43*(6), 612-622.

*Butler, E. A., Egloff, B., Wlhelm, F. H., Smith, N. C., Erickson, E. A., & Gross, J. J. (2003). The social consequences of expressive suppression. *Emotion*, *3*(1), 48-67.

*Cai, A., Lou, Y., Long, Q., & Yuan, J. (2016). The sex differences in regulating unpleasant emotion by expressive suppression: extraversion matters. *Frontiers in Psychology*, *7*(1011), 1-11.

Cappella, J. N. (1993). The facial feedback hypothesis in human interaction: Review and speculation. *Journal of Language and Social Psychology*, *12*(1-2), 13-29.

Carney, D. R., Cuddy, A. J., & Yap, A. J. (2010). Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science*, *21*(10), 1363-1368.

Carruthers, A., & Carruthers, J. (2017). The future of neuromodulators in aesthetic medicine. In Carruthers, A., & Carruthers, J. (Eds.), Botulinum Toxin E-Book: Procedures in Cosmetic Dermatology Series (4th ed.). London, UK: Elsevier.

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2017, June 30). Correcting for bias in psychology: A comparison of meta-analytic methods. Retrieved from osf.io/preprints/psyarxiv/9h3nu

*Ceschi, G., & Scherer, K. (2003). Children's ability to control the facial expression of laughter and smiling: Knowledge and behaviour. *Cognition & Emotion*, *17*(3), 385-411.

*Clapp, J. D. (2012). *Expressive inhibition following trauma: Implications for understanding posttraumatic numbing*. (Doctoral Dissertation). Retrievedfrom ProQuest Dissertations & Theses Global. (3541095)

Cochrane, T. (2017). The Double Intentionality of Emotional Experience. *European Journal of Philosophy*. https://doi.org/10.1111/ejop.12233

Cohen, J. (1987). *Statistical power analysis for the behavioral sciences*. Hillside, NJ: Lawrence Erlbaum Associates.

Darwin, C. (1965). *The expression of the emotions in man and animals* (Vol. 526). Chicago, IL: University of Chicago press. (Original work published in 1872)

*Davey, G. C., Sired, R., Jones, S., Meeten, F., & Dash, S. R. (2013). The role of facial feedback in the modulation of clinically-relevant ambiguity resolution. *Cognitive Therapy and Research*, *37*(2), 284-295.

*Davis, J. I. (2008). *The connection between facial expression and emotional experience.* (Doctoral Dissertation). Retrieved from ProQuest Dissertations & Theses Global. (3305213)

*Davis, J. I., Senghas, A., Brandt, F., & Ochsner, K. N. (2010). The effects of BOTOX injections on emotional experience. *Emotion*, *10*(3), 433-440.

*Davis, J. I., Senghas, A., & Ochsner, K. N. (2009). How does facial feedback modulate emotional experience?. *Journal of Research in Personality*, *43*(5), 822-829.

*Davis, J. D., Winkielman, P., & Coulson, S. (2015). Facial action and emotional language: ERP evidence that blocking facial feedback selectively impairs sentence comprehension. *Journal of Cognitive Neuroscience, 27*(11), 2269-2280.

Davis, J. D., Winkielman, P., & Coulson, S. (2017). Sensorimotor simulation and emotion processing: Impairing facial action increases semantic retrieval demands. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(3), 652-664.

Del Re, A. C., & Hoyt, W. T. (2014). MAd: Meta-Analysis with Mean Differences. R package version 0.8-2 [Computer software]. Retrieved from http://cran.r-project.org/web/packages/MAd

*Demaree, H. A., Robinson, J. L., Everhart, D. E., & Schmeichel, B. J. (2004). Resting RSA is associated with natural and self-regulated responses to negative emotional stimuli. *Brain and Cognition*, *56*(1), 14-23.

*Demaree, H. A., Schmeichel, B. J., Robinson, J. L., Pu, J., Everhart, D. E., & Berntson, G. G. (2006). Up-and down-regulating facial disgust: Affective, vagal, sympathetic, and respiratory consequences. *Biological Psychology*, *71*(1), 90-99.

*Dillon, D. G., Ritchey, M., Johnson, B. D., & LaBar, K. S. (2007). Dissociable effects of conscious emotion regulation strategies on explicit and implicit memory. *Emotion*, *7*(2), 354-365.

*Dimberg, U., & Söderkvist, S. (2011). The voluntary facial action technique: A method to test the facial feedback hypothesis. *Journal of Nonverbal Behavior*, *35*(1), 17-33.

Doi, S. A., & Thalib, L. (2008). A quality-effects model for meta-analysis. *Epidemiology*, *19*(1), 94-100.

*Duncan, J. W., & Laird, J. D. (1980). Positive and reverse placebo effects as a function of differences in cues used in self-perception. *Journal of Personality and Social Psychology*, *39*(6), 1024-1036.

*Duncan, J., & Laird, J. D. (1977). Cross-modality consistencies in individual differences in self-attribution. *Journal of Personality*, *45*(2), 191-206.

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*(2), 455-463.

*Dzokoto, V., Wallace, D. S., Peters, L., & Bentsi-Enchill, E. (2014). Attention to emotion and non-western faces: Revisiting the facial feedback hypothesis. *The Journal of General Psychology*, *141*(2), 151-168.

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

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

Ekman (1999) Basic emotions. In Dalgleish, T., & Power, M. J. (Eds.), *Handbook of cognition and emotion* (pp. 45-60). New York, NY: John Wiley & Sons Ltd.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review*, *3*(4), 364-370.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In Davidson, R. J., Scherer, K. R., & Goldsmith, H. (Eds.), *Handbook of affective sciences* (pp. 572-595). New York, NY: Oxford University Press.

Finzi, E., & Rosenthal, N. E. (2014). Treatment of depression with onabotulinumtoxinA: a randomized, double-blind, placebo controlled trial. *Journal of Psychiatric Research*, *52*, 1-6.

Finzi, E., & Rosenthal, N. E. (2016). Emotional proprioception: Treatment of depression with afferent facial feedback. *Journal of Psychiatric Research*, *80*, 93-96.

Fisher, Z., & Tipton, E. (2015). Robumeta: Robust variance meta-regression. R package version 1.6.

*Flack, W. (2006). Peripheral feedback effects of facial expressions, bodily postures, and vocal expressions on emotional feelings. *Cognition & Emotion*, *20*(2), 177-195.

*Flack, W. F., Laird, J. D., & Cavallaro, L. A. (1999). Emotional expression and feeling in schizophrenia: effects of specific expressive behaviors on emotional experiences. *Journal of Clinical Psychology*, *55*(1), 1-20.

