EMOTIONAL LABOR WITHIN A PERFORMANCE EPISODE: UNDERSTANDING WHEN AND WHY EMPLOYEES CHANGE BETWEEN EMOTION REGULATION TECHNIQUES WITH CUSTOMERS

A Dissertation

Presented to

The Graduate Faculty of The University of Akron

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

Allison S. Gabriel

May, 2013
EMOTIONAL LABOR WITHIN A PERFORMANCE EPISODE: UNDERSTANDING WHEN AND WHY EMPLOYEES CHANGE BETWEEN EMOTION REGULATION TECHNIQUES WITH CUSTOMERS

Allison S. Gabriel
Dissertation

Approved:

Advisor
Dr. James M. Diefendorff

Committee Member
Dr. Steven R. Ash

Committee Member
Dr. Joelle D. Elicker

Committee Member
Dr. Rosalie J. Hall

Accepteed:

Department Chair
Dr. Paul E. Levy

Dean of the College
Dr. Chand Midha

Dean of the Graduate School
Dr. George R. Newkome

Date

Dr. Rosalie J. Hall

Dr. Paul E. Levy
EMOTIONAL LABOR RESEARCH HAS A LONG-STANDING TRADITION OF UTILIZING PERSON-LEVEL, CROSS-SECTIONAL DATA WHEN ATTEMPTING TO UNDERSTAND HOW SURFACE ACTING AND DEEP ACTING OPERATE IN RELATION TO OUTCOMES SUCH AS EMOTIONAL EXHAUSTION AND EMOTIONAL DISPLAYS. HOWEVER, SUCH RESEARCH DESIGNS IGNORE THE POSSIBILITY OF MEANINGFUL VARIANCE IN BOTH SURFACE AND DEEP ACTING WITHIN SINGLE PERFORMANCE EPISODES. TO ADDRESS THIS, THE CURRENT CALL CENTER SIMULATION STUDY UTILIZED CONTINUOUS RATING METHODOLOGY (E.G., RUEF & LEVENSON, 2007) TO ASSESS THREE KEY QUESTIONS, EACH OF WHICH WAS A NEW CONTRIBUTION TO THE EMOTIONAL LABOR LITERATURE. THE FIRST QUESTION ASSESSED THE WITHIN-EPISTODE RELATIONSHIP OF SURFACE ACTING AND DEEP ACTING TO DETERMINE WHETHER THESE TWO CONSTRUCTS ARE ANTIPODES, OR EXCLUSIVE AND OPPOSING CONSTRUCTS (AS INDICATED BY A NEGATIVE CORRELATION), OR IF THEY ARE POSITIVELY RELATED SUCH THAT INDIVIDUALS CAN ENGAGE IN BOTH SIMULTANEOUSLY (E.G., BEAL & TROUGAKOS, 2013). THE SECOND QUESTION PROBED THE ROLE CUSTOMERS PLAY IN SHAPING EMPLOYEE EMOTIONAL LABOR. SPECIFICALLY, THE CURRENT STUDY BUILT FROM PREVIOUS WORK (E.G., CÔTÉ, 2005; DIEFENDORFF & GOSERAND, 2003) TO IDENTIFY HOW SOCIAL FEEDBACK FROM CUSTOMERS (I.E., CONFEDERATE CALLERS) INFLUENCED THE TYPE OF REGULATION THAT EMPLOYEES (I.E., PARTICIPANTS) UTILIZE. FINALLY, THE LAST QUESTION FOCUSED ON WHETHER THIRD-PARTY EVALUATIONS OF WITHIN-EPISTODE VARIATIONS IN THE VALENCE OF EMPLOYEE EMOTIONALITY CORRESPONDED TO EMPLOYEE REPORTS OF SPECIFIC EMOTION REGULATION STRATEGIES. IN A CALL CENTER SIMULATION WITH A SAMPLE OF UNDERGRADUATE STUDENTS, RESULTS INDICATED THAT COLLECTING CONTINUOUS,
momentary-level assessments offers unique information for emotional labor research. Both surface and deep acting strategies positively covaried with each other over time, though this effect was attenuated depending upon study condition (i.e., service failure, service recovery) and the phase of the performance episode (i.e., when the between-subjects manipulations were implemented). Further, results indicated that the social context (as dictated by confederate callers) impacted the extent to which participants utilized varying levels and combinations of surface and deep acting strategies. Finally, surface and deep acting were both found to be negatively related to emotional exhaustion and third-party affect ratings, though this effect also depended upon the study condition and the phase of the performance episode. Combined, the results depict a complex picture of the momentary influences and outcomes of surface and deep acting emotion regulation strategies for employee performance and well-being.
ACKNOWLEDGEMENTS

As a runner, I usually joke that I do not take the phrase “it’s like a marathon” lightly. However, when thinking of my time at UA, it certainly has been like a marathon, and there is a whole host of supporting individuals who aided me along the way.

To Jim Diefendorff: thank you for seeing something in me when I haven’t always believed in myself. You have always treated me as more of a colleague and friend than an advisee, and I thank your family and friends for doing the same. To Paul Levy, Rosalie Hall, Joelle Elicker, and Steve Ash: thank you for letting me pursue this idea and turn it into a reality. Your support has been incredibly meaningful. I would be remiss not to thank Alicia Grandey for encouraging me to enter the field of I/O Psychology, as well as Rebecca Erickson for allowing me to collaborate with her over the last five years, Andee Snell for the statistics/life pep talks throughout my time here, and Mike Plybon for helping develop the computer program used for my dissertation.

To Cecily Becker, Christina Moran, Sara Shondrick, and Darlene Thompson: I am forever fortunate for our lifelong friendships; you four (+ the I/O house husbands) make each day brighter. To Stephen and Caroline Hill: our endless studying turned into endless laughter and friendship; I value you both so much in my life. To Grace Leung, thank you for being a wonderful mentor. To Megan Chandler: thank you for always being the best listener I could have ever asked for. To my younger peers Ali Benedetti, Noelle Frantz, Chantale Wilson, Jessica Dinh, Sara Powers, and Ernest Hoffman: I am so sad to leave
you, but know great things await. To Matt Sloan and my faithful team of undergraduate RAs: thank you for helping me, and stepping up on more than one occasion to make sure the project ran smoothly in my absence.

I am thankful for my friends outside Akron who helped keep me sane. To Julia Covelli: thank you for always making me laugh when I needed it most. To Sam Kling, Morgan Barth, and Darian Leib: thank you for always helping me keep perspective.

To my parents and my brother, Michael: you have cheered for me from the start, tried to read my papers (though I think that was out of pity!) and have let me call at all hours of the day when I felt like I was in over my head. I could not have done this without you. To my Grandma, your phone calls have always made me laugh and have been a ray of sunshine during some stressful times. To the rest of my family: thank you for supporting me through the years and believing in me.

Finally, to my husband, Mike: the best part of coming here was meeting you. Thank you for trusting in our relationship enough to think proposing during comps was perfectly fine, that planning a wedding during my dissertation proposal was manageable, and that the last year of job searching would lead to our forever future. I am so thankful for your family who welcomed me with open arms, your calmness and advice when I haven’t always wanted to listen, your unconditional love, and for the journey we are just beginning as we head to Richmond.

I dedicate this to my late grandparents, Seymour Yodowitz, Laura Gabriel, and Dr. John Gabriel. Opi – I wish and dream that I turn into half the professor and mentor you were.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>viii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
</tbody>
</table>

CHAPTER

I. STATEMENT OF THE PROBLEM ................................................................. 1

II. LITERATURE REVIEW .............................................................................. 12

Surface Acting and Deep Acting ................................................................. 13

Person-Level Assessments of Emotion Regulation ...................................... 16

Within-Person Assessments of Emotion Regulation .................................... 19

Continuous Rating Methodology ................................................................. 25

Within-Episode Dynamics of Emotional Labor ........................................... 29

The Relationship Between Surface Acting and Deep Acting ..................... 29

The Role of Social Feedback in Shaping Emotional Labor ......................... 34

Outcomes of Within-Episode Emotional Labor .......................................... 46

Emotional Exhaustion .................................................................................. 46

Observer-Rated Participant Hedonic Tone ................................................ 47

III. METHOD .................................................................................................. 52

Participants .................................................................................................. 52

Procedure .................................................................................................... 53

Pre-Simulation Measures ............................................................................ 59
Surface Acting and Deep Acting Trajectories ............................................................. 142

V. SUMMARY .................................................................................................................. 161

Can Customers Influence Employee Emotion Regulation? ........................................ 163

What is the Relationship Between Surface Acting and Deep Acting? ...................... 167

Does Participant Emotionality Covary with Momentary Emotion Regulation? .......... 168

Are Surface and Deep Acting Harmful to Employee Well-Being? ........................... 169

Are Surface and Deep Acting Harmful to Employee Emotional Performance? .......... 172

Implications for Emotional Labor Theory .................................................................... 174

Practical Implications .................................................................................................. 177

Limitations and Future Directions ................................................................................ 181

Conclusion ..................................................................................................................... 185

REFERENCES ................................................................................................................. 187

APPENDICES ................................................................................................................ 198

APPENDIX A: TRAINING MATERIALS – SURFACE ACTING FIRST .................. 199

APPENDIX B: TRAINING MATERIALS – DEEP ACTING FIRST ....................... 205

APPENDIX C: PARTICIPANT INFORMATION ......................................................... 211

APPENDIX D: POWERPOINT MANUAL ................................................................. 213

APPENDIX E: SERVICE FAILURE SCRIPT ............................................................ 218

APPENDIX F: SERVICE RECOVERY SCRIPT ......................................................... 221

APPENDIX G: POST-SIMULATION SCRIPT ............................................................. 224

APPENDIX H: PRE-SIMULATION MEASURES ......................................................... 225

APPENDIX I: PHYSIOLOGICAL MEASURES ......................................................... 231
APPENDIX J: POST-SIMULATION MEASURES..................................................232

APPENDIX K: INSTITUTIONAL REVIEW BOARD APPROVAL..................234
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Summary of hypotheses</td>
</tr>
<tr>
<td>2</td>
<td>Independent samples t-tests for experimenter effects</td>
</tr>
<tr>
<td>3</td>
<td>One-way ANOVAs for confederate caller effects</td>
</tr>
<tr>
<td>4</td>
<td>Means and standard deviations of surface acting and deep acting split by counterbalancing</td>
</tr>
<tr>
<td>5</td>
<td>Independent samples t-tests for counterbalancing effects</td>
</tr>
<tr>
<td>6</td>
<td>Correlations among focal variables across the entire performance episode split by study condition and counterbalancing</td>
</tr>
<tr>
<td>7</td>
<td>Correlations among focal variables within the three phases of the performance episode split by study condition and counterbalancing</td>
</tr>
<tr>
<td>8</td>
<td>Means, standard deviations, and correlations across the entire performance episode</td>
</tr>
<tr>
<td>9</td>
<td>Cross-correlation analyses for focal variables across all participants</td>
</tr>
<tr>
<td>10</td>
<td>Cross-correlation analyses for focal variables for service failure and service recovery conditions</td>
</tr>
<tr>
<td>11</td>
<td>Cross-correlation analyses within phases of the performance episode across all participants</td>
</tr>
<tr>
<td>12</td>
<td>Cross-correlation analyses within phases of performance episode split by condition</td>
</tr>
<tr>
<td>13</td>
<td>Means, standard deviations, and correlations for main focal variables split by condition</td>
</tr>
<tr>
<td>14</td>
<td>Means, standard deviations, and correlations for each phase of the performance episode</td>
</tr>
</tbody>
</table>
Means, standard deviations, and correlations for each phase of the performance episode split by condition ..........................................................100
Correlations between average continuous surface and deep acting and post-simulation surface and deep acting ......................................................103
Means and standard deviations for participant emotionality, surface acting, and deep acting across performance episode by study condition ...........105
Proportions of time spent utilizing surface acting within performance episode by study condition .................................................................110
Proportions of time spent utilizing deep acting within performance episode by study condition ........................................................................114
Participant emotionality when neither surface nor deep acting are utilized ....117
Switching between surface acting and deep acting during the performance episode ......................................................................................120
Proportion of time spent utilizing surface or deep acting a greater amount ....124
Between-person correlations of focal variables with emotional exhaustion ......126
Regression models for emotional exhaustion across the entire episode split by study condition ........................................................................130
Regression models for emotional exhaustion by condition and phase of episode ..........................................................................................132
Correlations between emotional exhaustion and proportion of time spent surface/deep acting different levels ......................................................134
Correlations between emotional exhaustion and switching between regulation strategies .............................................................................136
Cross-correlations for third-party affect ratings .........................................137
Switching between emotion regulation strategies and third-party affect ratings ..............................................................................................141
Unconditional means models for focal variables ........................................143
HLM estimates for growth curve models ....................................................150
Average amount of change per second in participant emotionality, surface acting, and deep acting ratings ......................................................... 156

Summary of results .................................................................................. 159
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hypothetical relationship between surface acting and deep acting across two service episodes (one difficult [left panel] and one easy [right panel]).</td>
</tr>
<tr>
<td>2</td>
<td>Three-phase structure of confederate call.</td>
</tr>
<tr>
<td>3</td>
<td>Hypothetical model of emotion regulation variation and discontinuous change during the service episode.</td>
</tr>
<tr>
<td>4</td>
<td>Visual representation of the laboratory call center simulation.</td>
</tr>
<tr>
<td>5</td>
<td>Visual representation of counterbalancing procedure.</td>
</tr>
<tr>
<td>6</td>
<td>Participant emotionality instructions.</td>
</tr>
<tr>
<td>7</td>
<td>Participant emotionality rating screen.</td>
</tr>
<tr>
<td>8</td>
<td>Surface acting instructions.</td>
</tr>
<tr>
<td>9</td>
<td>Surface acting rating screen.</td>
</tr>
<tr>
<td>10</td>
<td>Deep acting instructions.</td>
</tr>
<tr>
<td>11</td>
<td>Deep acting rating scale.</td>
</tr>
<tr>
<td>12</td>
<td>Call segmenting instructions.</td>
</tr>
<tr>
<td>13</td>
<td>Call segmenting rating scale.</td>
</tr>
<tr>
<td>14</td>
<td>Third-party affect rating instructions.</td>
</tr>
<tr>
<td>15</td>
<td>Third-party affect rating screen.</td>
</tr>
<tr>
<td>16</td>
<td>Interaction between phase of confederate call and study condition predicting changes in participant emotionality.</td>
</tr>
</tbody>
</table>
17 Interaction between phase of confederate call and study condition predicting changes in surface acting (SA) .................................................................108

18 Interaction between phase of confederate call and study condition predicting changes in deep acting (DA) .................................................................112

19 Amount of switching ‘in general’ split by condition and phase of call ..........123

20 Within-person experience of emotionality, surface acting, and deep acting for the service failure condition .................................................................144

21 Within-person experience of emotionality, surface acting, and deep acting for the service recovery condition .................................................................145

22 Examples of variability in participant emotionality ratings .......................147

23 Examples of variability in participant surface acting ratings .......................148

24 Examples of variability in participant deep acting ratings .........................149

25 Growth curve model for participant emotionality ....................................151

26 Growth curve model for surface acting ....................................................153

27 Growth curve model for deep acting .......................................................154
CHAPTER I

STATEMENT OF THE PROBLEM

Emotional labor, or the study of employees managing their emotions to conform to organizational expectations (Rafaeli & Sutton, 1987), has become one of the key domains of emotions research in organizations (Barsade, Brief, & Spartaro, 2003; Grandey, 2008; Hochschild, 1983). As a general standard, employees in most service occupations are required to engage in “service with a smile,” where positive displays (i.e., smiling, positive vocal quality, eye contact) are encouraged and negative displays (i.e., anger, frustration) are discouraged (Fineman, 2003; Hochschild, 1983; Rafaeli & Sutton, 1987). These norms are often conveyed in organizations via emotional display rules which help communicate to employees what emotional displays are considered part of effective job performance (Ekman, 1973; Wharton & Erickson, 1993). In most cases, organizations desire positive expressions regardless of whether employees feel positive or not; after all, the central aim of emotional labor is for employees to induce organizationally-desired emotions in others (Grandey, Diefendorff, & Rupp, 2013) in order to meet broader performance objectives (e.g., service quality, sales). Thus, employees are, at times, forced to use regulation strategies that enable them to display the emotions that are being required by the organization (Grandey, 2000; Hochschild, 1983).

Researchers have outlined frameworks for how emotional labor operates, looking at emotional labor primarily as employees utilizing different types of regulation strategies
such as surface acting (i.e., displaying required emotions via “faking” positive emotions and suppressing negative emotions; internal feelings may be incongruent with organizational expectations) or deep acting (i.e., modifying internal emotions to align with organizational expectations) (Grandey, 2000; Hochschild, 1983). A great deal of research has outlined these processes, finding deep acting tends to be good for employee well-being and customer outcomes (e.g., Brotheridge & Lee, 2002; Chi, Grandey, Diamond, Krimmel, 2011; Grandey, 2003; Groth, Hennig-Thurau, & Walsh, 2009), whereas surface acting tends to be more detrimental to well-being and customer perceptions of the service experience (e.g., Brotheridge & Lee, 2002; Grandey, 2003; Hochschild, 1983; Rafaeli & Sutton, 1987). Thus, when deep acting, employees make a conscious effort to change their internal feelings to align with organizational demands; surface acting, conversely, occurs when employees display a positive expression that may be incongruent with internal feelings because it is expected as part of their job.

While there have been many advancements in emotional labor research, one limitation has been that the majority of studies utilize cross-sectional research designs, capturing only typical levels of emotional regulation via one-time self-report measures (e.g., Brotheridge & Lee, 2002; Brotheridge & Grandey, 2002; Diefendorff, Richard, & Yang, 2008; Grandey, 2003; Grandey, Fisk, & Steiner, 2005). Such general measures likely assess semantic information about emotion regulation that is filtered through the self and other knowledge structures (e.g., Cropanzano, Weiss, Hale, & Reb, 2003; Menon & Dubé, 2000), instead of episodic information about emotion regulation performed in specific customer interactions. For example, Brotheridge and Lee (2002) reported in their cross-sectional assessment of service employees that “in general” levels of surface acting
were positively correlated with depersonalization (i.e., sense of estrangement/separation from oneself; Hochschild, 1983; Leidner, 1993; Sheldon, Ryan, Rawsthorne, & Hardi, 1997), emotional exhaustion (i.e., feelings of being emotionally used up; Wharton & Erickson, 1993), and felt inauthenticity (i.e., displaying an emotional façade that is not authentically oneself; Brotheridge & Lee, 2002). Conversely, “in general” levels of deep acting were negatively correlated with depersonalization. Other studies have followed suit, utilizing “on average” assessments to measure emotional labor (e.g., Brotheridge & Grandey, 2002; Grandey, 2003) in cross-sectional research designs, with some studies pairing this data with other source ratings of emotional labor processes (e.g., Gosserand & Diefendorff, 2005; Grandey, Fisk, Matilla, Jansen, & Sideman, 2005). Though valuable, this work does not capture the dynamic variability that might exist within the emotional labor process.

In an attempt to address this limitation, studies have begun to examine emotional labor and related processes using experience sampling methodology (ESM; e.g., Beal, Trougakos, Weiss, & Green, 2006; Diefendorff, Becker, & Yang, under review; Judge, Woolf, & Hurst, 2009; Totterdell & Holman, 2003). Such approaches attempt to capture psychological and contextual factors that affect life as it is lived (e.g., Diener, Larsen, & Emmons, 1984), asking participants to provide ratings of how they feel at the moment, or to report on events and psychological processes that have just occurred. For instance, Diefendorff et al. (under review) used ESM to ask call center workers to rate various emotional labor constructs (including emotion regulation strategies) for the most recent call they had completed. Such ratings of the previous call are believed to better represent episodic information and to be less influenced by semantic or conceptual information.
Although utilizing ESM as a measurement approach is better at capturing event-level processes involved in emotional labor, such methods still lack the precision to adequately capture micro-level processes that are believed to be occurring when employees engage in emotional labor (e.g., Beal & Trougakos, 2013). For instance, while Diefendorff et al. (under review) asked employees to rate the extent to which they engaged in surface acting or deep acting during the previous call they attended to, this measurement technique does not capture dynamic emotion regulation processes such as (a) when during the call the emotion regulation strategies were used, (b) what specific circumstances triggered the use of each strategy, (c) how long the strategies were used, (d) whether participants vacillated between strategies, (e) whether participants used more than one strategy at any given point in time, or (f) whether there were particular sequencings or progressions of strategies that were more likely to occur than others.

A recent review by Beal and Trougakos (2013) highlighted the points outlined above, stating that no research to date has studied whether or not employees switch between regulation strategies within a single performance episode. Though Scott, Barnes, and Wagner (2012) did attempt to introduce the idea of “emotional labor variability” into the literature, their study relied on experience-sampled data, using within-person standard deviations of surface acting and deep acting as person-level predictors to reflect emotional labor variability. Thus, actual momentary research is still lacking. Beal and Trougakos further highlighted that correlational results in the literature show high, positive correlations between emotion regulation strategies at the within-person level of analysis. For instance, Beal, Trougakos, Dalal, and Weiss (2011) found within-person correlations ranging from .35 to .61 between reports of faked, hidden, and reappraised
emotions, raising the possibility that different emotion regulation strategies occur in tandem. Beal and Trougakos concluded that such results suggest that “individuals often use multiple strategies during a single episode” (p. 45). This idea is consistent with the dynamic, self-regulatory view of emotional labor proposed by Diefendorff and Gosserand (2003; see also Diefendorff & Richard, 2008) who argued for testing the links among emotional labor constructs at the within-event level of analysis, as well as theoretical work outlined by Côté (2005) stating that emotion regulation is a dynamic process that is dependent upon information from the broader social context in service exchanges. Integrating these ideas, within-event emotional labor processes likely function according to dynamic self-regulatory principles (e.g., discrepancy detection and discrepancy reduction) to determine when and why particular emotion regulation strategies are adopted (Côté, 2005; Diefendorff & Gosserand, 2003).

As an example, a situation that could create the need for dynamic changes in emotion regulation would be an interaction with a rude customer, an instance that has been well documented in the emotional labor literature (e.g., Goldberg & Grandey, 2007; Grandey, Dickter, & Sin, 2004; Rupp & Spencer, 2006; Sliter, Jex, Wolford, & McInerney, 2010). According to Beal and Trougakos (2013), employees in these situations are likely to detect the negative emotions of the rude/hostile customer and choose a regulation strategy accordingly (e.g., suppress their own negative emotions and fake a positive display). However, within that episode, employees may reevaluate the extent to which their chosen regulation strategy is working. According to Côté (2005), this reevaluation process is likely to be in response to the emotional, or social, feedback that employees receive from the customers (i.e., positive emotional reactions are
indicative of service success, whereas negative emotional reactions are indicative of service failure). In the event that employees detect a regulation failure, they are likely to switch to another regulation strategy to try to appease the customer (e.g., attempting to reappraise the situation to more genuinely experience desired emotions; Diefendorff & Richard, 2008), whereas detection of a success could result in employees retaining the emotion regulation strategy that was used when the positive emotional reaction from the customer occurred.

Given that no research has attempted to address this type of within-episode regulation, the current study was designed to address this theoretical gap. Continuous rating methodology (e.g., Larsen & Fredrickson, 1999; Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005; Mauss, Shallcross, Troy, et al., 2011; Ruef & Levenson, 2007) was used to collect within-episode, continuous ratings of surface acting and deep acting from interactions with customers in a simulated call center (modeled on recent work by McCance, Nye, Wang, Jones, and Chiu [2013]). The continuous rating methodology provided the precision needed to take a first look at how emotion regulation unfolds and progresses within a single performance episode, allowing for better documentation of how (and when) employees opt to change emotion regulation strategies during a service episode. In addition to documenting this dynamic process for the first time, there are three substantive questions the current study also addressed.

The first question pertained to gaining a better understanding of the relationship between surface acting and deep acting. Much of the theoretical discussion of surface acting and deep acting describes the constructs as being “opposite” or “contrasting” strategies for managing emotions, often treating them as though they exist along a
continuum from surface to deep acting (e.g., Austin, Dore, & O’Donovan, 2008; Kruml & Geddes, 2000; Zapf, 2002; Zapf, Vogt, Seifert, Mertini, & Isic, 1999). For example, a recent meta-analysis by Mesmer-Magnus, DeChurch, and Wax (2012) explicated a continuum of emotional disconcordance (i.e., surface acting) and emotional congruence (i.e., deep acting) in analyzing the relationships between emotion regulation and emotional labor outcomes. Such a view suggests that surface acting and deep acting should be negatively correlated because they are, in essence, antipodes (i.e., opposite in nature). However, a recent meta-analysis by Hülsheger and Schewe (2011) estimated that the population correlation between surface and deep acting was .22. If surface acting and deep acting are separate constructs on opposite sides of a continuum, how could such a positive population correlation be possible?

One potential explanation lies in the possibility that all research data up to this point (including both “in general” and ESM emotional labor assessments) have not been at the level of measurement that would enable a negative correlation between surface acting and deep acting to be observed. That is, a negative correlation may occur at the within-episode level where employees likely can only engage in one regulation strategy at a time, whereas the positive correlation may be emerging in the emotional labor literature due to courser levels of measurement. For example, a positive, event-level correlation using ESM could be due to a person vacillating between both approaches during the actual episode (i.e., a negative relationship), with the end-of-event ratings reflecting something about the difficulty of the encounter (high levels of both strategies may occur for difficult events, whereas low levels of both strategies may occur for easy events) and not the true covariation of the strategies at a given point in time. At the
person-level, a positive correlation could be due to stable individual differences, such as negative affectivity (e.g., high levels of which will result in the need to surface act and deep act more often), or stable situational demands such as occupational demands (e.g., high levels of display rules or demanding customers [e.g., patients interacting with nurses] will result in the need to surface act and deep act more often), rather than the strategies being used together.

As such, one thought is that continuous ratings could produce the elusive negative correlation between surface acting and deep acting. Of course, continuous ratings of emotional labor could also result in a positive relationship between surface acting and deep acting, suggesting that, counter to implicit assumptions in the emotional labor literature, individuals can and do use both strategies simultaneously. Such a finding would be a conceptual contribution given that it would suggest that homologous processes occur across multiple levels of analysis (e.g., person, event, and within-event levels; Chen, Bliese, & Mathieu, 2005). Thus, the use of continuous within-event ratings may help clarify the true nature of the relationship between the constructs. To test the coherence between surface acting and deep acting, cross-correlations for each participant can be calculated. Cross-correlations measure the extent to which two measures (i.e., surface acting, deep acting) covary together across time (i.e., a customer call) within a given individual (i.e., a participant) (Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005; Levenson, 1988). Further, cross-correlations account for lags between measures across a given time frame. More detail on the purpose and use of cross-correlations is discussed in Chapter III.
The second key question pertained to the impact of social information on emotion regulation strategies. Côté (2005) created the social interaction model of emotion regulation that outlined how employees and customers separately affect the emotion regulation process of the employee in a dynamic manner. According to Côté, characteristics of the social environment, such as customer emotional displays and customer reactions to employee emotional displays, can shape the emotion regulation techniques used by the employee (see Chapter II for more detail). For example, negative emotional reactions from the customer in response to the employee may signal a service failure (i.e., being unhappy with the service encounter, the quality of service being provided, the employee’s emotional displays, and so forth). Conversely, positive emotional reactions could be indicative of a service success or, following a service failure, a service recovery, thus conveying to the employee that he/she should continue to regulate his/her emotions in the manner he/she currently is since it appears to be working effectively. Both “in general,” cross-sectional surveys and ESM are ill equipped to provide evidence of such within-episode switching given their focus on person-level and interaction-level processes. Thus, in the current study, by combining a between-person manipulation of customer behavior (i.e., changes from neutral to negative affect within a service episode) with continuous emotion regulation ratings, a closer test of the social interaction model of emotion regulation and the role of customer reactions in the emotional labor process was possible.

Finally, the third substantive question focused on testing the relationship of these emotion regulation strategies with third-party observer ratings of employee positive and negative emotional displays, as well as participant ratings of the well-being outcome of
emotional exhaustion. Although these outcomes have been examined in many prior studies (e.g., Bono & Vey, 2007; Brotheidge & Grandey, 2002; Brotheridge & Lee, 2002; Diefendorff, Croyle, & Gosserand, 2005; Erickson & Wharton, 1997; Goldberg & Grandey, 2007; Hochschild, 1983; Morris & Feldman, 1996; Wharton, 1999), the current study’s use of real-time, within-episode ratings of surface acting and deep acting provides a unique opportunity to examine these links. In particular, in addition to obtaining continuous ratings of emotion regulation, continuous ratings of employee emotional vocal expressions by third-party observers (i.e., trained raters) will be obtained. Given that deep acting has been typically believed to equate to “real” emotional displays (i.e., reappraising a situation to truly feel positive emotions, thus displaying real positive emotions; Grandey, 2000; Hochschild, 1983) and surface acting has been thought to represent a “leakier” strategy that reveals true feelings (i.e., displaying a smile, but having negative affective tone leak out in vocal quality/content; Ekman & Friesen, 1969; Elfenbein & Ambady, 2002), one may suspect that deep acting would correspond to third-party ratings of positive emotions, while surface acting may equate to less positive ratings of emotional expressions. Although research on emotional labor has attempted to assess the relationship between surface and deep acting and emotional performance (see Hülsheger and Schewe’s [2011] meta-analysis), limited work is available where third-party “customer” ratings are obtained. Moreover, no work has attempted to relate in vivo ratings of emotion regulation to independent evaluations of emotional displays. As such, the current study attempted to examine the correspondence of emotion regulation strategies with independent ratings of verbal emotional content.
Additionally, the continuous emotion regulation ratings were linked to “end of episode” self-reports of emotional exhaustion, an outcome commonly considered in the emotional labor literature (e.g., Brotheridge & Lee, 2002; Wharton & Erickson, 1993). Hülsheger and Schewe’s (2011) meta-analysis reported strong correlations between surface acting and emotional exhaustion (\( \rho = .44 \)), and a rather weak relationship between deep acting and emotional exhaustion (\( \rho = .09 \)), suggesting that deep acting may result in lower levels of emotional exhaustion for employees. However, emotional labor researchers have suggested that, while deep acting may be viewed as a more beneficial type of acting, it is still an effortful process that may have negative effects for employees (e.g., Hochschild, 1983). Thus, the current study sought to reexamine this relationship at the within-episode level with more complete data on the actual amount of surface acting and deep acting that occurs within a single interaction, as well as other temporal and dynamic factors that can be derived from such assessments (e.g., mean, proportion, within-strategy variation, switching between strategies). These ideas are discussed in more detail in the subsequent chapter.
Arlie Hochschild sparked the discussion of emotional labor in organizations in her seminal 1983 book, *The Managed Heart*, which explicated the daily emotional demands that service employees face. Specifically, Hochschild (1983) highlighted the process of emotional labor, where service employees engage in different types of acting (i.e., surface acting, deep acting) to align their emotional displays with organizational expectations (i.e., emotional display rules). Hochschild (1983) further pointed out that all occupations requiring emotional labor (e.g., call center employees, nurses, front-line food service employees) have three common criteria. First, they require contact with external members of the organization (e.g., customers, patients), via face-to-face (i.e., nurse to patient) or voice-to-voice (i.e., call center employee to caller) contact. Second, they require employees to manage their emotions in such a way that an organizationally-desired emotional response is evoked in the recipient (i.e., positive displays from the employee lead to positive emotions in the customers). Finally, in such occupations, the employer exhibits control over employee emotional displays, typically via organizational practices such as training, supervisor evaluations, rewards, and rules or regulations (Pugh, Diefendorff, & Moran, 2013).

The notion of emotional labor has become embedded within organizations, particularly with growing expectations for “service with a smile” (Grandey, 2000; Pugh,
2001). Part of this stems from the service quality expectations that many customers have, in which customers desire service interactions that are not just efficient, but also possess a level of courtesy (e.g., politeness, respect, friendliness; Parasuraman, Zeithmal, & Berry, 1985). In turn, many organizations believe that customers who are satisfied with the service experience will be more likely to return to the organization and recommend the store to others (e.g., Pugh, 2001; Tsai, 2001; Tsai & Huang, 2002), thus creating a bottom-line impact. Fineman (2003) stated that organizations are beginning to take extreme measures, citing an example from McDonald’s corporation in which all members of management attend “Hamburger University” where they are instructed on how to gain employee compliance with emotional expectations. Given that service with a smile has become the norm and not the exception, Hochschild (1983) opined that employees are often placed in problematic situations in which their feelings are dissonant with what is being mandated by external factors from the organization. Ashford and Tomiuk (2000) came to a similar conclusion in their qualitative study of service agents, stating that the norms from organization emotional demands often conflict with internal feelings, creating a state of internal tension that service employees must cope with.

Surface Acting and Deep Acting

Within the emotional labor literature, a primary focus has been on how employees engage in the emotion regulation process. Specifically, the focus of empirical research has often been on the techniques or strategies employees use to handle emotional display rule expectations from the organization. As outlined by Hochschild (1983) and many others (Brotheridge & Grandey, 2002; Brotheridge & Lee, 2002; Diefendorff et al, 2005; Grandey, 2000; Grandey, 2003), the two most commonly studied strategies are surface
acting and deep acting. Although more specific regulation strategies have been identified (see Gross, 1998a, 1998b), Grandey (2000) argued that these two broad categories of emotion regulation appear to capture the essence of the emotion management occurring in emotional labor contexts.

For surface acting, the primary goal is for employees to modify their outward emotional expressions to align with organizational expectations (Grandey, 2003; Hochschild, 1983; Rafaeli & Sutton, 1987). Specifically, when an employee is surface acting, he/she may engage in positive displays such as smiling, while still internally feeling negative emotions such as anger or frustration. Typically, surface acting is focused on strategies of suppression, as when the emotion being felt by the employee (i.e., anger, frustration) is hidden (i.e., Grandey, 2003; Gross, 1998a, 1998b), and faking (or simulation; Diefendorff, Richard, & Yang, 2008; Ekman & Friesen, 1975) in which employees express an emotion that is not felt (i.e., happiness). Thus, surface acting occurs when a person (a) experiences a negative emotion and expresses a positive emotion (i.e., suppression, faking), (b) experiences a negative emotion and expresses nothing (i.e., suppression), or (c) experiences a neutral emotion but expresses a positive emotion (i.e., faking). In all instances, employees show emotions that are different from their internal emotional state (Grandey, 2000; Totterdell & Holman, 2003); as such, surface acting is more of a response-focused strategy (Grandey, 2000). Surface acting can be performed in “good faith” or in “bad faith,” according to Ashforth and Humphrey (1993). Under good faith circumstances, employees engage in surface acting for the betterment of the organization or the service recipient, such as when employees smile because they know they have to for the success of the encounter (and ultimately the
Under bad faith circumstances, there is little regard for the service recipient; rather, surface acting is done out of obligation for organizational norms. Because the discrepancy between feelings and expressions remains for those who engage in surface acting, levels of emotion-display dissonance (Grandey et al., 2013; Diefendorff et al., under review) should be high.

Counter to surface acting, deep acting involves actively changing one’s internal feelings so that they match the expectations of the display rules, resulting in emotional displays that more naturally match the expectations (Grandey, 2003; Hochschild, 1983). Individuals engaging in deep acting adopt what Gross (1998a, 1998b) terms an antecedent-focused coping strategy, resulting in alignment between internal and external emotional states before the outward display begins. Similarly, Grandey (2003) describes deep acting as grounded in the idea of reappraisal, where individuals are making an attempt to reappraise the situation or their current emotions to shift them in line with organizational expectations (Gross, 1998a, 1998b). Rafaeli and Sutton (1987) claimed that deep acting involves “acting in good faith” (p. 32) since employees are making a concerted effort to feel what they are supposed to express; Ashforth and Humphrey (1993) resonated with this, stating that deep acting is most often a good faith effort to show organizationally-desired emotions. Thus, while surface acting focuses on modifying expressions, deep acting focuses on modifying internal states (Grandey et al., 2013).

Studies of emotional labor that focus on surface and deep acting have covered a broad range of methodologies and levels of analysis. In the following sections, I provide an overview of the studies conducted at the person- and event-level analysis to help clarify the advancements and shortcomings of these endeavors.
Person-Level Assessments of Emotion Regulation

The majority of emotional labor research has studied emotional labor processes at the person-level of analysis via cross-sectional, one-time assessments with service employees (e.g., Brotheridge & Grandey, 2002; Brotheridge & Lee, 2002; Grandey, 2003) or via qualitative, case study designs (e.g., Hochschild, 1983; Shuler & Sypher, 2000). For example, Brotheridge and Grandey’s (2002) study of 238 full-time service employees relied on cross-sectional data, measuring the extent to which employees engaged in surface acting and deep acting in general. These global surface and deep acting assessments were then related to other important constructs, such as the burnout dimensions of emotional exhaustion, depersonalization, and personal accomplishment. Surface acting positively correlated with emotional exhaustion and depersonalization, and negatively related to personal accomplishment. Deep acting, on the other hand, positively correlated with personal accomplishment, but was not significantly related to either emotional exhaustion or depersonalization.