*Flack, W. F., Laird, J. D., & Cavallaro, L. A. (1999). Separate and combined effects of facial expressions and bodily postures on emotional feelings. *European Journal of Social Psychology*, *29*(2-3), 203-217.

Forscher, B. K. (1963). Chaos in the brickyard. *Science*, *142*(3590), 339.

Franco, A., Malhotra, N., & Simonovits, G. (2016). Underreporting in psychology experiments: Evidence from a study registry. *Social Psychological and Personality Science*, *7*(1), 8-12.

*Gan, S., Yang, J., Chen, X., & Yang, Y. (2015). The electrocortical modulation effects of different emotion regulation strategies. *Cognitive Neurodynamics*, *9*(4), 399-410.

Gellhorn, E. (1964). Motion and emotion: The role of proprioception in the physiology and pathology of the emotions. *Psychological Review*, *71*(6), 457-472.

Gerrards-Hesse, A., Spies, K., & Hesse, F. W. (1994). Experimental inductions of emotional states and their effectiveness: A review. *British Journal of Psychology*, *85*(1), 55-78.

*Goldin, P. R., McRae, K., Ramel, W., & Gross, J. J. (2008). The neural bases of emotion regulation: reappraisal and suppression of negative emotion. *Biological Psychiatry*, *63*(6), 577-586.

Gratiolet, P. (1865). *De la physionomie et des mouvements d'expression*. Paris: Hetzel.

*Gross, J. J. (1993). *Emotional suppression*. Doctoral Dissertation). Retrieved from ProQuest Dissertations & Theses Global. (9430509)

*Gross, J. J. (1998). Antecedent-and response-focused emotion regulation: divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology*, *74*(1), 224-237.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85*(2), 348-362.

*Gross, J. J., & Levenson, R. W. (1993). Emotional suppression: physiology, self-report, and expressive behavior. *Journal of Personality and Social Psychology*, *64*(6), 970-986.

*Gross, J. J., & Levenson, R. W. (1997). Hiding feelings: the acute effects of inhibiting negative and positive emotion. *Journal of Abnormal Psychology*, *106*(1), 95-103.

Han, C., Park, G. Y., Wang, S. M., Lee, S. Y., Lee, S. J., Bahk, W. M., & Pae, C. U. (2012). Can botulinum toxin improve mood in depressed patients? *Expert Review of Neurotherapeutics*, *12*(9), 1049-1051.

*Harris, C. R. (2001). Cardiovascular responses of embarrassment and effects of emotional suppression in a social setting. *Journal of Personality and Social Psychology, 81*(5), 886-897.

*Hawk, S. T., Fischer, A. H., & Van Kleef, G. A. (2012). Face the noise: embodied responses to nonverbal vocalizations of discrete emotions. *Journal of Personality and Social Psychology*, *102*(4), 796-814.

Hawlik, A. E., Freudenmann, R. W., Pinkhardt, E. H., Schönfeldt-Lecuona, C. J., & Gahr, M. (2014). Botulinum toxin for the treatment of major depressive disorder. *Fortschritte der Neurologie-Psychiatrie*, *82*(2), 93-99.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLOS Biology*, *13*(3), e1002106. https://doi.org/10.1371/journal.pbio.1002106

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39-65.

*Helt, M. S., & Fein, D. A. (2016). Facial feedback and social input: Effects on laughter and enjoyment in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *46*(1), 83-94.

*Hendricks, M. A. (2013). *An exploration of the relationship between emotional and cognitive control.* (Doctoral Dissertation). Retrieved from ProQuest Dissertations & Theses Global. (3587340)

*Hendricks, M. A., & Buchanan, T. W. (2016). Individual differences in cognitive control processes and their relationship to emotion regulation. *Cognition and Emotion*, *30*(5), 912-924.

*Henry, J. D., Green, M. J., de Lucia, A., Restuccia, C., McDonald, S., & O'Donnell, M. (2007). Emotion dysregulation in schizophrenia: reduced amplification of emotional expression is associated with emotional blunting. *Schizophrenia Research*, *95*(1), 197-204.

*Henry, J. D., Green, M. J., Restuccia, C., de Lucia, A., Rendell, P. G., McDonald, S., & Grisham, J. R. (2009). Emotion dysregulation and schizotypy. *Psychiatry Research*, *166*(2), 116-124.

*Henry, J. D., Rendell, P. G., Scicluna, A., Jackson, M., & Phillips, L. H. (2009). Emotion experience, expression, and regulation in Alzheimer's disease. *Psychology and Aging*, *24*(1), 252-257.

*Hess, U., Kappas, A., McHugo, G. J., Lanzetta, J. T., & Kleck, R. E. (1992). The facilitative effect of facial expression on the self-generation of emotion. *International Journal of Psychophysiology*, *12*(3), 251-265.

*Hofmann, S. G., Heering, S., Sawyer, A. T., & Asnaani, A. (2009). How to handle anxiety: The effects of reappraisal, acceptance, and suppression strategies on anxious arousal. *Behaviour Research and Therapy*, *47*(5), 389-394.

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. *Psychology of Aesthetics, Creativity, and the Arts*, *4*(1), 47-56.

Ioannidis, J. (2008). Interpretation of tests of heterogeneity and bias in meta-analysis. *Journal of Evaluation in Clinical Practice*, *14*(5), 951-957.

*Ito, T. A., Chiao, K. W., Devine, P. G., Lorig, T. S., & Cacioppo, J. T. (2006). The influence of facial feedback on race bias. *Psychological Science*, *17*(3), 256-261.

Itkes, O., Kimchi, R., Haj-Ali, H., Shapiro, A., & Kron, A. (2017). Dissociating Affective and Semantic Valence. *Journal of Experimental Psychology: General*, *146*(7), 924-942.

Izard, C. E. (1990). Facial expressions and the regulation of emotions. *Journal of Personality and Social Psychology*, *58*(3), 487-498.

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

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, *2*(3), 260-280.

James, W. (1884). What is an emotion? *Mind*, *9*(34), 188-205.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524-532.

Jonas, K. J., Cesario, J., Alger, M., Bailey, A. H., Bombari, D., Carney, D., ... & Jackson, B. (2017). Power poses–where do we stand? *Comprehensive Results in Social Psychology*, *2*(1), 139-141.

*Kalokerinos, E. K., Greenaway, K. H., & Denson, T. F. (2015). Reappraisal but not suppression downregulates the experience of positive and negative emotion. *Emotion*, *15*(3), 271-275.

*Kao, C. H., Su, J. C., Crocker, J., & Chang, J. H. (2016). The Benefits of Transcending Self-Interest: Examining the Role of Self-Transcendence on Expressive Suppression and Well-Being. *Journal of Happiness Studies*, 1-17.

*Kappas, A. (1989). *Control of emotion.* (Doctoral Dissertation). Retrieved from ProQuest Dissertations & Theses Global. (9102394)

Keltner, D., Ekman, P., Gonzaga, G. C., Beer, J. (2003). Facial expression of emotion. In Davidson, R. J., Scherer, K. R., & Goldsmith, H. H. (Eds.), Handbook of Affective Sciences (pp. 415-432). New York, NY: Oxford University Press.

*Kircher, T., Pohl, A., Krach, S., Thimm, M., Schulte-Rüther, M., Anders, S., & Mathiak, K. (2012). Affect-specific activation of shared networks for perception and execution of facial expressions. *Social cognitive and affective neuroscience*, *8*(4), 370-377.

*Korb, S., Grandjean, D., Samson, A. C., Delplanque, S., & Scherer, K. R. (2012). Stop laughing! Humor perception with and without expressive suppression. *Social Neuroscience*, *7*(5), 510-524.

Kraft, T. L., & Pressman, S. D. (2012). Grin and bear it: The influence of manipulated facial expression on the stress response. *Psychological Science*, *23*(11), 1372-1378.

Kring, A. M., Smith, D. A., & Neale, J. M. (1994). Individual differences in dispositional expressiveness: development and validation of the Emotional Expressivity Scale. *Journal of Personality and Social Psychology*, *66*(5), 934-949.

Kruger, T. H., & Wollmer, M. A. (2015). Depression–An emerging indication for botulinum toxin treatment. *Toxicon*, *107*, 154-157.

*Labott, S. M., & Teleha, M. K. (1996). Weeping propensity and the effects of laboratory expression or inhibition. *Motivation and Emotion*, *20*(3), 273-284.

LaFrance, M., Hecht, M. A., & Paluck, E. L. (2003). The contingent smile: a meta-analysis of sex differences in smiling. *Psychological Bulletin*, *129*(2), 305-334.

*Laird, J. D. (1974). Self-attribution of emotion: the effects of expressive behavior on the quality of emotional experience. *Journal of Personality and Social Psychology*, *29*(4), 475-486.

Laird, J. D. (1984). The real role of facial response in the experience of emotion: A reply to Tourangeau and Ellsworth, and others. *Journal of Personality and Social Psychology*, *47*(4), 909-917.

*Laird, J. D., & Crosby, M. (1974). Individual differences in self-attribution of emotion. In London H., & Nisbett, R. E. (Eds.), *Thought and feeling: The cognitive alteration of feeling states* (pp. 44-59). Chicago: Aldine Publishing Company.

Laird, J. D., & Lacasse, K. (2014). Bodily influences on emotional feelings: Accumulating evidence and extensions of William James's theory of emotion. *Emotion Review*, *6*(1), 27-34.

*Lalot, F., Delplanque, S., & Sander, D. (2014). Mindful regulation of positive emotions: a comparison with reappraisal and expressive suppression. *Frontiers in Psychology*, *5(243). 1-9*.

*Larsen, R. J., Kasimatis, M., & Frey, K. (1992). Facilitating the furrowed brow: An unobtrusive test of the facial feedback hypothesis applied to unpleasant affect. *Cognition & Emotion*, *6*(5), 321-338.

*Lee, E. A. (2011). *Expressive suppression of negative emotion: A comparison of asian american and european american norms for emotion regulation.* (Doctoral Dissertation). Retrieved from ProQuest Dissertations & Theses Global. (3483720)

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*(5), 834-855.

Lehrer, J. (2010). The truth wears off. *The New Yorker*, *13*(52), 229.

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, *27*(4), 363-384.

Lewinski, P., Fransen, M. L., & Tan, E. S. (2016). Embodied Resistance to Persuasion in Advertising. *Frontiers in Psychology*, *7*, 1202. http://doi.org/10.3389/fpsyg.2016.01202

*Lewis, M. B. (2012). Exploring the positive and negative implications of facial feedback. *Emotion*, *12*(4), 852-859.

*Lewis, M. B., & Bowler, P. J. (2009). Botulinum toxin cosmetic therapy correlates with a more positive mood. *Journal of Cosmetic Dermatology*, *8*(1), 24-26.

Lin, W., Hu, J., & Gong, Y. (2015). Is it helpful for individuals with minor depression to keep smiling? An event-related potentials analysis. *Social Behavior and Personality: An International Journal*, *43*(3), 383-396.

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*(3), 121-143.

Lumley, M. A., Cohen, J. L., Borszcz, G. S., Cano, A., Radcliffe, A. M., Porter, L. S., ... & Keefe, F. J. (2011). Pain and emotion: a biopsychosocial review of recent research. *Journal of Clinical Psychology*, *67*(9), 942-968.

*Ma, C. (2011). *Culture and emotional suppression.* (Doctoral Dissertation). Retrieved from ProQuest Dissertations & Theses Global. (3495692)

Magid, M., Finzi, E., Kruger, T. H. C., Robertson, H. T., Keeling, B. H., Jung, S., ... & Wollmer, M. A. (2015). Treating depression with botulinum toxin: a pooled analysis of randomized controlled trials. *Pharmacopsychiatry*, *25*(6), 205-210.

*Maldonado, R. C., DiLillo, D., & Hoffman, L. (2015). Can college students use emotion regulation strategies to alter intimate partner aggression-risk behaviors? An examination using I³ theory. *Psychology of Violence*, *5*(1), 46-55.

*Marmolejo-Ramos, F., & Dunn, J. (2013). On the activation of sensorimotor systems during the processing of emotionally-laden stimuli. *Universitas Psychologica*, *12*(5), 1515-1546.

*Martijn, C., Tenbült, P., Merckelbach, H., Dreezens, E., & de Vries, N. K. (2002). Getting a grip on ourselves: Challenging expectancies about loss of energy after self-control. *Social Cognition*, *20*(6), 441-460.

Matsumoto, D. (1987). The role of facial response in the experience of emotion: More methodological problems and a meta-analysis. *Journal of Personality and Social Psychology*, *54*(4), 769-774.

Martin, M. (1990). On the induction of mood. *Clinical Psychology Review*, *10*(6), 669-697.

*McCanne, T. R., & Anderson, J. A. (1987). Emotional responding following experimental manipulation of facial electromyographic activity. *Journal of Personality and Social Psychology*, *52*(4), 759-768.

*McCaul, K. D., Holmes, D. S., & Solomon, S. (1982). Voluntary expressive changes and emotion. *Journal of Personality and Social Psychology*, *42*(1), 145-152.

McIntosh, D. N. (1996). Facial feedback hypotheses: Evidence, implications, and directions. *Motivation and Emotion*, *20*(2), 121-147.