Similarly, Brotheridge and Lee (2002) collected cross-sectional data from a convenience sample of 236 employees in a variety of service occupations by asking students to distribute surveys to family members/friends in service jobs. In a series of path analyses, Brotheridge and Lee concluded that surface acting affected emotional exhaustion via decreased feelings of personal authenticity (i.e., behavior that is not authored or true to oneself). Further, deep acting was positively related to the experience of personal authenticity, which reduced emotional exhaustion. Fitting with results from Brotheridge and Grandey (2002), these findings concluded that surface acting is detrimental, whereas deep acting is beneficial, for well-being.
Other cross-sectional research has attempted to gain “outsider” perspectives of the emotional labor process. For example, Grandey (2003) collected self-reports of “in general” surface acting and deep acting (and other constructs) from administrative assistants and correlated these measures with coworker ratings of affective delivery (i.e., ratings of the level of sincerity, enthusiasm, warmth, friendliness, and courtesy of the employee when interacting with customers). Grandey found that surface acting was again positively related to emotional exhaustion, whereas there was no effect of deep acting. Further, surface acting was negatively related to coworker ratings of affective delivery, while deep acting was positively related to affective delivery ratings. Grandey concluded that her results helped demonstrate that deep acting was more beneficial than surface acting not only for employee well-being, but also for emotional performance.

Hülsheger and Schewe (2011) conducted a meta-analysis summarizing many of the findings from the cross-sectional emotional labor literature (for another meta-analysis, see Bono & Vey, 2005). Hülsheger and Schewe, when looking at just surface acting, found positive, population-level correlations with emotional exhaustion ($\rho = .44$), depersonalization ($\rho = .48$), psychosomatic complaints ($\rho = .44$) and psychological strain ($\rho = .42$). Additionally, they reported significant negative relationships with positive outcomes that they classified as “job-related well-being” such as job satisfaction ($\rho = - .33$) and organizational attachment ($\rho = -.31$). Further, negative relationships emerged between surface acting and task performance ($\rho = -.11$) and emotional performance ($\rho = -.14$). Conversely, when looking at deep acting, Hülsheger and Schewe found positive relationships with personal accomplishment ($\rho = .27$), emotional performance ($\rho = .18$) and customer satisfaction ($\rho = .37$). Interestingly, deep acting was related positively to
psychosomatic complaints ($\rho = .18$), which would be counter to expectations if deep acting is as beneficial as it seems. Such a finding, paired with lack of relationships between deep acting and other outcomes (i.e., emotional exhaustion) suggests that there might be more to the emotional labor story.

Integrating the above results, it appears as though surface acting is associated with lower well-being via processes such as depersonalization and reduced authenticity. Thus, based upon the inherent nature of surface acting (i.e., having to fake or cover up internal feelings), employees likely experience feelings of disconnect between internal emotions and their emotional expressions at work. This dissonance could make interactions with customers and clientele highly challenging, thus increasing employee emotional exhaustion and reducing the likelihood that customers are satisfied with their service experience. Conversely, deep acting appears to be beneficial, given that it is often linked to increased personal accomplishment and customer evaluations and largely unrelated to indicators of ill-being. However, as Grandey (2003) pointed out despite her positive findings, deep acting can be a highly effortful process given the need for employees to consciously alter their emotional experience. Further, the results from Hülsheger and Schewe’s (2011) meta-analysis provide mixed results for how deep acting operates; one would typically expect no negative outcomes from deep acting, yet a significant, positive relationship with psychosomatic complaints occurred. What might explain such incongruent findings?

Part of the problem may lie in the use of “in general,” cross-sectional assessments that do not directly tap into the event-level and within-person, dynamic emotional labor process that researchers have opined likely exists (e.g., Beal & Trougakos, 2013;
Diefendorff & Gosserand, 2003). In the following section, I review recent research aimed at addressing this shortcoming in emotional labor research by using experience sampling methodology (ESM) to assess event-level emotion regulation.

**Within-Person Assessments of Emotion Regulation**

As previously stated, the aforementioned cross-sectional research on emotional labor made substantial contributions to our understanding of how different emotional labor strategies (i.e., surface acting, deep acting) influence individuals. However, there has been a trend in the organizational psychology literature to focus more on within-person variability. Dalal and Hulin (2008) provided an extensive review of the arguments for focusing on both between-person and within-person variance in psychological constructs, stating that far too often researchers ignore within-person variance that has significant implications for understanding constructs of interest. Indeed, numerous researchers have found that meaningful within-person variance does exist in affect-related constructs, such as affective responses to daily-task satisfaction (Gabriel, Diefendorff, & Erickson, 2011), performance feedback (Ilies & Judge, 2005), and goal attainment (Harris, Daniels, & Briner, 2003; Henkel & Hinsz, 2004) as well as job satisfaction in relation to interpersonal and informational justice (Loi, Yang, & Diefendorff, 2009).

Diefendorff and Gosserand (2003) made an explicit call to emotional labor researchers to model within-person variation in emotional labor constructs by repeatedly assessing key constructs at the event-level of analysis. Specifically, they argued that static measures of emotional labor variables, “provide an understanding of the average level of variables, but ignore important within-person variability” (p. 957). Their treatment of
emotional labor in a control theory framework aims to articulate the mechanisms and processes by which emotional labor can unfold over time, describing how an individual’s emotional experience at Time 1 may influence the person’s emotional displays at Time 2. Numerous studies have since utilized event-level assessments of emotional labor constructs.

For example, Totterdell and Holman (2003) tested Grandey’s (2000) conceptual model of emotional labor with an experience sampling study. Data were collected from 18 employees who filled out surveys via palm pilots four times a day for two consecutive work weeks, resulting in a total of 537 event-level responses. The goal of the study was to identify various antecedents and consequences of the emotional labor process. At the event-level, Totterdell and Holman collected measures of emotion regulation, with deep acting conceptualized multidimensionally as positive refocus (i.e., attentional deployment; “thought about pleasant things;” p. 59) and perspective taking (i.e., cognitive change; “thought about how the customers feel;” p. 59) and surface acting conceptualized as faking emotions (i.e., response modulation; “expressed feelings you did not feel;” p. 59). Totterdell and Holman (2003) found that negative events involving customers predicted faking (surface acting); however, they also found that positive events (as opposed to negative events) from customers and coworkers were associated with greater use of positive refocus (deep acting). Positive events from customers were also associated with more perspective taking (deep acting). Given that negative events are thought to elicit more emotion regulation, since employees need to handle the disconnect between their own negative emotions and the positive emotional display rules required by the job, these findings were counter to emotional labor theory predictions. However, they
do show that emotion regulation (i.e., deep acting) can occur even during positive work interactions.

In terms of well-being and performance outcomes, Totterdell and Holman (2003) found surface acting was positively related to feeling emotionally drained and numb, whereas deep acting was not linked to either well-being outcome. Focusing just on performance outcomes (which are limited in value, as they are ratings provided by the employee), the researchers found that both types of deep acting measures were related to quality of job performance (i.e., ratings of performance on a continuum of good to bad) and displayed enthusiasm (i.e., ratings of enthusiasm on a continuum of bored to enthused), but only positive refocus was related to display happiness (i.e., ratings of happiness on a continuum of unhappy to happy); there were no relationships between these variables and surface acting. However, Totterdell and Holman (2003) did find that both perspective taking (deep acting) and faking emotions (surface acting) were positively related to proactivity in helping customers (i.e., ratings of how much employees believed they “put themselves out to help customers;” p. 60). The findings of positive performance outcomes for both surface and deep acting techniques again paints a complex picture for how surface and deep acting strategies may operate beyond the “typical” expectations of the two strategies.

Judge, Woolf, and Hurst’s (2009) study also considered within-person, event-level emotion regulation strategies. In their study, service employees from a variety of organizations provided measures of surface acting, deep acting, job satisfaction, emotional exhaustion, and mood state at the end of the working day for seven consecutive working days. At the within-person level of analysis, results revealed that
surface acting positively predicted both emotional exhaustion and negative affect and also negatively predicted job satisfaction. Deep acting was unrelated to both emotional exhaustion and job satisfaction. However, deep acting did relate not only to decreases in negative affect (which was expected), but also to decreases in positive affect, which was counter to expectations. Although many of these findings were not new to the emotional labor literature (i.e., linking surface acting to emotional exhaustion, decreased job satisfaction, etc.) they were important for showing these relationships at the within-person level. Further, they demonstrated yet again the complex picture surrounding deep acting, in which the strategy may not always be beneficial.

Using end-of-day assessments similar to those used by Judge et al. (2009), Scott and Barnes (2011) utilized a sample of bus drivers and tested a mediational model in which surface acting was proposed to lead to increased negative and decreased positive affect, which was expected to subsequently influence work withdrawal (e.g., behaviors indicative of being “removed” from the workplace, such as putting in less effort, thinking about leaving, daydreaming). Conversely, deep acting was proposed to positively relate to positive affect and to negatively relate to negative affect, thus leading to lower levels of work withdrawal. Their results indicated that surface acting was positively related to negative affect, while deep acting was positively associated with positive affect and negatively associated with negative affect. The authors concluded that their results supported the idea that surface acting has a deleterious effect on daily affective states for employees. Further, upon testing for indirect effects in linking surface and deep acting to work withdrawal, Scott and Barnes found that both surface and deep acting were indirectly related to withdrawal via negative, but not positive, affect. Thus, while deep
acting may lead to improved mood states, both strategies still have the potential to negatively affect withdrawal at the within-person level via negative affect.

As another example, Diefendorff et al. (under review) collected data from call center employees in China three times a day (immediately after the completion of specific calls) for ten consecutive working days, measuring factors such as surface acting, deep acting, emotion-rule dissonance, and customer difficulty. They found that event-level measures of surface acting were associated with higher felt inauthenticity, emotional deviance, emotion-display dissonance, and emotional exhaustion. Interestingly, there were no main effects of deep acting, which again fits with the rather limited findings that surround deep acting in comparison to surface acting in the emotional labor literature.

The existing ESM studies answer the call to focus on within-person processes, but they still do not capture the complexity of event-level emotional labor. For instance, Totterdell and Holman (2003) assessed surface and deep acting in two-hour increments (i.e., ratings every two hours reflected upon events that occurred in the previous two hours) within a given working day. This differs from techniques used by Judge et al. (2009) and Scott and Barnes (2011) who measured emotion regulation at the end of the day, and Diefendorff et al. (under review) who measured regulation at the end of a specific call with a customer. Thus, none of these studies, despite their methodological advancements, capture potential within-episode variance in emotion regulation. More specifically, as stated previously (see Chapter I), even event-level measures of surface and deep acting cannot address when during a specific encounter surface or deep acting are used. Though Scott et al. (2012) introduced the idea of “emotional labor variability”
into the emotional labor literature, their reliance on within-person standard deviations of surface acting and deep acting in an ESM study failed to demonstrate how both strategies unfold over time. That is, they treated within-person standard deviations as person-level predictors of outcomes, which still do not look at true momentary variability. Further, event-level measurement fails to consider what within the encounter may trigger the use of different types of emotion regulation (i.e., negative comment from customer, task demand, etc.), in addition to understanding how surface and deep acting “operate” within a single encounter (i.e., how long each strategy is used, if participants vacillate between strategies, if participants use multiple strategies at a given time, and so forth).

Therefore, regardless of the advancements to the field of emotional labor that these studies provide, they do not provide information on within-episode fluctuations in emotion regulation. When considering original work on affective events, Weiss and Cropanzano (1996) urged researchers to pay greater attention to aspects of events that can shift emotions; in the case of the current study, more attention to aspects of the situation that can shift emotion regulation is needed. Studies that use daily measures at one specific time-point (e.g., end of day, every two hours, after one call) fail to capture the variance in emotion regulation and miss whether this variance occurs in relation to specific work events. Recognizing this limitation in the emotional labor literature, Beal and Trougakos (2013) stated that researchers in the field of emotional labor have yet to fully consider and test whether or not individuals can fluctuate between emotion regulation strategies in a single service episode.

Thus, the question still remains whether individuals are capable of utilizing multiple emotion regulation techniques (e.g., suppression, reappraisal) within a single
episode, as opposed to reflecting on the whole day (Scott & Barnes, 2011) or even a
specific customer interaction (Diefendorff et al., *under review*). Indeed, it is not clear
how event-level measurement techniques (e.g., using a palm pilot) could capture such
within-episode fluctuations. To address this limitation, I used a continuous, uninterrupted
measure of emotion regulation that spans the entire regulation episode. In the subsequent
section, I review literature on the use of continuous measures in emotions research and
how this procedure was applied to the assessment of emotion regulation.

Continuous Rating Methodology

While the use of continuous ratings may be new to the emotional labor literature,
the approach is not new to the broader study of emotions. Larsen and Fredrickson (1999)
presented an extensive review of different types of emotion measures (e.g., self-report,
physiological, observer coding of emotions) and presented the shortcomings and
problems associated with each. Most applicable to the study of emotional labor is their
critique of self-reports. Specifically, Larsen and Fredrickson made it clear that both
single-item measures of emotions or multi-item measures of emotions are flawed in three
key ways: 1) global evaluations of emotions are subject to distortion and measurement
error, 2) making such static ratings can interrupt the “flow” of the emotional experience,
and 3) global assessments fail to address the duration of an emotion response (e.g.,
Fredrickson & Kahneman, 1993). As an example, Ruef and Levenson (2007) cited the
Positive and Negative Affectivity Schedule (PANAS; Watson, Clark, & Tellegen, 1988),
a scale that has been validated and used extensively throughout the literature across
multiple areas of psychology. While the PANAS is easy for participants to complete in
terms of time and effort, Ruef and Levenson stated that using such a scale in an
experiment can interrupt the flow of experimental procedures (i.e., having to administer the PANAS at multiple times throughout the experiment, such as between confederate calls in a call center simulation) and the scale may not be capable of capturing fluctuations between emotional states (i.e., changing from feelings of anger to feelings of pleasantness).

Accordingly, Larsen and Fredrickson (1999) proposed that a way to avoid these errors in measurement is to move towards “real-time ratings” in which data are collected “on a moment-by-moment basis, either on-line as the emotion is first experienced or retrospectively as the temporal dimension of the original episode is ‘replayed’ while real-time momentary self-report measures [are collected]” (p. 47). To collect such temporal ratings, Larsen and Fredrickson suggest that researchers should adapt their single-item measures by adding a continuous rating dial component or a sliding measure. Making this addition allows respondents to adjust the rating dial/slider as often as needed to reflect their current emotional state in relation to an event being experienced. The resulting data present unique analytic opportunities to the researcher, allowing for both nomothetic data analysis approaches (e.g., regression, correlation) and ideographic representation (e.g., charting of the emotion ratings throughout a given experience; Ruef & Levenson, 2007).

As an example of how this method can be used, recent work by Mauss, Levenson, McCarter, Wilhelm, and Gross (2005) focused on the coherence among continuous ratings of emotions (amusement, sadness), physiological indicators (heart rate, finger pulse amplitude, finger pulse transit time, blood volume, skin conductance level, somatic activity), and facial expressions when participants viewed movies. By collecting separate continuous ratings of amusement and sadness (i.e., having participants first rate
amusement/sadness, and then rate sadness/amusement) and collecting coded observations of facial behavior, Mauss and colleagues were able to map levels of coherence between self-reported amusement and positive facial expressions, as well as between self-reported sadness and negative facial expressions. Additionally, Mauss et al. calculated the level of coherence among the amusement and sadness experiences and their physiological indicators. Mauss et al. determined coherence between measures by calculating the average cross-correlations, which “index the average extent to which two measures covary across time within individuals” (p. 181). Based upon the results, Mauss et al. concluded that emotions play a critical role in linking emotional, behavioral, and physiological systems, and further claimed that their measures of continuous emotions had a strong advantage over retrospective emotion ratings through the removal of measurement error due to memory biases and the ability to measure emotions without interrupting the process.

Mauss et al. (2011) used similar real-time rating dial assessments to investigate the effects of a disconnect between expressed emotion behavior (continuously rated by others) and emotion experience (continuously rated by the self) on outcomes such as depressive symptoms and well-being. In a laboratory setting, participants provided continuous ratings of their emotional experience while watching a positive emotion inducing video clip. All participants were videotaped during the experimental session in order to record their physical behaviors, which were rated by observers. Dissociations between behaviors and emotions (i.e., the difference between self-ratings and observer ratings) were then calculated via cross-correlations between observed participant behavior and the participant’s continuous ratings within a range of -10 to +10 seconds.
This time frame was chosen “because theoretical considerations led [them] to expect that there might not be meaningful time lags among measures not greater than 10 seconds in either direction” (Mauss et al., p. 740). For each participant, this resulted in 21 cross-correlations, of which Mauss et al. chose the highest value “because it likely represents the most accurate index of association” (p. 740; see also Mauss et al., 2005). Results indicated that levels of dissociation between continuous assessments of emotions and facial behaviors predicted subsequent depressive symptoms and decreased well-being six months later. The authors concluded that incongruence between feelings and expressions, such as when one smiles when not feeling happy, can result in impaired well-being.

What is clear from the above studies is that continuous measures may have some advantages over person- and event-level measures of emotional labor. Continuous measurement of emotional labor processes would not have the time constraints nor reliance on retrospective memory that person-level and end-of-event-level measurements have, thus allowing such assessments to more accurately represent the emotional labor processes of individuals. More specifically, while person-level assessments likely tap into general self-conceptions of “typical” emotional labor and event-level assessments may tap into the average levels of regulation for a given event, continuous ratings would uniquely capture moment-to-moment fluctuations in experience. Beyond this advantage, Mauss et al. (2005) illustrated that multiple continuous ratings can accurately be collected even when a stimulus (i.e., a phone call, a movie) is replayed for a participant multiple times. For emotional labor, this would mean being able to accurately collect continuous ratings of deep acting and surface acting for the same, single performance episode.
Thus, I applied the same methodological principles outlined in the above continuous rating studies to collect continuous ratings of emotion regulation within a service episode. To date, no research has considered continuous ratings beyond emotional distinctions (e.g., happiness, sadness, amusement, etc.). Building from the suggestions of both Beal and Trougakos (2013) and Diefendorff and Gosserand (2003), I expected that continuous ratings may capture meaningful within-episode variation in surface acting and deep acting. In the current study, I used a call center simulation in which participants are exposed to two customer interactions that are designed to induce emotion regulation as well as to provide “social feedback” (e.g., Côté, 2005) regarding the effectiveness of participants’ emotional expressions. After the interaction was completed, participants listened to recordings of the calls and provided continuous ratings of their surface acting and deep acting during the interaction. Below, I further explicate the hypotheses I tested with continuous rating assessments of emotion regulation.

Within-Episode Dynamics in Emotional Labor

Using a within-person, within-episode continuous measurement approach to assessing surface acting and deep acting will enable the proposed study to answer three critical questions.

The Relationship Between Surface Acting and Deep Acting

The first question pertained to the true nature of the relationship between surface acting and deep acting. Much of the emotional labor literature assumes that surface acting and deep acting are polar opposites, with some theorists going as far as placing them on opposite ends of an emotional labor continuum (e.g., Austin et al., 2008; Kruml & Geddes, 2000; Mesmer-Magnus et al., 2012; Zapf, 2002; Zapf et al., 1999). Such a
relationship would be reflected by a negative correlation between simultaneous measures of surface acting and deep acting, such that when surface acting is high, deep acting is low (and vice versa). In essence, this perspective treats surface and deep acting as antipodes, in which employees are not able to perform both surface and deep acting at the same time when interacting with an employee.