*McIntosh, D.N., Zajonc, R.B., Vig, P.S., & Emerick, S.W. (1997). Facial movement, breathing, temperature, and affect: Implications of the vascular theory of emotional efference. *Cognition & Emotion*, *11*(2), 171-196.

McRae, K., Ochsner, K. N., Mauss, I. B., Gabrieli, J. J., & Gross, J. J. (2008). Gender differences in emotion regulation: An fMRI study of cognitive reappraisal. *Group Processes & Intergroup Relations*, *11*(2), 143-162.

*Meeten, F., Ivak, P., Dash, S., Knowles, S., Duka, T., Scott, R., ... & Davey, G. C. (2015). The effect of facial expressions on the evaluation of ambiguous statements. *Journal of Experimental Psychopathology*, *6*(3), 253-263.

*Miyamoto, Y. (2006). *Culturally situated cognition: The influence of communication practices, physical environments, and intra-individual motor actions* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (3224701)

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLOS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

*Moore, S. A., & Zoellner, L. A. (2012). The effects of expressive and experiential suppression on memory accuracy and memory distortion in women with and without PTSD. *Journal of Experimental Psychopathology*, *3*(3), 368-392.

*Ohira, H., & Kurono, K. (1993). Facial feedback effects on impression formation. *Perceptual and Motor Skills*, *77*, 1251-1258.

Nicholas, M. K., & Ashton-James, C. (2017). Embodied pain: grasping a thorny problem? *Pain*, *158*(6), 993-994.

Nolen-Hoeksema, S., & Aldao, A. (2011). Gender and age differences in emotion regulation strategies and their relationship to depressive symptoms. *Personality and Individual Differences*, *51*(6), 704-708.

Nosek, B. A., & Lakens, D. (2014). Registered reports. *Social Psychology*, *45*, 137-141.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716. https://doi.org/10.1126/science.aac4716

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, *76*(6), 972.

Panasiti, M. S., Porciello, G., & Aglioti, S. M. (2017). The bright and the dark sides of motor simulation. *Neuropsychologia*. http://dx.doi.org/10.1016/j.neuropsychologia.2017.05.020

*Paredes, B., Stavraki, M., Briñol, P., & Petty, R. E. (2013). Smiling after thinking increases reliance on thoughts. *Social Psychology*, *44*, 349-353.

*Paul, S., Simon, D., Kniesche, R., Kathmann, N., & Endrass, T. (2013). Timing effects of antecedent-and response-focused emotion regulation strategies. *Biological Psychology*, *94*(1), 136-142.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277-293.

*Pedder, D. J., Terrett, G., Bailey, P. E., Henry, J. D., Ruffman, T., & Rendell, P. G. (2016). Reduced facial reactivity as a contributor to preserved emotion regulation in older adults. *Psychology and Aging*, *31*(1), 114-125.

Pennebaker, J. W., & Roberts, T. A. (1992). Toward a his and hers theory of emotion: Gender differences in visceral perception. *Journal of Social and Clinical Psychology*, *11*(3), 199-212.

*Phillips, L. H., Henry, J. D., Hosie, J. A., & Milne, A. B. (2008). Effective regulation of the experience and expression of negative affect in old age. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *63*(3), 138-145.

Piderit, T. (1858). *Grundsätze der mimik und physiognomik*. Braunschweig: Vieweg und Sohn.

Poincaré, H. (1902). La science et l'hypothese. Flammarion.

Pustejovsky, J. (2016). clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections. R package version 0.2.2 [Computer software]. Retrieved from https://rdrr.io/cran/clubSandwich/

Price, T. F., & Harmon-Jones, E. (2015). Embodied emotion: the influence of manipulated facial and bodily states on emotive responses. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(6), 461-473.

Price, T. F., Peterson, C. K., & Harmon-Jones, E. (2012). The emotive neuroscience of embodiment. *Motivation and Emotion*, *36*(1), 27-37.

Prochazkova, E., & Kret, M. E. (2017). Connecting Minds and Sharing Emotions through Mimicry: A Neurocognitive Model of Emotional Contagion. *Neuroscience & Biobehavioral Reviews*, *80*, 99-114.

Quirin, M., Kazén, M., & Kuhl, J. (2009). When nonsense sounds happy or helpless: the implicit positive and negative affect test (IPANAT). *Journal of Personality and Social Psychology*, *97*(3), 500.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ranehill, E., Dreber, A., Johannesson, M., Leiberg, S., Sul, S., & Weber, R. A. (2015). Assessing the robustness of power posing: No effect on hormones and risk tolerance in a large sample of men and women. *Psychological Science*, *26*(5), 653-656.

Raudenbush, S. W., Becker, B. J., & Kalaian, H. (1988). Modeling multivariate effect sizes. *Psychological Bulletin*, *103*(1), 111-120.

Reisenzein, R., Horstmann, G., & Schützwohl, A. (2017). The Cognitive-Evolutionary Model of Surprise: A Review of the Evidence. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12292

*Reisenzein, R., & Studtmann, M. (2007). On the expression and experience of surprise: no evidence for facial feedback, but evidence for a reverse self-inference effect. *Emotion*, *7*(3), 612-627.

*Richards, J. M., & Gross, J. J. (1999). Composure at any cost? The cognitive consequences of emotion suppression. *Personality and Social Psychology Bulletin*, *25*(8), 1033-1044.

*Richards, J. M., & Gross, J. J. (2000). Emotion regulation and memory: the cognitive costs of keeping one's cool. *Journal of Personality and Social Psychology*, *79*(3), 410-424.

*Richards, J. M., & Gross, J. J. (2006). Personality and emotional memory: How regulating emotion impairs memory for emotional events. *Journal of Research in Personality*, *40*(5), 631-651.

*Richards, J. M., Butler, E. A., & Gross, J. J. (2003). Emotion regulation in romantic relationships: The cognitive consequences of concealing feelings. *Journal of Social and Personal Relationships*, *20*(5), 599-620.

*Roberts, N. A., Levenson, R. W., & Gross, J. J. (2008). Cardiovascular costs of emotion suppression cross ethnic lines. *International Journal of Psychophysiology*, *70*(1), 82-87.

*Robinson, J. L., & Demaree, H. A. (2009). Experiencing and regulating sadness: Physiological and cognitive effects. *Brain and Cognition*, *70*(1), 13-20.

*Roemer, R. J. (2014). *Smiling into happiness: Conditioning positive affect.* (Doctoral Dissertation). Available from ProQuest Dissertations & Theses Global. (3639733)

Rohatgi, A. (2011). WebPlotDigitizer [Computer software]. Retrieved from http://arohatgi. info/WebPlotDigitizer/app

*Rohrmann, S., Hopp, H., Schienle, A., & Hodapp, V. (2009). Emotion regulation, disgust sensitivity, and psychophysiological responses to a disgust-inducing film. *Anxiety, Stress, & Coping*, *22*(2), 215-236.

Rosenberg, M. S. (2005). The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*, *59*(2), 464-468.

Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, *99*(3), 400-406.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. New York, NY: John Wiley & Sons.

*Rummer, R., Schweppe, J., Schlegelmilch, R., & Grice, M. (2014). Mood is linked to vowel type: The role of articulatory movements. *Emotion*, *14*(2), 246-250.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161-1178.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145-172.

Salovey, P., Mayer, J. D., Goldman, S. L., Turvey, C., & Palfai, T. P. (1995). Emotional attention, clarity, and repair: Exploring emotional intelligence using the Trait Meta-Mood Scale. In J. W. Pennebaker (Ed.), *Emotion, disclosure, & health* (pp. 125-154).

Schimmack, U., & Chen, Y. (2017, August 14) The power of the pen paradigm: a replicability analysis. Retrieved from https://replicationindex.wordpress.com/category/facial-feedback-hypothesis/

*Schmeichel, B. J., Vohs, K. D., & Baumeister, R. F. (2003). Intellectual performance and ego depletion: role of the self in logical reasoning and other information processing. *Journal of Personality and Social Psychology*, *85*(1), 33-46.

*Schmeichel, B. J., Volokhov, R. N., & Demaree, H. A. (2008). Working memory capacity and the self-regulation of emotional expression and experience. *Journal of Personality and Social Psychology*, *95*(6), 1526-1540.

Schmitz, B. (2016). *Art-of-living*. Cham, Switzerland: Springer International Publishing.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*(4), 813-838.

Smith, K. M., & Apicella, C. L. (2017). Winners, losers, and posers: The effect of power poses on testosterone and risk-taking following competition. *Hormones and Behavior*, *92*, 172-181.

* Söderkvist, S., Ohlén, K., & Dimberg, U. How the experience of emotion is modulated by facial feedback; Study 1. (submitted).

* Söderkvist, S., Ohlén, K., & Dimberg, U. How the experience of emotion is modulated by facial feedback; Study 2. (submitted).

* Söderkvist, S., & Dimberg, U. Unpublished work (in prep).

*Soussignan, R. (2002). Duchenne smile, emotional experience, and autonomic reactivity: a test of the facial feedback hypothesis. *Emotion*, *2*(1), 52-74.

Soussignan, R. (2004). Regulatory function of facial actions in emotion processes. *Advances in Psychology Research*, *31*, 171-196.

Stanley, T. D. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychological and Personality Science*. https://doi.org/10.1177/1948550617693062

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60-78.

Steinmetz, J., & Posten, A. C. (2017). Physical temperature affects response behavior. *Journal of Experimental Social Psychology*, *70*, 294-300.

*Stel, M., van den Heuvel, C., & Smeets, R. C. (2008). Facial feedback mechanisms in autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*(7), 1250-1258.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2010). Meta-analysis of free-response studies, 1992–2008: Assessing the noise reduction model in parapsychology. *Psychological Bulletin*, *136*(4), 471-485.

Strack, F. (2016). Reflection on the smiling registered replication report. *Perspectives on Psychological Science*, *11*(6), 929-930.

*Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: a nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, *54*(5), 768-777.

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., ... & Tetzlaff, J. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, *343*. https://doi.org/10.1136/bmj.d4002

*Tamir, M., Robinson, M. D., Clore, G. L., Martin, L. L., & Whitaker, D. J. (2004). Are we puppets on a string? The contextual meaning of unconscious expressive cues. *Personality and Social Psychology Bulletin*, *30*(2), 237-249.

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, *5*(1), 13-30.

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, *2*(1), 85-112.

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, *22*(13), 2113-2126.

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

*Tourangeau, R., & Ellsworth, P. C. (1979). The role of facial response in the experience of emotion. *Journal of Personality and Social Psychology*, *37*(9), 1519-1531.

*Trent, J. (2010). *Intuition and facial feedback* (Unpublished doctoral dissertation). University of Missouri, Columbia.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105-110

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*(3), 419-435.

Verduyn, P., Delaveau, P., Rotgé, J. Y., Fossati, P., & Van Mechelen, I. (2015). Determinants of emotion duration and underlying psychological and neural mechanisms. *Emotion Review*, *7*(4), 330-335.

*Vieillard, S., Harm, J., & Bigand, E. (2015). Expressive suppression and enhancement during music-elicited emotions in younger and older adults. *Frontiers in Aging Neuroscience*, *7*, 11. http://doi.org/10.3389/fnagi.2015.00011

*Wagenmakers, E. J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams Jr, R. B., ... & Zwaan, R. A. (2016). Registered Replication Report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science*, *11*(6), 917-928.

*Wittmer, V. T. (1985). Defensive style: Manipulation of facial expressiveness and its impact on psychophysiological reactivity and self-reported anxiety (repression-sensitization). (Doctoral Dissertation). Retrieved from ProQuest Dissertations & Theses Global. (8601507)

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063-1070.

Waynbaum, I. (1907). *La physionomie humaine: Son mecanisme et son roˆle social*. Paris: Alcan.

Webb, T. L., Miles, E., & Sheeran, P. (2012). Dealing with feeling: a meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin*, *138*(4), 775-808.

Westermann, R., Stahl, G., & Hesse, F. (1996). Relative effectiveness and validity of mood induction procedures: a meta-analysis. *European Journal of Social Psychology*, *26*, 557-580.

Winton, W. M. (1986). The role of facial response in self-reports of emotion: A critique of Laird. *Journal of Personality and Social Psychology*, *50*(4), 808-812.

Wollmer, M. A., de Boer, C., Kalak, N., Beck, J., Götz, T., Schmidt, T., ... & Sönmez, D. (2012). Facing depression with botulinum toxin: a randomized controlled trial. *Journal of Psychiatric Research*, *46*(5), 574-581.

Whissell, C. M. (1985). The role of the face in human emotion: first system or one of many? *Perceptual and Motor Skills*, *61*(1), 3-12.

*Yartz, A. R. (2003). *Individual differences in disgust sensitivity and voluntary emotion regulation: Subjective, physiological, and behavioral responses to disgust and pleasant pictures.* (Doctoral Dissertation). Retrieved from ProQuest Dissertations & Theses Global. (3102423)

*Zajonc, R. B., Murphy, S. T., & Inglehart, M. (1989). Feeling and facial efference: Implications of the vascular theory of emotion. *Psychological Review, 96*(3), 395-416.

*Zariffa, J., Hitzig, S. L., & Popovic, M. R. (2014). Neuromodulation of emotion using functional electrical stimulation applied to facial muscles. *Neuromodulation: Technology at the Neural Interface*, *17*(1), 85-92.