However, despite the theoretical support for this idea, actual results show a positive relationship occurring at both the event-level and person-level of analysis. As stated previously, Hülsheger and Schewe (2011) found a population correlation between surface and deep acting of .22, which suggests that surface acting and deep acting may be more overlapping than the literature would suggest. Other experience sampling studies that have more accurately tried to capture daily emotion regulation also tend to find positive relationships between the constructs; two of the daily-level studies outlined above (Judge et al., 2009; Scott & Barnes, 2011) that looked at surface and deep acting at the daily level of analysis (i.e., end-of-day emotional labor evaluations) found positive correlations between the two strategies ($r = .20$ and $.07$, respectively).

Integrating these ideas, it becomes clear that there is disconnect between emotional labor theory (i.e., an employee can surface act or deep act) and what has been empirically demonstrated in the emotional labor literature (i.e., surface acting and deep acting covary). I contend that the inability to demonstrate the theoretical negative relationship between surface acting and deep acting may be due to limitations in the precision of existing measurement approaches. The psychological literature has produced several examples of how the nature of the relationship between two variables may change when moving across levels of analysis. For example, work by Vancouver, Thompson,
Tischner, and Putka (2002) found that the relationship between self-efficacy and performance was positive at the person-level of analysis, but negative at the event-level of analysis. Applying a similar framework, the negative relationship between surface and deep acting may be primarily observed when in vivo measures are employed, instead of the typical “end-of-event” or “in general” assessments used in the literature.

Theoretical support for the negative relationship between surface and deep acting at the within-episode level of measurement can be drawn from research related to goals and attentional processes. One theory, action identification theory (Vallacher & Wegner, 1987, 1989), states that individuals, when asked about the goals they are pursuing on a momentary basis, will only report one goal (though many may be activated; e.g., DeShon, Brown, & Greenis, 1996) due to limited attentional resources. Thus, if asked via in vivo assessments of emotion regulation what they are doing, it is likely that individuals will state that they are surface acting or deep acting, even though there may be varying levels of both types of emotion regulation occurring. This idea resonates with Allport’s (1987) selection-for-action process model, stating that individuals have limited attentional resources, thus creating the need for individuals to monitor their environments to switch between their goals (and goal-related actions) efficiently. Further, DeShon and Gillespie (2005) stated that “at any fixed time point, action is directed toward reducing a discrepancy on a single goal” (p. 1110). If one were to treat surface acting and deep acting as independent strategies (or alternative subgoals), this further builds support for individuals reporting doing one or the other, but not both simultaneously.

Thus, I proposed that when surface acting and deep acting are aggregated to a higher level of analysis (i.e., averaged across an entire performance episode) there would
be a positive correlation between the regulation strategies since people can report doing both over the course of the entire episode. However, from moment-to-moment within an episode, the relationship between surface acting and deep acting would be negative since individuals likely only have the attentional and cognitive resources to consciously pursue one regulation strategy at a time.

Figure 1 depicts how this type of relationship variation at the within-episode level and aggregated-episode level could exist; whether an individual is experiencing a highly stressful encounter (left panel) or low stress encounter (right panel), while both types of emotion regulation may occur, within-episode the two strategies will be enacting opposite of each other (i.e., not surface acting or deep acting at the same time). However, if one were to take the average of surface acting and deep acting across persons presented (i.e., average of surface acting and deep acting across each person’s entire episode), the two regulation strategies would be positively correlated. This helps clarify how theory could suggest a negative relationship, yet typical measures of surface acting and deep acting produce positive intercorrelations.
Thus, the following predictions were made:

_Hypothesis 1:_ At the momentary level of analysis, there is a negative correlation between surface acting and deep acting.

_Hypothesis 2:_ Episode-level mean surface acting and mean deep acting are positively correlated across subjects.

Moreover, in a more exploratory sense, this continuous-rating measurement approach will also enable the examination of the meaning of low levels of surface acting and deep acting. A vexing problem in the literature (Grandey et al., 2013) is the meaning of low levels of each strategy. High levels clearly indicate that a person is using the strategy and low levels indicate that he/she is not using the strategy, but this could be for one of three reasons. First, low levels of one strategy could mean that the person is using
the other strategy; this possibility can be captured with continuous assessments. Second, low levels of both strategies could mean that the person does not need to regulate his/her emotions, which would occur when the person naturally feels the emotions needed and can display appropriate emotions in the interaction (e.g., Diefendorff et al., 2005). Finally, low levels of both strategies could mean that the person should regulate his/her emotions (i.e., the feelings are different from display rules), but does not have the desire to do so, resulting in displays of inappropriate or negative emotions. I plan to explore the meaning of low levels of surface acting and deep acting in the process of examining the relationship between the strategies, in addition to linking these strategies to independent, third-party evaluations of employee hedonic tone. I return to this idea in a later section.

The Role of Social Feedback in Shaping Emotional Labor

Another advantage of utilizing continuous ratings of emotional labor is the ability to look at the influence of customer behaviors on changes in employee emotion regulation. To understand the influence customers may have, Côté’s (2005) social interaction model of emotion regulation provides a good starting point. Côté’s (2005) model highlights the dynamic, relational process of emotion regulation by exploring the influence of both the employee and the customer on the emotion regulation process. Specifically, Côté states that his social interaction model of emotion regulation is “based on social functional accounts of emotion positing that senders’ public displays of emotion communicate rich and important information to receivers during social encounters” (p. 514). In this model, the emotions sent by the employee (the sender) are designed to provide information to the customer (the receiver) in regards to the service exchange that is occurring. In return, Côté contends that customers have the ability to influence the
subsequent emotional displays and regulation of employees; this also fits with theoretical ideas presented by Rafaeli and Sutton (1989). Thus, employee and customer emotions operate in a reciprocally causal, dynamic fashion with emotional displays from each party influencing the subsequent behavior of the other party.

Côté (2005) stated that there are three parts to the feedback loop within the social interaction model. In the first part, the type of emotion regulation the employee is using (i.e., surface acting, deep acting) influences the actual emotional display that is presented to the customer. This sentiment fits with previous work that has demonstrated that emotional displays vary in relation to the type of emotion regulation being used (e.g., Grandey, 2003; Gross, 1998a; Harris, 2001; Totterdell & Holman, 2003). The second part of the feedback loop, according to Côté, builds from affective events theory (Weiss & Cropanzano, 1996) in which the customer interprets the employee’s emotional display as an affective event; in turn, the customer will proceed to appraise and react to the event. Researchers have long documented the idea that employees use emotions in an attempt to control the behavior of customers (e.g., Rafaeli & Sutton, 1987), and this part of the social interaction model acknowledges this process. Finally, in the third part of the model, Côté stated that the customer generates a response to the emotional display of the employee. According to Côté, the customer response is the critical piece in determining the strain felt by the employee. As an example, Côté stated that a customer who responds positively to an employee’s positive emotional display is likely to reduce the employee’s felt strain since the customer reaction will be viewed as supportive.

Côté’s (2005) model outlines circumstances under which customers are likely to negatively react to employee emotions. For example, when a customer perceives that an
employee is not engaging in authentic emotional displays (i.e., is surface acting), he or she is more likely to respond negatively. A laboratory study by Butler and colleagues (2003) demonstrated this effect: receivers who detected that their interaction partner was suppressing or faking emotions felt lower levels of rapport with the sender, in addition to lower levels of liking and increases in blood pressure. Other research by Grandey (2003) also demonstrated that surface acting is negatively related to coworker ratings of affective delivery, further suggesting that individuals not only detect inauthentic emotions from others, but that the detection of such inauthentic displays negatively influences evaluations on the part of the sender.

Conversely, employees who engage in deep acting, which should create more authentic emotional displays (Grandey, 2003; Hochschild, 1983), should invoke more positive reactions from customers. Authentic displays of positive emotions, such as happiness, for example, have been positively linked to more favorable customer reactions (Tsai, 2001; Tsai & Huang, 2002) as well as positive, post-encounter customer mood (Barger & Grandey, 2006). Displays of happiness communicate liking and a desire to affiliate, which leads recipients of such emotional displays to want to respond positively to the emotion sender (Van Kleef, de Dreu, & Manstead, 2010). Thus, Côté (2005) posits that customers will respond favorably to deep acting, resulting in reduced strain on the part of employees. Together, these ideas suggest that service employees will be sensitive to customer reactions that pertain to their own emotional displays when trying to effectively regulate their emotions during an encounter. Applied to the within-episode approach that Beal and Trougakos (2013) highlighted, I expected that the tenets of Côté’s
(2005) model could help explain when and why employees choose to engage in different types of emotion regulation within the same encounter.

Further explanation comes from the control theory principles (e.g., Carver & Scheier, 1999) that Diefendorff and Gosserand (2003) previously integrated into emotional labor research. Control theory explicates that individuals continuously search for any discrepancies between a desired goal/standard and their current state in order to determine whether or not adjustments (i.e., change behavior, change goal) need to be made. Diefendorff and Gosserand (2003) argued that employees in service settings operate in a similar manner. For example, service employees often strive to meet specified emotional display rules by comparing their current displays (and felt emotions; Diefendorff & Richard, 2008) against the display rule and trying to remove the discrepancy by engaging in emotion regulation. Diefendorff and Gosserand further stated, “if the customer suddenly appears irritated, a discrepancy will be detected for the goal of creating a positive emotion in the customer” (p. 949), which may lead the employee to actively do something to manage the situation. If the employee was not engaging in any emotion regulation, then he/she may be predicted to either surface act or deep act, depending on a variety of factors (e.g., the size and nature of the discrepancy, characteristics of the interaction partner, default strategies of the employee). However, if the employee was already engaging in some form of emotion regulation, this new discrepancy may suggest to the person that a different approach is needed, resulting in the person switching to a different regulation strategy. Linking to Côté’s (2005) social interaction model and theory from Rafaeli and Sutton (1989), these ideas fit with the notion that employees use the customers as feedback, helping to dictate how employees
alter their regulation strategies. For instance, Rafaeli and Sutton (1989) stated that customer social feedback can “alter the intensity, or shift the content of emotions expressed by a role occupant [i.e., employee]” (p. 24).

Focusing on a within-episode analysis of emotion regulation, in the context of the current study, I proposed that when an employee began an interaction with a customer, there is likely to be little discrepancy between the emotional experience and the emotional expectations. Part of this expectation stems from the manipulation being used in the current call center study design (see Figure 2 for visual representation).

Figure 2. Three-phase structure of confederate call.

Specifically, when participants (i.e., an “employee”) in the current study begin a phone call with a confederate caller (i.e., a “customer”), confederates were instructed to have a neutral affective tone, thus providing minimal social feedback about the participant’s emotion regulation. Although this is scripted, this service interaction likely operates in a manner similar to most service exchanges: at the beginning, there is little information for customers or employees to use to judge the success (or failure) of the service encounter. Thus, at the initial stages of a customer interaction, reports of both

38
surface acting and deep acting were theorized to be relatively low in comparison to reports of surface and deep acting at other stages of the call.

Midway through the service encounter in the current study (“Phase II”), the customer demeanor shifted to becoming negative (i.e., express anger, irritation, frustration) for all interactions. In this circumstance, employees will likely detect a discrepancy between the desired customer reaction (i.e., satisfaction, pleasantness) and the current state of the customer. According to control theory and the integrated theory outlined by Diefendorff and Gosserand (2003), individuals would begin to either adjust their goal, or change their behaviors. While employees could technically change the goal (i.e., abandon the goal of having happy customers and engage in display deviance; Tschan, Rochat, & Zapf, 2005), this would be problematic for two key reasons. First, customers will expect positive emotional displays since this expectation has become embedded in evaluations of satisfactory service (Parasuraman et al., 1985). Additionally, if an employee engages in display deviance, this would create a discrepancy that would cause problems at a higher level of goal attainment. For example, building from literature on goals (e.g., Carver & Scheier, 1999; Diefendorff & Chandler, 2010), goals are hierarchically arranged, and discrepancies at lower levels of the goal hierarchy create discrepancies and problems for attainment of higher-order goals. Thus, an employee who deviates from engaging in positive emotional displays (i.e., a lower level goal) likely offsets the attainment of higher-order goals, such as satisfactory service performance, happy customers, and so forth. Thus, employees should not engage in display deviance.

Rather, I expected that employees would change their own behavior by increasing their emotion regulation. For example, if an employee reports minimal surface acting or
deep acting at the beginning of the encounter when the customer is neutral, when the customer becomes irritated, the employee may increase one strategy (or both) in an attempt to align subsequent customer emotional displays with service quality standards (i.e., create positive emotions in the customer; Grandey et al., 2013). Thus, in comparison to the beginning of the encounter, in response to a service failure, I proposed that emotion regulation levels would increase.

_Hypothesis 3_: In response to a service failure (i.e., heightened negative affective displays from the customer), mid-interaction levels of reported (a) surface acting and (b) deep acting are higher than those reported at the beginning of the interaction.

After the introduction of a negative demeanor change on the part of customers for all participants (“Phase II”), the customer encounter entered a third phase where the between-subjects manipulation of customer demeanor occurred (“Phase III”). Depending upon the condition participants were randomly placed in, the customer either became more hostile than he was previously (reflecting increased service failure and more negative social feedback to employees) or became positive (reflecting a service recovery and positive social feedback to employees). I theorized that any subsequent changes in emotion regulation in terms of magnitude or vacillation between regulation strategies will depend upon which condition the participant is experiencing.

In the service recovery condition, participants were made to feel as though customers were positively responding to their level of effort and emotion regulation being exerted. In an actual service exchange, when employees increase (or change) their emotion regulation, according to theory from Côté (2005) (see also Diefendorff &
Gosserand, 2003; Rafaeli & Sutton, 1989), the heightened levels of emotion regulation should impact actual employee emotional displays, which will be perceived and evaluated by customers. Essentially, employees are thought to be putting forth more effort to instill positive emotions in the customer in an attempt to improve the service interaction (e.g., Rupp & Spencer, 2006). Thus, the increase in emotion regulation from the employee is aimed at improving customer reactions, leading to a “service recovery.”

In the face of a service recovery (i.e., decreased negative affect and increased positive emotionality from the customer), levels of emotion regulation from the employee may be maintained since the customer deemed the current level of emotion regulation “acceptable” (hence the reduction of negative emotions). Theory from Côté (2005) and Rafaeli and Sutton (1989) fits with this, stating that positive customer feedback indicates that the current level of emotion regulation is effective. Additionally, it may be that employees are fearful that changing their emotion regulation (i.e., lowering it) will result in the customer no longer being satisfied, hence increasing their desire to retain their current regulation strategy. While it is also possible that employees may report lower levels of emotion regulation in response to a service recovery (i.e., feeling that they can relax their effortful regulation), I proposed that it will be kept at the same level as mid-interaction emotion regulation in line with Côté’s theory and the fact that employees may be inclined to continue regulating in a manner that “works.” I also suggested that employees would not experience high levels of fluctuation between strategies given that they will be sticking with the strategy that “works” based upon customer feedback.

In contrast to the service recovery condition, I expected that the escalating service failure condition might lead employees to engage in greater switching between emotion
regulation strategies. In actual service exchanges, there is the possibility that customers will not react positively to the increased levels of emotion regulation after a service failure. Côté (2005), for example, stated that if customers detect inauthentic displays, they are likely to continue negatively reacting to the service experience. In these instances, Côté proposed that employees will detect the continued negative reactions and will make the choice to change emotion regulation strategies to improve the interaction. The nature of this emotion regulation strategy change will depend upon the type of regulation that the employee used immediately after the initial service failure, and will also likely be dependent on the regulation strategy that the employee reports using most frequently during the episode. For example, for employees who engaged in deep acting in response to initial service failure, when customers continue to respond negatively, employees may resign themselves to simply get through the situation by abandoning deep acting in favor of surface acting, which requires less effort. Thus, the increased negative response from the customer could be interpreted as meaning that the previous emotion regulation strategy (i.e., deep acting) was not working, causing the employee to try a different strategy. Additionally, employees may report switching back and forth between the two strategies more so than an employee who experiences positive social feedback from the customer in an attempt to identify what type of regulation will cease the customer’s negative emotional feedback.

Thus, as opposed to making formal predictions about which types of switches between emotion regulation strategies will occur in the increased service failure condition, I proposed a between-subject comparison (i.e., service recovery condition compared to increased service failure condition) in the emotion regulation process.
Specifically, in comparison to employees who experience service recovery, I first expected that overall levels of emotion regulation would be higher in the increased service failure condition in comparison to those in the service recovery condition. Regardless of the changes being made in regulation strategy, negative feedback indicating failure from the customer will exacerbate the need for regulation on behalf of the employee beyond the previous level being exerted. Further, in comparison to the service recovery condition, I expected that the amount of vacillation between strategies would be higher for the group that experiences continued service failure. Employees experiencing failure would be more likely to try different strategies (or combinations of strategies) to alleviate the negative feedback coming from the customer.

Figure 3 provides a visual representation of what the differences between surface acting and deep acting could look like across the three points of the experimental study.
Figure 3. Hypothetical model of emotion regulation variation and discontinuous change during the service episode.

As can be seen, in the first portion of the experiment (‘Phase I’), participants are likely to engage in low levels of emotion regulation and to switch strategies at a relatively low level. At the point when customers become hostile (indicating service failure; ‘Phase II’), a rapid, discontinuous change in emotion regulation is expected, such that levels of surface acting and deep acting rapidly increase. Further, there may be relatively more switching between strategies at this point in the process. Then, at the point of the between-subjects manipulation (service recovery versus increased service failure; ‘Phase III’), the participants in the service recovery condition are expected to maintain their levels of regulation consistently with the previous phase. However, for those in the
increased service failure condition, another discontinuous change is likely to occur, with participants increasing their levels of emotion regulation again. Further, the amount of switching may also increase for the increased service failure participants.

In addition to modeling these divergent levels of overall regulation between conditions, Figure 3 also helps depict how emotion regulation strategies can be negatively correlated within a performance episode. As depicted by the separate lines for deep acting and surface acting, while the overall levels may be positively correlated between participants (or across phases of the episode), the strategies are happening opposite of each other at any given moment, resulting in a negative correlation at the momentary level of analysis. Moreover, there would likely be more vacillation and changing occurring as the service episode becomes harder (i.e., increases in service failure) since employees may be attempting different regulation strategies to manage the situation and social feedback from the customer.

Thus, in conjunction with the theory outlined, I predicted:

*Hypothesis 4:* The levels of (a) surface acting and (b) deep acting are higher during the third part of the service encounter for those participants in the increased service failure condition compared to participants in the service recovery condition.

*Hypothesis 5:* Participants in the increased service failure condition are more likely to switch between surface acting and deep acting strategies in the third phase compared to participants in the service recovery condition.
Outcomes of Within-Episode Emotional Labor Processes

Emotional Exhaustion

Consistent with past emotional labor research, it is expected that within-episode surface acting and deep acting will relate to employee outcomes. The outcome that has received the most attention in emotional labor research is emotional exhaustion. Researchers have found strong support that surface acting is positively related to emotional exhaustion (Brotheridge & Grandey, 2002; Brotheridge & Lee, 2002; Grandey, 2003; Hülsheger & Schewe, 2011; Judge et al., 2009; Pugliesi, 1999; Totterdell & Holman, 2003) and that deep acting is not significantly related to emotional exhaustion (Brotherlidge & Lee, 2002; Brotheridge & Grandey, 2002; Grandey, 2003; Hülsheger & Schewe, 2011). In the context of the current study, I tested these relationships by using a more complete assessment of the amount and type of regulation occurring during the episode. Generally, I expected that the average level of surface acting across the performance episode will be positively related to emotional exhaustion, whereas the average level of deep acting across the performance episode will be non-significantly related to emotional exhaustion, leading to the following prediction:

Hypothesis 6: The level of surface acting averaged across a performance episode is positively related to end of call emotional exhaustion.

Although support for this hypothesis would not necessarily be a new finding, in the current study I proposed exploring the relationships of different within-episode indicators of emotion regulation (e.g., proportion of time using a strategy, switching between strategies) with emotional exhaustion. Such analyses may help to shed light on
the ways in which surface acting leads to emotional exhaustion, and perhaps the conditions under which deep acting relates to emotional exhaustion.

In assessing both surface and deep acting, I examined the mean levels of each type of acting at each separate segment of the service episode (i.e., Phase I, Phase II, and Phase III). Additionally, I examined the proportion of time during the episode that an individual spent using minimal, low, and high levels of surface or deep acting, as well as the proportion of time individuals engaged in minimal, low, and high levels of surface and deep acting simultaneously.

Conversely, when considering both regulation strategies at the same time, a factor that may be important is the number of times an individual switches between surface acting and deep acting. Frequently switching between surface and deep acting during an episode may require more regulatory resources, resulting in higher levels of emotional exhaustion (compared to if a person maintains one strategy throughout). Moreover, it may be that the lowest levels of emotional exhaustion happen when both regulation strategies are low, or perhaps when deep acting is high and surface acting is low. These ideas, and the ideas outlined above, were examined in addition to testing Hypothesis 6.

Observer-Rated Participant Hedonic Tone

In addition to end-of-episode emotional exhaustion, I linked the continuous ratings of emotion regulation to third-party continuous ratings of employee verbal emotional expressions (i.e., hedonic tone). Specifically, two trained research assistants continuously rated the positive and negative emotions that participants expressed (i.e., via vocal quality) during the call with the confederate caller. This assessment helped test two ideas in the literature that have received limited attention. On one hand, there is the
possibility that an evaluation of employee emotional performance (i.e., positive emotional displays) is independent of the type of internal regulation being utilized. As outlined by Hochschild (1983), surface and deep acting are both done with the intent on displaying required emotions. In theory, both should result in successful emotional performance, meaning that while employees may be engaging in a variety of regulation mechanisms to cope with the demands of a service exchange, customers may not be able to detect a difference. This idea is analogous to that of equifinality in the self-regulation literature and the notion that individuals may engage in a variety of internal processes with the intent of achieving the same outcome (Carver & Scheier, 1998).

However, there is the possibility that surface and deep acting may differentially relate to emotional expressions. For example, surface acting, given the incongruence between internal feelings and external displays, might allow for internal negative feelings to “leak” out (e.g., Ekman & Friesen, 1969; Elfenbein & Ambady, 2002). Therefore, in cases of surface acting, it may be easier to detect negative emotions in terms of vocal quality since individuals may be experiencing higher levels of strain and a discrepancy between their internal and external states. Conversely, given that deep acting is aimed at reappraising the situation and feeling the emotions that need to be displayed (Gross, 1998a), it may be that emotional expressions will sound more positive and genuine. Together, these ideas suggest that surface acting and deep acting are not interchangeable regulation strategies, with one strategy (deep acting) being more effective for producing positive emotional expressions than the other (surface acting).

Though research on these linkages is limited, some emotional labor researchers have attempted to link surface acting and deep acting to emotional performance. For
example, Bono and Vey (2007) utilized a laboratory simulation where participants were either in an “enthusiasm” condition (i.e., playing the role of a university tour guide and recruiting students) or an “anger” condition (i.e., a debt collector trying to collect overdue rent from students). After learning about their role and emotional requirements, participants engaged in a role-play in front of a video camera. Post-performance, participants completed Grandey’s (2003) measure of surface and deep acting. After the lab portion was completed, third-party observers watched the videos and rated the extent to which participants were genuine and effective with their emotional displays, as well as the extent to which they engaged in certain emotional expressions (i.e., enthusiasm, anger). Bono and Vey then created an overall index of emotional performance. Results indicated that there was a significant, positive correlation between deep acting and emotional performance ($r = .30, p < .01$) but there was no relationship between surface acting and emotional performance ($r = -.05, n.s.$), providing initial support for the idea that surface and deep acting differentially relate to evaluations of emotionality.

Additional research linking emotion regulation to emotional performance comes from Hülsheger and Schewe’s (2011) meta-analysis. Hülsheger and Schewe found a positive population correlation between deep acting and emotional performance ($\rho = .18$, 95% confidence interval = .047 to .303) as well as a negative relationship between surface acting and emotional performance ($\rho = -.14$, 95% confidence interval = -.264 to -.017). These findings help shed further light on the relationship between emotion regulation and ratings of emotional performance, and even suggest that surface acting may be related to negative hedonic tone.
Although these findings are useful, they do not focus specifically on the type (i.e., positive, negative) of emotional expression being rated, but rather an overall rating of emotional performance. By focusing just on the emotionality of participants’ displays, the current study can shed light on how different types of emotion regulation within a performance episode link to the expression of positive and negative emotions. Building on the findings that deep acting is generally positively related to emotional performance (where positive displays are part of the definition of good performance), and Hülsheger and Schewe’s (2011) findings that surface acting is negatively related to emotional performance, I predicted:

Hypothesis 7: At the within-episode level of analysis, (a) deep acting is positively related to the hedonic tone of employee emotional expressions and (b) surface acting is negatively related to the hedonic tone of employee emotional expressions.

Given the unique data that will be collected, I explored whether switching between surface acting and deep acting results in a change in third party ratings of hedonic tone. For example, it might be that the more individuals fluctuate between emotion regulation techniques within episode (regardless of whether it is surface acting or deep acting), the more likely it is that negative emotions will leak out to customers as participants shift between regulation strategies. These switches in emotion regulation may represent a period where the employee is likely not regulating as he/she tries to reassess the situation to determine how best to deliver positive emotional displays. As a result, negative emotions may be more likely to “leak” during or around the switching timeframe.
In sum, the seven proposed hypotheses have been designed to test three significant questions in the emotional labor literature that have received theoretical attention, but little empirical testing and/or support: 1) are surface acting and deep acting positively or negatively related, 2) can customers influence employee emotion regulation within a performance episode, and 3) is surface acting bad for well-being and emotional performance, whereas deep acting is beneficial? A summary of the proposed hypotheses can be found in Table 1.

Table 1. Summary of hypotheses.

<table>
<thead>
<tr>
<th>Hypothesis 1</th>
<th>At the momentary level of analysis, there is a negative correlation between surface acting and deep acting.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis 2</td>
<td>Episode-level mean surface acting and mean deep acting are positively correlated across subjects.</td>
</tr>
<tr>
<td>Hypothesis 3</td>
<td>In response to a service failure (i.e., heightened negative affective displays from the customer), mid-interaction levels of reported (a) surface acting and (b) deep acting are higher than those reported at the beginning of the interaction.</td>
</tr>
<tr>
<td>Hypothesis 4</td>
<td>The levels of (a) surface acting and (b) deep acting are higher during the third part of the service encounter for those participants in the increased service failure condition compared to participants in the service recovery condition.</td>
</tr>
<tr>
<td>Hypothesis 5</td>
<td>Participants in the increased service failure condition are more likely to switch between surface acting and deep acting strategies in the third phase compared to participants in the service recovery condition.</td>
</tr>
<tr>
<td>Hypothesis 6</td>
<td>Levels of surface acting averaged across a performance episode are positively related to emotional exhaustion.</td>
</tr>
<tr>
<td>Hypothesis 7</td>
<td>At the within-episode level of analysis, (a) deep acting is positively related to the hedonic tone of employee emotional expressions and (b) surface acting is negatively related to the hedonic tone of employee emotional expressions.</td>
</tr>
</tbody>
</table>
CHAPTER III

METHOD

Participants

Eighty-four undergraduate students from a large Midwestern university were randomly recruited from Introduction to Psychology courses to participate in the current study. Participants were told that they would be participating in a call center simulation and would be interacting with student callers from another university. In exchange for participation, participants were awarded extra credit points and the chance to win one of two $50.00 gift cards to Amazon.com. The original anticipated sample size was 60 (30 in the service failure condition; 30 in the service recovery condition) based upon previous experimental research utilizing continuous ratings (e.g., Mauss et al., 2005, 2011).

Of the 84 students, 77 (91.67%) passed the manipulation checks by accurately identifying the different emotional components of the confederate call (i.e., whether the [confederate] caller was pleasant or rude at the beginning, middle, or end of the call). Of the eight students who did not pass the manipulation check, seven indicated that there was no emotional variability from the confederate caller during the call (i.e., the caller was rude at the beginning, middle, and end) and one indicated that the caller started the call rude/not polite (i.e., the caller should have been rated as polite/not rude). One additional participant was dropped for not being able to recall the meaning of surface
acting or deep acting (i.e., participant asked ‘What is surface acting and deep acting?’ after having already gone through the ratings and instructions). This resulted in a final sample of 76 participants (90.47% retained), with 38 participants in the service failure condition and 38 participants in the service recovery condition. Women comprised 65.8% of the sample. The majority of the sample was Caucasian (69.7%) and were first-year students at the university (57.9%) with an average age of 20.94 ($SD = 4.99$). Out of the 76 participants, 60.5% were employed at the time of their participation, with 59.2% of those employed working a part-time job. On average, those employed worked 17.90 hours a week ($SD = 11.35$). Across the entire sample, 69.7% indicated they had, at one point, worked a service-based occupation (i.e., retail, health care, banking, food services), and only 10.5% indicated they had ever worked in a call center setting.

As a note, five additional people were run through the call center but did not follow the call center role correctly (i.e., did not answer the caller’s questions correctly, stated that the caller’s problem could not be solved). This prevented the five participants from fully receiving the manipulations in the study. Hence, they were removed from the original data set (i.e., not included in the 85 noted above).

Procedure

The call center simulation used in the current study was adapted from McCance et al. (2013). The benefit of using a call center simulation, besides the prevalence of call centers in today’s service industry (e.g., Winiecki & Wigman, 2007) and following past precedent with other research (e.g., Goldberg & Grandey, 2007; Rupp & Spencer, 2006), is the ability to systematically and consistently impact the employee experience through the scripted actions of a confederate customer. Further, according to Goldberg and
Grandey (2007), having a call center simulation for studies of emotional labor removes confounds from the research, such as the appearance of the confederate customers, as well as adding increased realism (i.e., it is more realistic to simulate a call center than an over-the-counter service interaction).

The current study had three stages of data collection. In Stage I, participants completed an online survey through SurveyMonkey.com to assess basic demographic information (e.g., age, gender, ethnicity, college major), previous service work experience, and previous call center work experience. Participants also completed a variety of individual difference variables that were part of a larger data collection (i.e., to help the generation of future research projects).

In Stage II, participants came to the laboratory to partake in a 90-minute call center simulation (see Figure 4 for a visual representation of the lab portion of the call center simulation).

Figure 4. Visual representation of the laboratory call center simulation.
Upon arriving at the lab, participants were provided an informed consent, and then were introduced to their trainer (myself, or another graduate student who I personally trained on all procedures; see Appendix A for Training Script for Surface Acting first, and Appendix B for Training Script for Deep Acting first). Participants were told that the current study was being conducted in partnership with a local call center (University Tech Support) located in the same Midwestern town as the school. Participants were then informed that they would be playing the role of a service representative, and would be answering calls over the course of an hour via Skype, which is an internet based phone service that University Tech Support uses in their call centers.

After learning about University Tech Support, participants were told that the current study was operating in tandem with another study occurring at another university in the Midwest. Participants in that study (i.e., the confederate callers) were completing PowerPoint presentations in a timed competition to be entered to win $100 if they were the first in their session to finish. Thus, students role playing in the current study as call center representatives were told they would be fielding calls from these students (i.e., confederate callers) and assisting them through step-by-step instructions in a PowerPoint manual that was created by University Tech Support for the current study (see Appendix C for Participant Information, and Appendix D for PowerPoint training manual). Participants were assured that the materials were pilot tested, and that it was rare that questions were asked outside of the list of steps in the manual (hence supporting why the document has so few pages).

Once the scenario was described, the requirement for positive emotional displays in the call center was enforced. Participants were told that because they were working for
University Tech Support, it was imperative to follow their customer service guidelines. Participants were told that they had to be happy, positive, and respectful for all customers who were calling into the hotline, especially since students in the other study (i.e., the confederates) thought they were calling a real technology hotline. Adherence to the positive emotional display rules was encouraged three different ways. First, participants were told that their calls would be monitored and recorded via Skype for service quality, a standard practice in many call centers. Second, they were told that customers calling into the call center would also be rating them on the customer service they provided. Thus, we would be able to make sure they were expressing happy and positive emotions no matter what happened during the call. As a final reinforcement, participants were told that they would only be entered into the lottery to win one of the two $50.00 gift cards to Amazon.com if they had good customer service ratings; in the end, all participants were entered into the raffle.

Following the one-on-one training, participants were left alone for five minutes to review the training materials (Appendices C and D). Inside the training materials were reminders of the typical behaviors of call center employees (i.e., “We want our customers to know you are smiling at them even though they are a phone line away!”), instructions for how to take a call, (i.e., specific greeting, request for caller ID number for “matching” in the other study, follow-up questions, and a closing statement), as well as step-by-step instructions for different types of formatting in Microsoft PowerPoint. Participants read again that prior to their participation, pilot studies had been conducted about the typical questions that emerge from customers in order to generate the scenarios that have been provided, and that it is rare for customers to call with problems outside of that list.
After the five minutes, the trainer returned to the room and asked if there were any questions. If there were questions (e.g., how many people would call, how to answer the phone), they were addressed. If there were no questions, participants were then told that they would be doing a practice call with the trainer to make sure they understood how to handle customer calls. The trainer then went up the hall to a second lab (i.e., the confederate lab) and placed a practice call via Skype. The call lasted approximately 82 seconds (range = 45 – 125 seconds; script for practice call can be found in Appendix A or B [it was the same for both conditions]). During the call, the trainer asked for assistance with inserting a table into a PowerPoint presentation. This allowed participants to a) go through the mandatory greeting, follow-up, and closing statements required (see Appendix D), b) read through a set of step-by-step PowerPoint instructions, and c) experience a fairly neutral-to-pleasant ‘customer’ encounter.

As soon as the practice call was completed, the trainer returned to the call center lab and corrected any mistakes that were made (i.e., not providing the full greeting statement, not providing the full closing statement) and further reinforced the need to be pleasant during the call. Participants were then told that there would be a second portion of the study that involved providing continuous ratings of their experience during the calls. Participants then practiced providing the three continuous ratings (e.g., emotionality, surface acting, deep acting; see Simulation Measures section for more detail) on the practice call.

Following the completion of the practice rating, participants began the actual call center simulation. Prior to receiving calls, participants were wired to three sets of physiological sensors (see Appendix I). At this time, participants were informed that the
trainer would be staying in the room to monitor the physiological equipment. They were also told that the trainer was prohibited from helping them during the call center simulation based upon guidelines from University Tech Support. In reality, the trainer remained in the room to place a marker in the physiological data software indicating when the call began. Additionally, participants were told that they would be fielding an unknown number of calls (i.e., it depended on how many students were participating in the other study) and that they would be taking calls until they were told by the trainer to stop. The trainer then called the ‘researcher’ running the other study to let them know that the call center was ready. The researcher being called was actually the confederate caller. The confederate callers placed their phone call (i.e., service failure or service recovery call) approximately 30 seconds after the call center went live.

Participants received one of two confederate phone calls: service failure (see Appendix E), or service recovery (see Appendix F). Following precedent from Goldberg and Grandey (2007), all confederate callers were male. Both calls were the same in terms of the emotionality expressed and vocal content by the confederate in the first two phases of the call, in addition to the three questions asked: 1) how to insert a text box, 2) how to put a border around the text box, and 3) how to change the width of the border around the text box. The only difference was during Phase III of the call (i.e., the between-subjects manipulation), where the confederate either became increasingly more negative and hostile (service failure) or became calmer and more pleasant with the participant (service recovery). The confederate scripts were crafted to ensure that participants felt the need to truly regulate their emotions by having the content of the confederate script challenge the participation from a relational standpoint. One could argue that emotion regulation would
be minimal if the call was strictly about technical aspects of creating a PowerPoint. By adding relational components (i.e., “You’re doing a really terrible job!”), it was more likely that participants would truly have to put forth effort and utilize a variety of regulation techniques to handle the caller.

After the confederate call was complete, the trainer stopped participants after waiting approximately 30 seconds. Participants were told that the researcher from the other study had no more students participating, and that the one phone call was the only one they would have to take (see Post-Simulation script in Appendix G). At this point, the physiological measures were removed from participants, and they completed the final set of continuous ratings. After the continuous ratings, they completed a final survey battery (see Post-Simulation Measures), and were debriefed and dismissed.

Pre-Simulation Measures

Several individual difference variables were assessed as part of a larger data collection (i.e., demographics, personality, dispositional affectivity). All pre-simulation measures were collected online via Surveymonkey.com prior to participants coming in for the lab portion of the study. Given the lack of emphasis on these in the current study, all measure descriptions, items, and reliabilities can be found in Appendix H.

Physiological Measures

Though not a primary part of the current study, as part of an exploratory analysis, continuous physiological measures were recorded during the actual performance episode (i.e., as the phone call with the confederate caller was occurring). Given their lack of emphasis in the current study, a more thorough description of the measures can be found in Appendix I.
Simulation Measures

All simulation measures (i.e., continuous ratings) were collected via a program in E-Prime 2.0 that was designed for the current study. This program allowed continuous rating assessments to be collected every 200ms as participants moved a mouse along a continuous rating scale on a computer screen, and were synchronously recorded with the playing of an audio file of the confederate phone call. In total, three continuous ratings were collected during the practice ratings (i.e., after the trainer call) and after the actual confederate call; only ratings obtained from the confederate phone call are used in the current study. Ratings were counterbalanced, such that half of the participants provided continuous ratings of surface acting first, and half provided continuous ratings of deep acting first; both groups, started with rating their felt emotions during the call. Figure 5 is a visual representation of the counterbalancing procedure used for the current study; this counterbalancing was maintained for the practice and actual ratings.