*Zhu, N., Cai, Y. H., Sun, F. W., & Yang-yang, Y. F. (2015). Mapping the emotional landscape: The role of specific emotions in conceptual categorization. *Acta Psychologica*, 159, 41-51.

Appendices

Appendix A: Figures

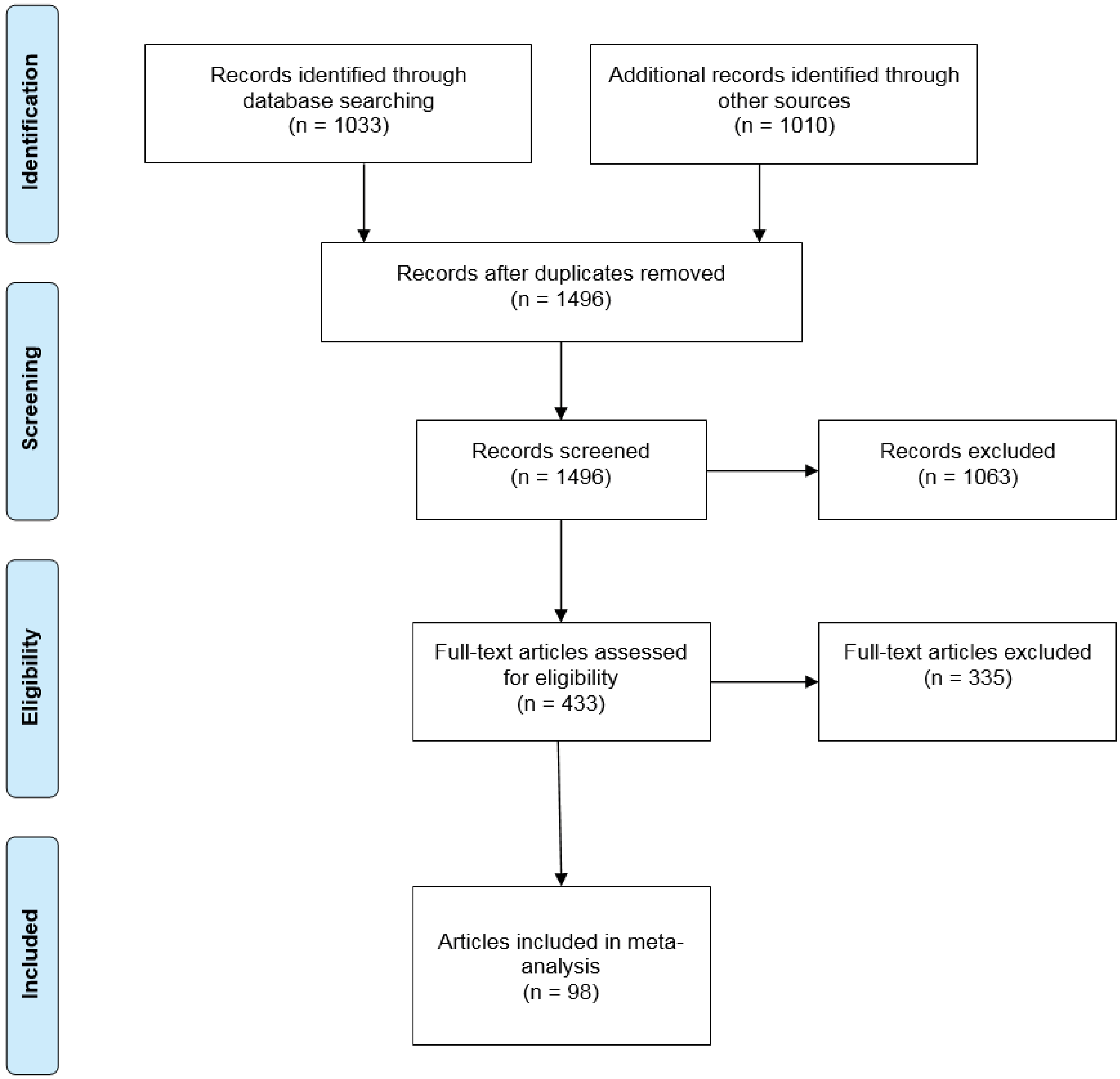

Figure 1. PRISMA-style flow chart showing selection of studies for the present meta-analysis on the facial feedback literature.

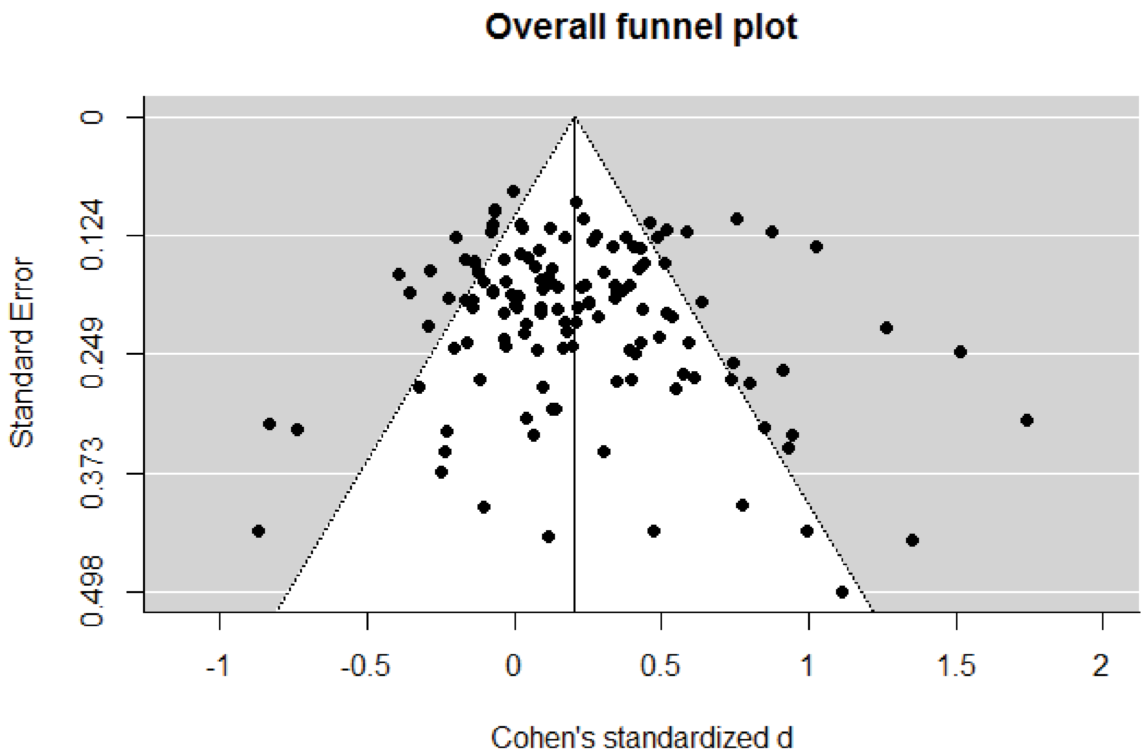

Figure 2. Overall funnel plot for studies examining the impact of facial expressions on affective experience.