<table>
<thead>
<tr>
<th>Condition Assignment</th>
<th>Continuous Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group I:</strong></td>
<td></td>
</tr>
<tr>
<td>Service Failure</td>
<td></td>
</tr>
<tr>
<td>(n = 38)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Emotionality</td>
</tr>
<tr>
<td></td>
<td>Emotionality</td>
</tr>
<tr>
<td><strong>Group II:</strong></td>
<td></td>
</tr>
<tr>
<td>Service Recovery</td>
<td></td>
</tr>
<tr>
<td>(n = 38)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Emotionality</td>
</tr>
<tr>
<td></td>
<td>Emotionality</td>
</tr>
</tbody>
</table>

*Note.* Each condition had 38 participants. This means that, for each condition, 19 rated surface acting first after participant emotionality, and 19 rated deep acting first after participant emotionality.

Figure 5. Visual representation of counterbalancing procedure.
Prior to providing each of the continuous ratings, participants were verbally read the instructions for each rating scale (see Appendices A and B). These instructions contained the definition of the construct, the type of scale the construct was being assessed on, and a statement stating that participants could move the mouse as much or as little as they liked. After receiving the verbal instructions, participants were prompted with instructions on the screen that reinforced the verbal instructions. Participants then proceeded to the rating screen and clicked ‘A’ to begin the audio recording of their phone call. All continuous assessments were recorded immediately at the onset of the audio.

As a note, each of the continuous rating scales had twenty rating point ‘boxes’ paired with rating anchors that were connected together (i.e., no space between them). Having the anchors connected together guaranteed that at any given time, participants had the mouse hovering over a rating. The participant’s mouse also was restricted to only appear in the rating box area of the screen.

Participant Emotionality

Though there were no hypothesized effects for participant emotionality in the current study, the first continuous rating assessment measured the within-episode variability of felt emotions during the phone call. Figure 6 is a screen shot of the instructions for participant emotionality, and Figure 7 is the screen shot of the participant emotionality screen.
How did you feel during the call?

For this continuous rating, please rate how you felt during the call. This will range from very negative on the left hand side of the rating scale to very positive on the right hand side. If you felt neutral, keep the mouse in the middle.

Please move the mouse as much or as little as you like as you listen to the call. The movement of the mouse should reflect how you were actually feeling during the call.

Please press the space bar to continue.

Figure 6. Participant emotionality instructions.

How did you feel during the call?

Please hover over one of the rating numbers at all times.

<table>
<thead>
<tr>
<th>Very Negative</th>
<th>Negative</th>
<th>Neutral</th>
<th>Positive</th>
<th>Very Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
</tbody>
</table>

Figure 7. Participant emotionality rating screen.

Each participant emotionality rating began with the mouse hovering over ‘10’ in order to start participants at the neutral emotionality part of the scale. Participants were told to move the mouse to the left to reflect feeling negative emotions during the call, and...
to the right to reflect feeling positive emotions during the call. Participants were encouraged to move the mouse as much or as little as they liked depending on their own personal emotional experience.

Surface Acting

Surface acting assessed the extent to which participants felt like they had to hide their actual emotions from the caller. Figure 8 is of the surface acting instruction screen, and Figure 9 is of the surface acting rating screen.

**How much did you hide your emotions?**

Please continuously rate the extent to which you attempted to hide your true emotions during the call. This scale ranges from not at all on the left hand side, a moderate amount in the middle, and to a very large extent on the right hand side.

If you did not try to hide your felt emotions, you will move the mouse to the far left side of the rating scale. If you tried to hide your emotions to a very large extent, you will move the mouse to the far right side of the scale.

Please press the space bar to continue.

Figure 8. Surface acting instructions.
Each participant rating began with the mouse hovering over the ‘1’ in order to start participants at the ‘None’ indicator of surface acting. Participants were told to move the mouse over the number that represented the extent to which they felt they were ‘hiding their felt emotions from the caller’ (i.e., surface acting) during the phone call while listening back to the recording. In defining surface acting to participants, they will be told that surface acting means that they experienced hiding emotions from the customer. The ‘hiding’ aspect of surface acting is a key dimension in common assessments of surface acting (e.g., Diefendorff et al., 2005). Participants were told that they were to provide ratings on the 1-20 scale, with a rating of “1” indicating not surface acting at all, and a rating of “20” indicating surface acting as much as possible. Participants were told to continuously rate the extent to which they used surface acting.
during the call, providing ratings as infrequently or as often as they felt it accurately reflected their emotional experience.

*Deep Acting*

Deep acting assessed the extent to which participants felt like they put forth effort to change their emotions in order to be more positive with the caller. Figure 10 is of the deep acting instructions screen, and Figure 11 is of the deep acting rating scale.

**How much did you change your emotions?**

Please continuously rate the extent to which you tried to change how you felt during the call (i.e., to be more positive). This scale ranges from not at all on the left hand side, a moderate amount in the middle, and to a very large extent on the right hand side.

If you did not try to change your felt emotions, you will move the mouse to the far left side of the rating scale. If you tried to change your felt emotions to a very large extent, you will move the mouse to the far right side of the scale.

Please press the space bar to continue.

Figure 10. Deep acting instructions.
How much did you change your emotions?

Figure 11. Deep acting rating scale.

Each participant rating began with the mouse hovering over the ‘1’ in order to start participants at the ‘None’ indicator of deep acting. Participants were told to move the mouse over the number that represented the extent to which they felt they were ‘trying to change their emotions’ (i.e., deep acting) during the phone call while listening back to the recording. In defining deep acting to participants, they were told that deep acting means that one is trying to actively change his or her emotions to feel more positively, in order to be pleasant with the caller. This process of actively changing one’s emotions is a key dimension in common assessments of deep acting (e.g., Diefendorff et al., 2005). Participants were told that they were to provide ratings on the 1-20 scale, with a rating of “1” indicating not deep acting at all, and a rating of “20” indicating deep acting as much as possible. Participants were told to continuously rate the extent to which
they used deep acting during the call, providing ratings as infrequently or as often as they felt it accurately reflected their emotional experience.

Post-Simulation Measures

All post-simulation measures were collected from participants after the continuous ratings. Emotional exhaustion was the primary dependent variable for the current study. Additional variables were, however, collected as part of a larger data collection. Only emotional exhaustion is presented below; a description of other post-simulation measures can be found in Appendix J.

Emotional Exhaustion

Emotion exhaustion was assessed using six of the seven items adapted from Erickson and Ritter’s (2001) emotional exhaustion scale. “I dread getting up in the morning and having to face another day on the job” was removed from the scale since it is not applicable to the one-time call center experience of the participants. All items were rated on a 7-point scale (1 = “strongly disagree;” 7 = “strongly agree.” An example item is “I feel used up right now” (see Appendix J for all items). The six items were averaged together to form a composite (Cronbach’s alpha = .87).

Third-Party Continuous Ratings

Call Segmenting

In order to distinguish the three phases of the phone call (i.e., Phase I – neutral/positive, Phase II – negative, Phase III – more negative or neutral/positive), a trained undergraduate research assistant utilized a continuous rating screen to divide the call into the three time segments. This rating program was in E-Prime 2.0 computer software, and collected ratings every 200ms to correspond to the continuous ratings.
provided by study participants. When making the ratings that divided the call into segments, the research assistant listened to each recorded phone call blind to the study condition and marked the two time points during the performance episode where the emotionality of the confederate’s vocal tone changed to being negative, ‘more’ negative, or positive/pleasant. The research assistant was only able to hear the confederate’s voice, helping to ensure that the change in confederate emotional tone was the primary driving factor in distinguishing the different time segments. These markings were then used when analyses were divided into the different phases of the phone call. Instructions for the call segmenting instructions are in Figure 12; Figure 13 is the rating screen.

**Call Segmentation**

You are about to listen to a participant’s phone conversation with one of our confederate callers. The goal of this is to **identify different emotional portions of the phone call based upon the emotional vocal content of the confederate.**

Each call in the study had three segments that vary between being the confederate caller being friendly, difficult, and very difficult. You will use a 1-20 rating scale that has been divided into three sections. Please hover below the appropriate section **depending on the confederate’s emotional vocal content.**

Figure 12. Call segmenting instructions.
Call Segmentation

Figure 13. Call segmenting rating screen.

In an attempt to provide evidence that the call segmenting ratings were producing the needed information, a second rater listened to a subset of the calls ($n = 6$; 3 in each condition, one call per confederate caller) and provided call segmenting ratings. The difference between the primary and second raters was minimal, with approximately 5 seconds difference in switching from Phase I to Phase II of the call, and 4 seconds difference in switching from Phase II to Phase III of the call.

Third-Party Affect Ratings

Two trained undergraduate research assistants separately provided continuous ratings of the affective tone (i.e., positive, neutral, negative) of the participants. These ratings were done in order to assess the hedonic vocal tone of participants. The program
used was in E-Prime 2.0, and collected ratings every 200ms to align with the continuous ratings provided by the participants. Each research assistant provided his or her ratings separately, were blind to the study condition, and only heard the participant’s voice to remove the possibility of the confederate caller biasing the ratings. The third-party affect instructions screen is presented in Figure 14; Figure 15 is the rating screen.

**Participant Emotionality**

You are about to listen to a participant’s phone conversation with one of our confederate callers. Because you are focusing on the participant, the confederate’s portion of the conversation has been muted. This means you will only be listening to the participant’s voice.

Please rate the extent to which you believe the participant was exhibiting positive or negative emotionality. Since this is a phone call (i.e., you cannot see what the emotional displays were), try to capture positive and negative emotions in either the participant’s vocal quality.

Press space bar to continue.

Figure 14. Third-party affect rating instructions.
Both research assistants were trained on how to detect differences in vocal emotional tone by reading articles on the topic and participating in training sessions with the lead experimenter. As part of the training, the two research assistants made practice ratings together on calls not used in the current study; as they were providing ratings, they discussed what they were hearing with each other and the experimenter and why they would rate vocal quality of the call in a particular way.

Given that the ratings were continuous for each research assistant, it was difficult to utilize traditional tests of inter-rater reliability given the continuous nature of the data. As such, I looked at the cross-correlation between the two ratings to determine the extent to which the ratings were covarying with each other. Each research assistant’s rating was aggregated to the 1-second time interval (from 200ms) in order to obtain second-by-second.
second averages that were used in my analyses for hypothesis testing (see Analytic Approach for more detail). The cross-correlation between the two raters’ third-party affect ratings was .62, which was significant at \( p < .001 \), suggesting high levels of concordance between the two ratings. Thus, both ratings were averaged together to form a composite third party rating.

**Analytic Approach**

Procedures followed were similar to those utilized by Gottman and Levenson (1985) and Mauss et al. (2005, 2011). Continuous rating assessments were collected every 200ms per participant. Following past precedent (Gottman & Levenson, 1985; Mauss et al., 2011), all 200ms data was aggregated to 1 second sum averages to create second-by-second continuous rating assessments. This resulted in a mean of 264.75 data points per participant (\( SD = 48.09 \), range = 173.00 – 404.00), or approximately 4 minutes and 24 seconds, on average, for the entire call. Calls in the service failure condition lasted 255.37 seconds on average (\( SD = 51.28 \)), and calls in the service recovery condition lasted 274.13 seconds on average (\( SD = 43.33 \)); the difference between conditions was not statistically significant (\( t_{(74)} = 1.72, n.s. \)).

To assess the extent to which continuous rating assessments covaried (e.g., tests of Hypothesis 1 and 7), cross-correlational analyses were used. A cross-correlation is a measure of the extent to which two constructs (i.e., surface acting, deep acting) covary with each other across time while accounting for lag/lead effects, and is an appropriate method of assessing the relationship between two variables which are measured with time series data (i.e., continuous ratings) that are non-stationary. Stated differently, cross-correlations assess the extent to which “two measures covary across time within a given
individual, while taking into account lags between measures” (Mauss et al., 2005, p. 181). This differs from an autocorrelation, which is a cross-correlation of a lagged variable with itself as opposed to another construct. Mauss et al. (2005) stated that cross-correlations should be calculated within a time window of -10 seconds to +10 seconds because “theoretical considerations lead us to expect meaningful time lags among measures…are not greater than 10 seconds in either direction” (p. 181; see also Gratton, 2000; Levenson, 1988); a similar metric of -10 seconds to +10 seconds was utilized in the current study (specified within the SPSS statistical software), providing a total of 21 cross-correlations per participant. Cross-correlations can be calculated utilizing SPSS statistical software by specifying the length of the lag/lead effect desired (i.e., 10 seconds).

To account for the within-person nesting of the data, Mauss et al. (2005) indicate the importance of calculating each cross-correlation separately for each participant. Then, to determine the average cross-correlation between two continuous ratings, I followed the guidelines by Mauss et al. (2005). Because the sampling distribution of correlations is not normally shaped (i.e., it is negatively skewed; Bobko, 2001), one must first transform the cross-correlations for each participant into Fisher’s Z values. The transformed Fisher’s Z values were then averaged across all 21 within-person correlations to arrive at one Fisher’s Z per person. The average Fisher’s Z was then calculated across participants, with this final value converted back into r for interpretation. Positive cross-correlations (e.g., .76) indicate the two continuous ratings follow a similar pattern within individuals; negative cross-correlations (e.g., -.76) indicate the ratings follow the opposite pattern within individuals. An alternative approach in the literature is to take the highest single
cross-correlation per person (instead of the average of the 21; Mauss et al., 2011), but doing so can upwardly bias the results by picking the strongest correlation and not representing the full range of correlations. As such, I followed the more conservative approach of averaging all 21 cross-correlations.

Statistical significance of the cross-correlations was calculated by testing the values against 0, or in the case in which correlations are compared between experimental groups, by comparing the cross-correlations between conditions using the test of independent correlations by Preacher (2002), which is based on a formula by Cohen and Cohen (1983; Cohen, Cohen, Aiken, & West, 2003). The calculator converts $r$s to Fisher’s $Z$ and compares the two correlations based upon the sample size specified (in the case of comparing conditions, $n = 38$). Specifically, the formula used by the calculator is:

$$z = \frac{z'_a - z'_b}{\sqrt{1/(n_a - 3) + 1/(n_b - 3)}}$$

In this formula, $z'_a$ equals the Fisher’s $z$ transformed correlation between $X$ (e.g., surface acting) and $Y$ (e.g., deep acting) for one group (e.g., service recovery group) and $z'_b$ equals the Fisher’s $z$ transformed correlation between $X$ (e.g., surface acting) and $Y$ (e.g., deep acting) for the other group (e.g., service failure condition). Tests can be interpreted with one- or two-tailed significance tests; given that no direction of the difference was specified in my hypotheses, two-tailed significance tests were always utilized.

For analyses in which standard correlation (e.g., Pearson product moment correlations), regression, and analysis of variance techniques were used, aggregated results from the continuous rating assessments were used. Mean levels of surface acting,
deep acting, participant emotionality, and third-party affect ratings were aggregated for each person to obtain one overall score per person.

Data Screening

Removal of Continuous Rating Outliers

Although there is precedent to not alter the continuous ratings in a significant manner (i.e., one cannot determine whether or not an extreme high/low score is error caused by the participant, or a meaningful rating; Laurans, Desmet, & Hekkert, 2009), preliminary data screening was conducted for each of the 76 participants individually to determine if any of the continuous rating data points for participant emotionality, surface acting, and deep acting were found to be extreme scores. To do so, continuous rating assessments were transformed into standardized scores within persons; scores that were greater than 3 SDs above the mean were inspected to see whether they fit the pattern of results before and after the extreme rating or not (i.e., a rating of ‘1’ when the three scores before and after were ‘15’). This resulted in data points being changed for three participants; three data points were changed with the average of the point above and below for deep acting, and one data point was changed with the average of the point above and below for participant emotionality. Given that there were a total of 20,121 data points for the entire sample, this resulted in only .0001% of the data being altered for deep acting, and .00005% of the data being altered for emotionality.

Experimenter Effects

Due to the fact that two experimenters trained participants (Experimenter 1 = 56 participants, 73.7%; Experimenter 2 = 20 participants, 26.3%), a series of independent samples t-tests were conducted to determine if there was a difference between the means
and standard deviations of participant-rated emotionality, surface acting, and deep acting dependent on the trainer. These analyses were conducted two ways: once across all participants, and once with participants split between study condition (i.e., service failure, service recovery). Results are presented in Table 2.

Table 2. Independent samples t-tests for experimenter effects.

<table>
<thead>
<tr>
<th></th>
<th>All Participants</th>
<th>Service Failure</th>
<th>Service Recovery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>t</em></td>
<td><em>t</em></td>
<td><em>t</em></td>
</tr>
<tr>
<td>Participant Emotionality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-1.46</td>
<td>-1.33</td>
<td>-0.83</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>-1.34</td>
<td>1.38</td>
<td>0.39</td>
</tr>
<tr>
<td>Surface Acting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.46</td>
<td>0.50</td>
<td>1.74</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.40</td>
<td>0.15</td>
<td>0.47</td>
</tr>
<tr>
<td>Deep Acting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.02</td>
<td>0.44</td>
<td>1.87</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.59</td>
<td>0.44</td>
<td>0.33</td>
</tr>
</tbody>
</table>

*Note.* Degrees of freedom for analyses across all participants are 74; degrees of freedom for analyses within each of the study conditions are 36. All values are t-statistics for experimenter differences in the means and standard deviations of participant emotionality, surface acting, and deep acting.  
* **p < .05; *** p < .001.

Across all 76 participants, there was no difference between experimenters on average participant rated emotionality (*t* = -1.46, *n.s.*), average participant rated surface acting (*t* = 1.46, *n.s.*), or average participant rated deep acting (*t* = 1.02, *n.s.*). For the service failure condition, there was no difference between experimenters for participant emotionality (*t* = -1.33, *n.s.*), surface acting (*t* = 0.50, *n.s.*), or deep acting (*t* = 0.44, *n.s.*). Finally, for the service recovery condition (*n* = 38), there was no difference between experimenters on average participant rated emotionality (*t* = -0.83, *n.s.*), average participant rated surface acting (*t* = 1.74, *n.s.*), or average participant rated deep acting (*t* = 1.87, *n.s.*).
Because experimenters could have differentially emphasized the level of variability participants had with the continuous ratings (i.e., over-emphasized that a person can move the mouse as much or as little as he/she wanted), the same series of analyses used on the means for participant emotionality, surface acting, and deep acting were used for the standard deviations (see Table 2). Standard deviations help capture the amount of variability exhibited in the continuous scales, and have been used by scholars to capture the idea of variability in event-level data collections (e.g., Scott et al., 2012). No differences were found in the standard deviations of participant emotionality ($t_{(74)} = -1.34, n.s.$), average participant rated surface acting ($t_{(74)} = 0.40, n.s.$), or average participant rated deep acting ($t_{(74)} = 0.59, n.s.$). Similarly, there were no effects within the service failure (participant emotionality: $t_{(36)} = 1.38, n.s.$; surface acting: $t_{(36)} = 0.15, n.s.$; deep acting: $t_{(36)} = 0.44, n.s.$) or service recovery conditions (participant emotionality: $t_{(36)} = 0.39, n.s.$; surface acting: $t_{(36)} = 0.47, n.s.$; deep acting: $t_{(36)} = 0.33, n.s.$).

**Confederate Effects**

Calls were made by six confederate callers that were randomly assigned to the service failure or service recovery conditions (Confederate 1 = 12 calls, 15.8%; Caller 2 = 20 calls, 26.3%; Caller 3 = 9 calls, 11.8%; Caller 4 = 9 calls, 11.8%; Caller 5 = 15 calls, 19.7%; Caller 6 = 10 calls, 13.2%); a seventh confederate was used for just one call in the service recovery condition (1.3%) due to a last minute cancellation of a confederate caller. Similar to the analyses run for between-trainer differences, four sets of one-way ANOVAs (analysis of variance) were conducted: between-confederate differences on overall means, on means within each condition, on the standard deviations across all
participants, and the standard deviations within each condition. All analyses are presented in Table 3.

Table 3. One-way ANOVAs for confederate caller effects.

<table>
<thead>
<tr>
<th></th>
<th>All Participants</th>
<th>Service Failure</th>
<th>Service Recovery</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participant Emotionality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.07</td>
<td>1.07</td>
<td>1.21</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.11</td>
<td>2.43</td>
<td>0.35</td>
</tr>
<tr>
<td><strong>Surface Acting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.48</td>
<td>0.19</td>
<td>0.77</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.76</td>
<td>3.15</td>
<td>0.06</td>
</tr>
<tr>
<td><strong>Deep Acting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.56</td>
<td>0.18</td>
<td>0.48</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.79</td>
<td>1.81</td>
<td>0.12</td>
</tr>
</tbody>
</table>

*Note. Degrees of freedom for analyses across all participants are 6, 69; degrees of freedom for the service failure condition are 5, 32; degrees of freedom for the service recovery condition are 6, 31. The degrees of freedom are different between the two conditions because the service recovery condition had an extra confederate caller for one call (i.e., a seventh confederate). All values are F-statistics for testing confederate effects on the means and standard deviations of participant emotionality, surface acting, and deep acting. * p < .05; ** p < .01; *** p < .001.*

Across all confederate callers, no differences were found for participant emotionality ($F_{(6, 69)} = 1.07, n.s.$), surface acting ($F_{(6, 69)} = 0.48, n.s.$), or deep acting ($F_{(6, 69)} = 0.56, n.s.$). This was replicated with no differences between confederate callers in the service failure (participant emotionality: $F_{(5, 32)} = 1.07, n.s.$; surface acting: $F_{(5, 32)} = 0.19, n.s.$; deep acting: $F_{(5, 32)} = 0.18, n.s.$) or service recovery conditions (participant emotionality: $F_{(6, 31)} = 1.21, n.s.$; surface acting: $F_{(6, 31)} = 0.77 n.s.$; deep acting: $F_{(6, 31)} = 0.48, n.s.$).

For between-caller differences in overall standard deviations, there were also no differences for participant emotionality variability ($F_{(6, 69)} = 1.11, n.s.$), surface acting variability ($F_{(6, 69)} = 0.76, n.s.$), or deep acting variability ($F_{(6, 69)} = 0.79, n.s.$). Moreover,
there were no differences between callers in the service failure condition (participant emotionality: $F_{(5, 32)} = 2.43, n.s.$; surface acting: $F_{(5, 32)} = 3.15, n.s.$; deep acting: $F_{(5, 32)} = 1.81, n.s.$) or service recovery condition (participant emotionality: $F_{(6, 31)} = .35, n.s.$; surface acting: $F_{(6, 31)} = .06, n.s.$; deep acting: $F_{(6, 31)} = .12, n.s.$).

Counterbalancing Effects

Given that half of the participants rated surface acting first and the other half rated deep acting first, independent samples t-tests were conducted to determine if the counterbalancing had any influence on the mean level and standard deviation of surface acting and deep acting for the entire performance episode across all participants and with participants split between condition. Participant emotionality was not included given that it was rated first for both conditions. Means and standard deviations for surface and deep acting for participants across the entire call and split by counterbalancing are in Table 4.
Table 4. Means and standard deviations of surface acting and deep acting split by counterbalancing.

<table>
<thead>
<tr>
<th></th>
<th>Entire Call</th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Surface Acting First</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>All Participants</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Acting</td>
<td>8.57</td>
<td>3.74</td>
<td>5.34</td>
<td>4.28</td>
</tr>
<tr>
<td>Deep Acting</td>
<td>7.95</td>
<td>3.93</td>
<td>4.91</td>
<td>3.71</td>
</tr>
<tr>
<td><em>Service Failure</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Acting</td>
<td>9.18</td>
<td>4.28</td>
<td>5.51</td>
<td>5.05</td>
</tr>
<tr>
<td>Deep Acting</td>
<td>7.89</td>
<td>4.54</td>
<td>5.03</td>
<td>4.39</td>
</tr>
<tr>
<td><em>Service Recovery</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Acting</td>
<td>7.97</td>
<td>3.11</td>
<td>4.79</td>
<td>3.01</td>
</tr>
<tr>
<td>Deep Acting</td>
<td>8.01</td>
<td>3.34</td>
<td>5.17</td>
<td>3.49</td>
</tr>
<tr>
<td><strong>Deep Acting First</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>All Participants</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Acting</td>
<td>7.95</td>
<td>3.73</td>
<td>4.39</td>
<td>3.89</td>
</tr>
<tr>
<td>Deep Acting</td>
<td>7.97</td>
<td>3.95</td>
<td>5.51</td>
<td>4.54</td>
</tr>
<tr>
<td><em>Service Failure</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Acting</td>
<td>8.08</td>
<td>3.33</td>
<td>3.81</td>
<td>3.18</td>
</tr>
<tr>
<td>Deep Acting</td>
<td>8.17</td>
<td>4.30</td>
<td>5.46</td>
<td>5.34</td>
</tr>
<tr>
<td><em>Service Recovery</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Acting</td>
<td>7.81</td>
<td>4.18</td>
<td>4.97</td>
<td>4.49</td>
</tr>
<tr>
<td>Deep Acting</td>
<td>7.76</td>
<td>3.67</td>
<td>5.57</td>
<td>3.72</td>
</tr>
</tbody>
</table>

*Note. M = mean. SD = standard deviation.*

At first glance of the means and standard deviations in Table 4, the means and standard deviations appear to be unaffected by the counterbalancing in the current study. Given the centrality of this counterbalancing to my study design, I conducted independent samples t-tests for counterbalancing effects across the entire call and Phases I, II, and III for all participants and split by condition. Results are depicted in Table 5.
Table 5. Independent samples t-tests for counterbalancing effects.

<table>
<thead>
<tr>
<th></th>
<th>All Participants</th>
<th>Service Failure</th>
<th>Service Recovery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
<td>t</td>
<td>t</td>
</tr>
<tr>
<td><strong>Entire Performance Episode</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Surface Acting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.73</td>
<td>.86</td>
<td>.13</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>-.62</td>
<td>-1.38</td>
<td>.64</td>
</tr>
<tr>
<td><strong>Deep Acting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-.02</td>
<td>-.20</td>
<td>.22</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.40</td>
<td>-.80</td>
<td>1.37</td>
</tr>
<tr>
<td><strong>Phase I</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Surface Acting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.02</td>
<td>1.25</td>
<td>.15</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.33</td>
<td>-1.39</td>
<td>1.82</td>
</tr>
<tr>
<td><strong>Deep Acting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-.63</td>
<td>-.27</td>
<td>-.71</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.24</td>
<td>.62</td>
<td>1.15</td>
</tr>
<tr>
<td><strong>Phase II</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Surface Acting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.72</td>
<td>.53</td>
<td>.48</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>-.52</td>
<td>-1.66</td>
<td>.64</td>
</tr>
<tr>
<td><strong>Deep Acting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.74</td>
<td>.04</td>
<td>1.00</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>-.21</td>
<td>-1.34</td>
<td>1.04</td>
</tr>
<tr>
<td><strong>Phase III</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Surface Acting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-.38</td>
<td>-.55</td>
<td>-.05</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>.37</td>
<td>-.25</td>
<td>.77</td>
</tr>
<tr>
<td><strong>Deep Acting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-.51</td>
<td>-.57</td>
<td>-.17</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.39</td>
<td>-2.09*</td>
<td>2.94**</td>
</tr>
</tbody>
</table>

**Note.** Degrees of freedom for analyses across all participants are 74; degrees of freedom for analyses within each of the study conditions are 36. All values are t-statistics for counterbalancing effects on the means and standard deviations of surface acting, and deep acting. * p < .05; ** p < .01; *** p < .001.

As shown in Table 5, there were no effects of counterbalancing on the means and standard deviations for surface acting and deep acting for the entire performance episode. These non-significant effects occurred across all participants and when participants were split between conditions. Additionally, when participants were split by phase of the performance episode, the non-significant effects continued for the means and standard
deviations of Phases I and II. For Phase III, there were two significant effects that emerged for the standard deviations of deep acting in the service failure ($t_{(36)} = -2.09, p < .05$) and service recovery ($t_{(36)} = 2.94, p < .01$) conditions. In the service failure condition, those who utilized deep acting first had slightly more variability ($M = .94, SD = .21$) than those who rated surface acting first ($M = .57, SD = .13$). In the service recovery condition, the opposite effect was found, with those who rated surface acting first having slightly more variability in deep acting ($M = 1.66, SD = .38$) than those who rated deep acting first ($M = 1.32, SD = .30$). Thus, given these conflicting findings and the largely non-significant differences for the means and standard deviations across all participants and participants split by condition for the entire performance episode and the three phases of the call, the counterbalancing of surface and deep acting was ruled as not affecting the means and standard deviations in any meaningful way.

As a final test, I calculated correlations among the focal variables split by the counterbalancing effect and study condition. Results from this analysis are presented in Table 6, with values for the service failure condition below the diagonal, and results for the service recovery condition above the diagonal, and in Table 7 for the performance episode split by phase of call and study condition (service failure is below diagonal; service recovery is above).
Table 6. Correlations among focal variables across the entire performance episode split by study condition and counterbalancing.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Surface Acting First</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Emotionality Avg.</td>
<td>--</td>
<td>.14</td>
<td>-.26</td>
<td>.09</td>
<td>-.13</td>
<td>.06</td>
<td>.16</td>
<td>-.17</td>
<td>-.47**</td>
</tr>
<tr>
<td>2. Emotionality SD</td>
<td>-.14</td>
<td>--</td>
<td>.21</td>
<td>.81**</td>
<td>.53*</td>
<td>.82**</td>
<td>.45</td>
<td>.12</td>
<td>.36</td>
</tr>
<tr>
<td>3. Surface Acting Avg.</td>
<td>-.53*</td>
<td>.39</td>
<td>--</td>
<td>.32</td>
<td>.67**</td>
<td>.17</td>
<td>.24</td>
<td>-.16</td>
<td>.25</td>
</tr>
<tr>
<td>4. Surface Acting SD</td>
<td>.02</td>
<td>.77**</td>
<td>.18</td>
<td>--</td>
<td>.58**</td>
<td>.88**</td>
<td>.13</td>
<td>.20</td>
<td>.33</td>
</tr>
<tr>
<td>5. Deep Acting Avg.</td>
<td>-.32</td>
<td>.27</td>
<td>.76**</td>
<td>-.04</td>
<td>--</td>
<td>.61**</td>
<td>.26</td>
<td>.08</td>
<td>.20</td>
</tr>
<tr>
<td>6. Deep Acting SD</td>
<td>-.37</td>
<td>.49*</td>
<td>.30</td>
<td>.46*</td>
<td>.46*</td>
<td>--</td>
<td>.19</td>
<td>.28</td>
<td>.45</td>
</tr>
<tr>
<td>7. T-P Affect Avg.</td>
<td>-.02</td>
<td>.12</td>
<td>-.01</td>
<td>.03</td>
<td>.03</td>
<td>.37</td>
<td>--</td>
<td>-.49*</td>
<td>-.01</td>
</tr>
<tr>
<td>8. T-P Affect SD</td>
<td>-.24</td>
<td>-.10</td>
<td>.39</td>
<td>-.15</td>
<td>-.15</td>
<td>-.24</td>
<td>-.25</td>
<td>--</td>
<td>.23</td>
</tr>
<tr>
<td>9. Emo. Exh.</td>
<td>-.62**</td>
<td>.11</td>
<td>.51*</td>
<td>.20</td>
<td>.20</td>
<td>.40</td>
<td>.24</td>
<td>.23</td>
<td>--</td>
</tr>
<tr>
<td><strong>Deep Acting First</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Emotionality Avg.</td>
<td>--</td>
<td>-.14</td>
<td>-.31</td>
<td>-.19</td>
<td>-.51*</td>
<td>-.46*</td>
<td>.59**</td>
<td>-.19</td>
<td>.07</td>
</tr>
<tr>
<td>2. Emotionality SD</td>
<td>-.16</td>
<td>--</td>
<td>.33</td>
<td>.60**</td>
<td>.17</td>
<td>.62**</td>
<td>.10</td>
<td>.22</td>
<td>.30</td>
</tr>
<tr>
<td>3. Surface Acting Avg.</td>
<td>.21</td>
<td>.52*</td>
<td>--</td>
<td>.17</td>
<td>.82**</td>
<td>.03</td>
<td>-.16</td>
<td>.15</td>
<td>.11</td>
</tr>
<tr>
<td>4. Surface Acting SD</td>
<td>.02</td>
<td>.73**</td>
<td>.66**</td>
<td>--</td>
<td>-.09</td>
<td>.76**</td>
<td>.07</td>
<td>-.09</td>
<td>.19</td>
</tr>
<tr>
<td>5. Deep Acting Avg.</td>
<td>.07</td>
<td>.48*</td>
<td>.85**</td>
<td>.41</td>
<td>--</td>
<td>.02</td>
<td>-.38</td>
<td>.27</td>
<td>.15</td>
</tr>
<tr>
<td>6. Deep Acting SD</td>
<td>-.29</td>
<td>.83**</td>
<td>.39</td>
<td>.73**</td>
<td>.27</td>
<td>--</td>
<td>-.09</td>
<td>.01</td>
<td>.36</td>
</tr>
<tr>
<td>7. T-P Affect Avg.</td>
<td>.05</td>
<td>.23</td>
<td>.19</td>
<td>.41</td>
<td>.13</td>
<td>.29</td>
<td>--</td>
<td>-.32</td>
<td>.05</td>
</tr>
<tr>
<td>8. T-P Affect SD</td>
<td>.16</td>
<td>-.17</td>
<td>-.22</td>
<td>-.41</td>
<td>-.06</td>
<td>-.23</td>
<td>-.45</td>
<td>--</td>
<td>.19</td>
</tr>
<tr>
<td>9. Emo. Exh.</td>
<td>-.45</td>
<td>.61*</td>
<td>.17</td>
<td>.34</td>
<td>.28</td>
<td>.49*</td>
<td>.03</td>
<td>-.35</td>
<td>--</td>
</tr>
</tbody>
</table>

*Note.* Service failure correlations are below the diagonal; service recovery conditions are above the diagonal. Avg. = average. SD = standard deviation. T-P = third-party. Emo. Exh. = emotional exhaustion. *p < .05; **p < .01.
Table 7. Correlations among focal variables within the three phases of the performance episode split by study condition and counterbalancing.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Surface Acting First</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phase I: Emotionality</td>
<td>--</td>
<td>-.28</td>
<td>-.21</td>
<td>.54*</td>
<td>.17</td>
<td>.17</td>
<td>.27</td>
<td>.43</td>
</tr>
<tr>
<td>2. Phase I: SA</td>
<td>-.46*</td>
<td>--</td>
<td>.64**</td>
<td>-.04</td>
<td>.24</td>
<td>.35</td>
<td>.02</td>
<td>.26</td>
</tr>
<tr>
<td>3. Phase I: DA</td>
<td>-.27</td>
<td>.89**</td>
<td>--</td>
<td>-.01</td>
<td>.06</td>
<td>.48*</td>
<td>.39</td>
<td>.19</td>
</tr>
<tr>
<td>4. Phase I: T-P Affect</td>
<td>-.10</td>
<td>.21</td>
<td>.24</td>
<td>--</td>
<td>-.28</td>
<td>.30</td>
<td>.43</td>
<td>.76**</td>
</tr>
<tr>
<td>5. Phase II: Emotionality</td>
<td>.61**</td>
<td>-.35</td>
<td>-.16</td>
<td>-.18</td>
<td>--</td>
<td>-.42</td>
<td>-.39</td>
<td>-.11</td>
</tr>
<tr>
<td>6. Phase II: SA</td>
<td>-.23</td>
<td>.55*</td>
<td>.51*</td>
<td>.09</td>
<td>-.60**</td>
<td>--</td>
<td>.70**</td>
<td>.14</td>
</tr>
<tr>
<td>7. Phase II: DA</td>
<td>-.10</td>
<td>.65**</td>
<td>.71*</td>
<td>.32</td>
<td>-.30</td>
<td>.74**</td>
<td>--</td>
<td>.26</td>
</tr>
<tr>
<td>8. Phase II: T-P Affect</td>
<td>.01</td>
<td>.02</td>
<td>.11</td>
<td>.65**</td>
<td>-.03</td>
<td>-.07</td>
<td>.45</td>
<td>--</td>
</tr>
<tr>
<td>9. Phase III: Emotionality</td>
<td>.32</td>
<td>-.27</td>
<td>-.24</td>
<td>-.17</td>
<td>.84**</td>
<td>-.59</td>
<td>-.49*</td>
<td>-.16</td>
</tr>
<tr>
<td>10. Phase III: SA</td>
<td>-.01</td>
<td>.25</td>
<td>.31</td>
<td>.24</td>
<td>-.43</td>
<td>.84**</td>
<td>.65**</td>
<td>.14</td>
</tr>
<tr>
<td>11. Phase III: DA</td>
<td>-.05</td>
<td>.29</td>
<td>.49</td>
<td>.31</td>
<td>-.26</td>
<td>.52*</td>
<td>.86**</td>
<td>.55*</td>
</tr>
<tr>
<td>12. Phase III: T-P Affect</td>
<td>.17</td>
<td>-.40</td>
<td>-.39</td>
<td>.29</td>
<td>.01</td>
<td>-.20</td>
<td>-.08</td>
<td>.70**</td>
</tr>
<tr>
<td>13. Emo. Exh.</td>
<td>-.56*</td>
<td>.30</td>
<td>.27</td>
<td>.31</td>
<td>-.56*</td>
<td>.58*</td>
<td>.40</td>
<td>.20</td>
</tr>
<tr>
<td><strong>Deep Acting First</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phase I: Emotionality</td>
<td>--</td>
<td>-.11</td>
<td>-.21</td>
<td>.58**</td>
<td>.67**</td>
<td>-.18</td>
<td>-.45</td>
<td>.55</td>
</tr>
<tr>
<td>2. Phase I: SA</td>
<td>.12</td>
<td>--</td>
<td>.84**</td>
<td>-.11</td>
<td>.06</td>
<td>.38</td>
<td>.30</td>
<td>-.02</td>
</tr>
<tr>
<td>3. Phase I: DA</td>
<td>.22</td>
<td>.84**</td>
<td>--</td>
<td>-.26</td>
<td>.01</td>
<td>.33</td>
<td>.35</td>
<td>-.16</td>
</tr>
<tr>
<td>4. Phase I: T-P Affect</td>
<td>.17</td>
<td>-.12</td>
<td>-.07</td>
<td>--</td>
<td>.43</td>
<td>-.13</td>
<td>-.30</td>
<td>.85**</td>
</tr>
<tr>
<td>5. Phase II: Emotionality</td>
<td>.63**</td>
<td>.25</td>
<td>.17</td>
<td>.01</td>
<td>--</td>
<td>-.59</td>
<td>-.78</td>
<td>.52*</td>
</tr>
<tr>
<td>6. Phase II: SA</td>
<td>.53*</td>
<td>.39</td>
<td>.38</td>
<td>.21</td>
<td>.18</td>
<td>--</td>
<td>.82**</td>
<td>-.21</td>
</tr>
<tr>
<td>7. Phase II: DA</td>
<td>.33</td>
<td>.51*</td>
<td>.60**</td>
<td>.26</td>
<td>-.12</td>
<td>.83**</td>
<td>--</td>
<td>-.34</td>
</tr>
<tr>
<td>8. Phase II: T-P Affect</td>
<td>.10</td>
<td>-.08</td>
<td>-.03</td>
<td>.58**</td>
<td>-.05</td>
<td>.45</td>
<td>.38</td>
<td>--</td>
</tr>
<tr>
<td>9. Phase III: Emotionality</td>
<td>.17</td>
<td>.28</td>
<td>.21</td>
<td>-.11</td>
<td>.82**</td>
<td>-.07</td>
<td>-.30</td>
<td>-.19</td>
</tr>
<tr>
<td>10. Phase III: SA</td>
<td>.42</td>
<td>.16</td>
<td>.25</td>
<td>.13</td>
<td>-.19</td>
<td>.80**</td>
<td>.75**</td>
<td>.36</td>
</tr>
<tr>
<td>11. Phase III: DA</td>
<td>.32</td>
<td>.23</td>
<td>.32</td>
<td>.21</td>
<td>-.29</td>
<td>.74**</td>
<td>.88**</td>
<td>.32</td>
</tr>
<tr>
<td>12. Phase III: T-P Affect</td>
<td>-.07</td>
<td>-.03</td>
<td>-.02</td>
<td>.36</td>
<td>-.18</td>
<td>.26</td>
<td>.26</td>
<td>.92**</td>
</tr>
<tr>
<td>13. Emo. Exh.</td>
<td>.05</td>
<td>-.12</td>
<td>.06</td>
<td>-.12</td>
<td>-.46*</td>
<td>.12</td>
<td>.27</td>
<td>.15</td>
</tr>
</tbody>
</table>
Table 7. Correlations among focal variables across the entire performance episode split by study condition and counterbalancing (cont.).