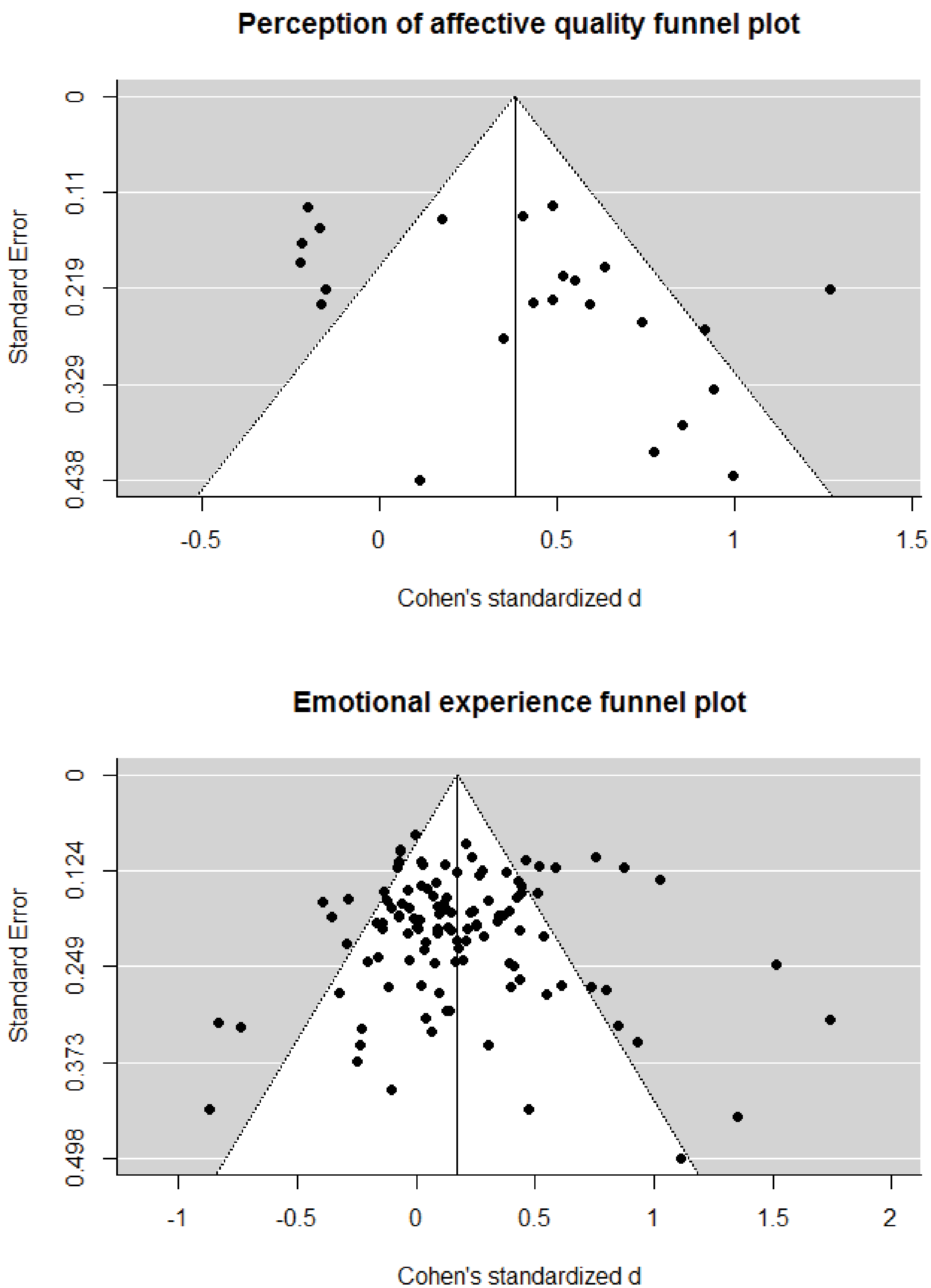

Figure 3. Funnel plots for studies examining the effect of facial feedback on perceptions of affective quality and the effect of facial feedback on emotional experience.

Appendix B: Tables

Table 1. Moderator coding criteria.

Moderator (bolded) and level	Criteria
Type of affective response	
Perception of affective quality	Participants reported their affective reaction to the stimulus (e.g., *How funny is the photo?*).
Emotional experience	Participants reported their emotional experience (e.g., *How amused did the photo make you feel?*).
Modulation vs. initiation	
Modulation	Emotional stimuli were presented.
Initiation	Either no stimuli were presented, or non-evocative stimuli/tasks were presented (e.g., neutral images & filler tasks).
Discrete vs. dimensional emotion measure	
Discrete	Measures of discrete emotions (such as anger or happiness) were collected.
Discrete emotion	
Anger Disgust Fear Happiness Sadness Surprise	Classified according to the discrete emotions identified in Ekman and Cordaro's basic emotion theory (2011).[a] Some studies measured emotions that were similar, but not included in Ekman and Cordaro's classification. We categorized these cases into their most similar discrete emotion.
Dimensional	Bipolar measures or measures of positive or negative affect, were reported.
Dimensional emotion	
Positivity	If the facial feedback manipulation was positive in nature (e.g., smiling), *or* the facial feedback manipulation was neither positive nor negative (e.g., suppression) but the stimuli were positive.
Negativity	If the facial feedback manipulation was negative in nature (e.g., frowning), *or* the facial feedback manipulation was neither positive nor negative (e.g., suppression) but the stimuli were negative.
Between vs. within-subjects design	
Between	Effect size estimates from between-subject comparisons.
Within	Effect size estimates from within-subject comparisons.
Facial feedback manipulation	
Botox-control Exaggeration-control Posing-control Incidental-control Suppression-control Posing-posing Posing-suppression Incidental-incidental Incidental-suppression Suppression-exaggeration	All procedures were coded in a manner that captures both the procedure used in the experimental group and the procedure used in the comparison group.

Table 1. (Continued)

Moderator (bolded) and level	Criteria
Stimuli	
Audio	
Film	
Imagined Scenarios	
Pictures	
Sentences	
Social context	
Stories	
Proportion of women (0-100)	Calculated using each study's reported gender composition for their entire sample. If studies excluded participants and reported the gender composition of their remaining sample, we used these updated values.
Timing of measurement	
During manipulation	Methodology stated participants held the manipulation while providing self-reports, *or* participants were instructed to hold the manipulation throughout the experiment.
After manipulation	Methodology stated participants did not hold the manipulation while giving self-reports, there was a break between the manipulation and self-reports, *or* participants were instructed to hold the manipulation at a specific moment in the experiment.
Awareness of video recording	
Yes	Participants were told they were going to be recorded *or* the methodology stated that a video camera was placed within participant view.
No	Methodology stated that participants were unaware of video recording, that the video camera was hidden, *or* that there was not a video camera.
Publication year	
Publication status	
Unpublished	Dissertations, unpublished data, and in-prep manuscripts.
Published	Peer-reviewed articles.

Note. [a] Ekman and Cordaro (2011) included contempt in their list of basic emotions, but no facial feedback studies have investigated contempt

Table 2. Subgroup and moderator(bolded) analyses

Moderator (bolded) and level	*s*	*k*	*d*	β_1	*F*	95% CI	*p*
Type of affective response	136	286		.23		[.03, .43]	.02
Perception of affective quality	24	39	0.38			[0.19, 0.57]	.0004
Emotional experience	117	247	0.17			[0.11, 0.23]	.0000004
Modulation vs. initiation	116	246		.19		[-.37, -.01]	.04
Modulation	92	179	0.13			[0.07, 0.19]	.00007
Initiation	28	67	0.32			[0.15, 0.49]	.0005
Discrete vs. dimensional emotion measure	116	246		.05		[-.08, .18]	.43
Discrete	57	130	0.19			[0.09, 0.29]	.0003
Discrete emotion	56	129			.77		.60
Anger	11	18	0.53			[0.19, 0.87]	.006
Disgust	14	23	0.29			[0.03, 0.56]	.03
Fear	11	15	0.13			[-0.05, 0.3]	.13
Happiness	36	44	0.23			[0.08, 0.37]	.004
Sadness	16	20	0.30			[0.06, 0.55]	.02
Surprise	2	9	-0.31			[-5.57, 4.95]	.59
Dimensional	63	116	0.14			[0.06, 0.21]	.0005
Dimensional emotion	58	109		.04		[-0.12, 0.19]	.63
Positivity	35	57	0.18			[0.07, 0.29]	.002
Negativity	37	52	0.12			[0.01, 0.22]	.03
Between vs. within-subjects design	136	286		.10		[-.02, .22]	.12
Between	79	150	0.15			[0.07, 0.23]	.0002
Within	59	136	0.25			[0.16, 0.35]	.000001
Facial feedback manipulation	134	284			1.78[b]		.16
Botox-control	3	6	0.71			[-1.07, 2.49]	.23
Exaggeration-control	15	29	-0.04			[-0.41, 0.33]	.82
Posing-control	9	20	0.30			[-0.16, 0.76]	.17
Incidental-control	14	31	0.13			[-0.05, 0.31]	.15
Suppression-control	57	96	0.15			[0.04, 0.25]	.006
Posing-posing	13	33	0.54			[0.29, 0.8]	.0006
Posing-suppression	3	5	0.26			[-0.55, 1.08]	.30
Incidental-incidental	10	14	0.43			[0.22, 0.63]	.001
Incidental-suppression	30	43	0.07			[-0.02, 0.16]	.11
Suppression-exaggeration	4	7	0.34			[-0.68, 1.36]	.36
Stimuli	110	217			92.1		<.005
Audio	3	10	0.72			[-0.82, 2.27]	.18

Table 2. (Continued)

Moderator (bolded) and level	*s*	*k*	*d*	β_1	*F*	95% CI	*p*
Film	42	94	0.13			[0.03, 0.22]	.009
Imagined Scenarios[a]	1	5	1.28			[-0.98, 3.53]	.27
Pictures	52	84	0.16			[0.08, 0.23]	.0002
Sentences[a]	2	4	0.70			[0.43, 0.96]	.0000003
Social context	10	18	-0.14			[-0.74, 0.46]	.61
Stories[a]	2	2	0.41			[-.29, 1.10]	.25
Proportion of women (0-100)	121	261		.16		[-.09, .42]	.21
Timing of measurement	113	237		-.03		[-.17, .11]	.65
During manipulation	42	81	0.18			[0.09, 0.26]	.0001
After manipulation	71	156	0.22			[0.12, 0.33]	.00008
Awareness of video recording	125	265		-.06		[-.20, .07]	.36
Yes	54	116	0.17			[0.06, 0.28]	.003
No	72	149	0.23			[0.15, 0.32]	.0000007
Publication year	133	283		-.01		[-0.01, -0.001]	.03
Publication status				-.05		[-.18, .08]	.43
Unpublished	20	57	0.15			[0.04, 0.26]	.01
Published	117	229	0.21			[0.14, 0.28]	.00000003

Note. *k* = number of effect size estimates; *s* = number of studies; *d* = Cohen's standardized difference*;* β_1 coefficients are from separate meta-regressions with RVE where a continuous moderator was entered in the model as a predictor or a categorical moderator with two levels was dummy-coded and entered into the model as a predictor; *F* values are from Approximate Hotelling-Zhang with small sample correction omnibus tests of the effects of moderators with more than two levels; *95% C.I* corresponds to the β_1 coefficient for moderators or *d* values for individual levels of moderators; *p* corresponds to the β_1 coefficient or *F* value for moderators, or *d* values for individuals levels of a moderator.

[a] For cases with too few observations for the RVE approach, we calculated their mean effect size using a traditional random-effects meta-regression model.

[b] *F*-test is comparing all types of methodologies. *F*-test that compares only studies featuring a true control condition yielded similar results ($F(10.4) = .62$, $p = .66$).

The number of effect size estimates and studies often do not add up as expected because some studies provided multiple effect size estimates and/or did not provide data for a level of a moderator.

Vita

Nicholas Coles was born and raised in Longwood, FL, where he attended Highlands Elementary School, Greenwood Lakes Middle School, and Lake Mary High School, graduating in 2011. He subsequently attended the University of Central Florida, where he began researching emotion at the Applied Cognitive and Technology Lab, the UCF Medical School, and the Institution of Simulation and Training. He completed an undergraduate thesis, titled *A psychophysiological investigation of the paradoxical effects of valuing happiness*, under the supervision of Dr. Valerie Sims, and received a Bachelor of Science degree in Interdisciplinary in 2015, with multiple distinctions.

Nicholas joined the Psychology graduate program at the University of Tennessee-Knoxville in 2015, where he continues to research emotion and meta-science under the supervision of Dr. Jeff Larsen. Nicholas graduated with a Masters of Arts in Psychology in December 2017, and is continuing to work towards a Doctor of Philosophy in Psychology.