<table>
<thead>
<tr>
<th></th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Surface Acting First</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phase I: Emotionality</td>
<td>.47*</td>
<td>-.13</td>
<td>.03</td>
<td>.34</td>
<td>-.20</td>
</tr>
<tr>
<td>2. Phase I: SA</td>
<td>-.01</td>
<td>.40</td>
<td>.04</td>
<td>.14</td>
<td>-.05</td>
</tr>
<tr>
<td>3. Phase I: DA</td>
<td>-.01</td>
<td>.33</td>
<td>.46*</td>
<td>-.10</td>
<td>-.05</td>
</tr>
<tr>
<td>4. Phase I: T-P Affect</td>
<td>.23</td>
<td>.36</td>
<td>.39</td>
<td>.70**</td>
<td>.05</td>
</tr>
<tr>
<td>5. Phase II: Emotionality</td>
<td>.59**</td>
<td>-.56**</td>
<td>-.56**</td>
<td>.13</td>
<td>-.57**</td>
</tr>
<tr>
<td>6. Phase II: SA</td>
<td>-.00</td>
<td>.52*</td>
<td>.45</td>
<td>-.06</td>
<td>.32</td>
</tr>
<tr>
<td>7. Phase II: DA</td>
<td>.18</td>
<td>.34</td>
<td>.61**</td>
<td>.07</td>
<td>.34</td>
</tr>
<tr>
<td>8. Phase II: T-P Affect</td>
<td>.13</td>
<td>.17</td>
<td>.22</td>
<td>.74**</td>
<td>.34**</td>
</tr>
<tr>
<td>9. Phase III: Emotionality</td>
<td>--</td>
<td>-.52*</td>
<td>-.35</td>
<td>.21</td>
<td>-.26</td>
</tr>
<tr>
<td>10. Phase III: SA</td>
<td>-.58*</td>
<td>--</td>
<td>.72**</td>
<td>.01</td>
<td>.30</td>
</tr>
<tr>
<td>11. Phase III: DA</td>
<td>-.52*</td>
<td>.57*</td>
<td>--</td>
<td>-.03</td>
<td>.18</td>
</tr>
<tr>
<td>12. Phase III: T-P Affect</td>
<td>.00</td>
<td>-.03</td>
<td>.22</td>
<td>--</td>
<td>-.11</td>
</tr>
<tr>
<td>13. Emo. Exh.</td>
<td>-.43</td>
<td>.47*</td>
<td>.35</td>
<td>.06</td>
<td>--</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deep Acting First</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Phase I: Emotionality</td>
<td>.84**</td>
<td>-.43</td>
<td>-.35</td>
<td>.54*</td>
<td>.05</td>
</tr>
<tr>
<td>2. Phase I: SA</td>
<td>-.03</td>
<td>.51*</td>
<td>.54*</td>
<td>-.06</td>
<td>.23</td>
</tr>
<tr>
<td>3. Phase I: DA</td>
<td>-.20</td>
<td>.58**</td>
<td>.78**</td>
<td>-.24</td>
<td>.21</td>
</tr>
<tr>
<td>4. Phase I: T-P Affect</td>
<td>.61**</td>
<td>-.34</td>
<td>-.44</td>
<td>.81**</td>
<td>.12</td>
</tr>
<tr>
<td>5. Phase II: Emotionality</td>
<td>.69**</td>
<td>-.59**</td>
<td>-.43</td>
<td>.46*</td>
<td>-.00</td>
</tr>
<tr>
<td>6. Phase II: SA</td>
<td>-.25</td>
<td>.83**</td>
<td>.56*</td>
<td>-.21</td>
<td>-.03</td>
</tr>
<tr>
<td>7. Phase II: DA</td>
<td>-.57*</td>
<td>.85**</td>
<td>.68**</td>
<td>-.41</td>
<td>.08</td>
</tr>
<tr>
<td>8. Phase II: T-P Affect</td>
<td>.59**</td>
<td>-.30</td>
<td>-.46</td>
<td>.88**</td>
<td>.04</td>
</tr>
<tr>
<td>9. Phase III: Emotionality</td>
<td>--</td>
<td>-.34</td>
<td>-.53*</td>
<td>.59*</td>
<td>.23</td>
</tr>
<tr>
<td>10. Phase III: SA</td>
<td>-.45</td>
<td>--</td>
<td>.68**</td>
<td>-.33</td>
<td>.05</td>
</tr>
<tr>
<td>11. Phase III: DA</td>
<td>-.59**</td>
<td>.87**</td>
<td>--</td>
<td>-.52*</td>
<td>.04</td>
</tr>
<tr>
<td>12. Phase III: T-P Affect</td>
<td>-.26</td>
<td>.26</td>
<td>.23</td>
<td>--</td>
<td>.00</td>
</tr>
<tr>
<td>13. Emo. Exh.</td>
<td>-.74**</td>
<td>.35</td>
<td>.45</td>
<td>.24</td>
<td>--</td>
</tr>
</tbody>
</table>

*Note*. Service failure correlations are below the diagonal; service recovery conditions are above the diagonal. Avg. = average. SD = standard deviation. T-P = third-party. Emo. Exh. = emotional exhaustion. *p < .05, **p < .01.

Although Table 6 offers initial insight, the results in Table 7 are more important to interpret given the differences in social context within the two study conditions. As a note, given that there were only 19 people per group being analyzed, it was less of a concern about significance of effects, and more of a concern about effects of differing directions. For the service recovery condition, there were no differences to report across the three phases of the call. There were only a couple notable exceptions worth
mentioning for the service failure condition. Specifically, during Phase I, for those who rated surface acting first, there was a negative relationship between surface acting and emotionality \((r = -.46, p < .05)\) and between deep acting and emotionality \((r = -.27, n.s.)\). For those who rated deep acting first, the relationship was non-significant between emotionality and surface acting \((r = .12, n.s.)\), and deep acting \((r = .22, n.s.)\), with the relationships being positive in direction. However, these conflicting results could be due to the lack of variation and social information occurring during the early phase of the call. Additionally, for Phase II, for those who rated surface acting first, there was a negative relationship between surface acting and emotionality \((r = -.60, p < .05)\) and deep acting and emotionality \((r = -.30, n.s.)\); for service recovery, the relationships were again non-significant (surface acting and emotionality: \(r = .18, n.s.;\) deep acting and emotionality: \(r = -.12, n.s.\)). No differences were found in Phase III. Thus, although there were some differences, there was no consistent pattern that emerged that would raise concerns about the effects of counterbalancing.
CHAPTER IV
RESULTS

Means, standard deviations, and correlations for the focal variables (i.e., duration of the performance episode in seconds, averages of the continuous ratings, emotional exhaustion, switching between emotion regulation strategies) are presented in Table 8.

Table 8. Means, standard deviations, and correlations across the entire performance episode.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Call Duration (Seconds)</td>
<td>264.75</td>
<td>48.09</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Participant Emotionality (Avg.)</td>
<td>10.87</td>
<td>3.01</td>
<td>-17</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Emotionality SD</td>
<td>3.05</td>
<td>1.64</td>
<td>-09</td>
<td>-06</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Surface Acting (Avg.)</td>
<td>8.26</td>
<td>3.73</td>
<td>-06</td>
<td>-30**</td>
<td>.33**</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Surface Acting SD</td>
<td>4.32</td>
<td>2.11</td>
<td>-04</td>
<td>--</td>
<td>.72**</td>
<td>.29*</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Deep Acting (Avg.)</td>
<td>7.96</td>
<td>3.91</td>
<td>-00</td>
<td>-23**</td>
<td>.36**</td>
<td>.76**</td>
<td>.20</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Deep Acting SD</td>
<td>3.89</td>
<td>2.04</td>
<td>-09</td>
<td>--</td>
<td>.71**</td>
<td>.20</td>
<td>.70**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Third-Party Affect (Avg.)</td>
<td>10.66</td>
<td>0.65</td>
<td>-37**</td>
<td>0.19</td>
<td>.20</td>
<td>.05</td>
<td>.14</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Third-Party Affect SD</td>
<td>0.70</td>
<td>0.28</td>
<td>14</td>
<td>-16</td>
<td>.03</td>
<td>.09</td>
<td>--</td>
<td>.09</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Emotional Exhaustion</td>
<td>2.50</td>
<td>1.10</td>
<td>10</td>
<td>-36**</td>
<td>.28*</td>
<td>.27*</td>
<td>.24*</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Switching ‘In General’</td>
<td>3.02</td>
<td>2.64</td>
<td>01</td>
<td>.04</td>
<td>.26*</td>
<td>.07</td>
<td>.27*</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Switching DA → SA</td>
<td>1.46</td>
<td>1.45</td>
<td>-01</td>
<td>.02</td>
<td>.22</td>
<td>.10</td>
<td>.25*</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Switching SA → DA</td>
<td>1.57</td>
<td>1.36</td>
<td>02</td>
<td>.05</td>
<td>.26*</td>
<td>.03</td>
<td>.26*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Means, standard deviations, and correlations are for focal (i.e., hypothesis) variables only. Avg. = average. SD = standard deviation. SA = surface acting. DA = deep acting. Emotional exhaustion was the only variable assessed at the end of the session.

*p < .05; **p < .01.
In regards to descriptive statistics, on average participants reported slightly lower than “moderate” levels (on the rating scale) of surface acting ($M = 8.26$, $SD = 3.73$) and deep acting ($M = 7.96$, $SD = 3.91$) across the entire phone call. Further, they reported moderately neutral emotionality ($M = 10.87$, $SD = 3.01$), and were rated as fairly neutral in the third-party affect ratings ($M = 10.66$, $SD = 0.43$). Participants experienced moderate levels of emotional exhaustion at the end of the call ($M = 2.50$, $SD = 1.10$).

Hypothesis 1

Hypothesis 1 proposed a negative relationship between surface acting and deep acting at the momentary level of analysis (i.e., within the performance episode). Cross-correlations are presented in Table 9. Though no effects were hypothesized, I also present results from participant emotionality during the performance episode.

Table 9. Cross-correlation analyses for focal variables across all participants.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Surface Acting</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Deep Acting</td>
<td>.64***</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>3. Participant Emotionality</td>
<td>- .51***</td>
<td>-.48***</td>
<td>--</td>
</tr>
</tbody>
</table>

*Note. Cross-correlations were tested with $df = 74$. † $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.*

Contrary to Hypothesis 1, surface acting and deep acting were positively related within the performance episode ($r_{(74)} = .64$, $p < .001$). Although no specific effects were hypothesized, both surface acting and deep acting negatively covaried with participant ratings of emotionality across the entire call ($r_{(74)} = -.51$, $p < .001$ and $r_{(74)} = -.48$, $p <$
suggesting that when participants felt more negatively during the call, their levels of surface and deep acting were higher.

To better understand whether the study manipulation impacted the average cross-correlations, I also examined cross-correlations separately for each between-subjects condition (i.e., service failure vs. service recovery), for each within-subjects condition (i.e., the three study phases), and for each of the resulting six cells in the study design. Results indicated that the cross-correlation between surface acting and deep acting across the entire performance episode remained positive for both the service failure \( r_{(34)} = .63, p < .001 \) and service recovery \( r_{(34)} = .65, p < .001 \) conditions, both of which were roughly the same as the overall cross-correlation noted above \( r_{(74)} = .64, p < .001 \).

Table 10. Cross-correlation analyses for focal variables for service failure and service recovery conditions.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Surface Acting</td>
<td>--</td>
<td>.65***</td>
<td>-.48***</td>
</tr>
<tr>
<td>2. Deep Acting</td>
<td>.63***</td>
<td>--</td>
<td>-.49***</td>
</tr>
<tr>
<td>3. Participant Emotionality</td>
<td>-.54***</td>
<td>-.47***</td>
<td>--</td>
</tr>
</tbody>
</table>

Note. Cross-correlations were tested with \( df = 34 \). Cross-correlations for service failure are below the diagonal; cross-correlations for service recovery are above the diagonal. † \( p < .10; * p < .05; ** p < .01; *** p < .001 \).
Indeed, the correlations between surface and deep acting in the two conditions were not significantly different from each other ($z = .07, n.s.$). Further, I found that surface and deep acting were negatively correlated with participant emotionality in the service failure condition (surface acting: $r(34) = -.54, p < .001$; deep acting: $r(34) = -.47, p < .001$) and service recovery condition (surface acting: $r(34) = -.48, p < .001$; deep acting: $r(34) = -.49, p < .001$); the relationships of emotionality with surface acting ($z = -.34, n.s.$) and deep acting ($z = .11, n.s.$) did not differ across the two conditions.

Next, I examined whether the cross-correlations differed as a function of the within-subject manipulation of phases (i.e., first, second, third) across all participants. Results can be found in Table 11.

Table 11. Cross-correlation analyses within phases of the performance episode across all participants.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phase I</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Surface Acting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Deep Acting</td>
<td>.26*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Participant Emotionality</td>
<td>-.15</td>
<td>-.06</td>
<td></td>
</tr>
<tr>
<td><strong>Phase II</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Surface Acting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Deep Acting</td>
<td>.41***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Participant Emotionality</td>
<td>-.33***</td>
<td>-.28**</td>
<td></td>
</tr>
<tr>
<td><strong>Phase III</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Surface Acting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Deep Acting</td>
<td>.34***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Participant Emotionality</td>
<td>-.31***</td>
<td>-.29**</td>
<td></td>
</tr>
</tbody>
</table>

*Note. Cross-correlations were tested with $df = 74$. † $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$. 

90
Ignoring the between-subject manipulation, the relationship between surface acting and deep acting remained positive across all three phases of the call (Phase I: \( r_{(74)} = .26, p < .05 \); Phase II: \( r_{(74)} = .41, p < .001 \); Phase III: \( r_{(74)} = .34, p < .001 \)), which remains counter to Hypothesis 1. To test whether the cross-correlations (which were dependent, not independent; Bobko, 2001) were significantly different from each other, I conducted a repeated measures analysis of variance, with the average Fisher’s z values (used prior to the conversion back to average cross-correlation values), with ‘Phase of Call’ as the factor of interest. Results revealed a non-significant within-subjects effect \((F_{(2,86)} = 1.16, n.s.)\), demonstrating that the relationship between surface acting and deep acting did not vary as a function of the call phase.

Moreover, when looking at the relationships of surface acting and deep acting with participant emotionality across the three phases (see Table 11), the relationships of emotionality with surface acting and deep acting were non-significant in Phase I \( (r_{(34)} = -.15, n.s., \text{and} \ r_{(34)} = -.06, n.s., \text{respectively}) \), significant in Phase II \( (r_{(34)} = -.33, p < .001, \text{and} \ r_{(34)} = -.28, p < .01, \text{respectively}) \), and significant in Phase III \( (r_{(34)} = -.31, p < .001, \text{and} \ r_{(34)} = -.29, p < .01, \text{respectively}) \). Analyses were conducted to determine if these correlations were significantly different from one another across the three phases; results were not supportive of a difference for surface acting \( (F_{(2,88)} = 2.46, n.s.) \), but they were supportive for deep acting \( (F_{(2,90)} = 6.51, p < .01, \eta^2_p = .13) \). Pairwise comparisons revealed that the cross-correlation between participant emotionality and deep acting during Phase I (see Table 11) was significantly different from Phase II \( (p < .05) \) and Phase III \( (p < .01) \), with the relationship being stronger during Phases II and III than in
Phase I; the cross-correlation at Phase II was not significantly different from the cross-
correlation at Phase III.

Though these results suggest that participant emotionality and deep acting
covaried more strongly as the call became more difficult, it is important to also factor in
the between-subject manipulation of service recovery and service failure, which took
place in the third phase of each call. Thus, a final set of supplemental cross-correlation
analyses was examined by dividing the data by both the within-subject variable of call
phase and the between-subject variable of recovery-failure. Results from this way of
splitting the dataset are in Table 12; cross-correlations for service failure are below the
diagonal, and cross-correlations for service recovery are above the diagonal.

Table 12. Cross-correlation analyses within phases of the performance episode split by
condition.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phase I</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Surface Acting</td>
<td>--</td>
<td>.36*</td>
<td>-.19</td>
</tr>
<tr>
<td>2. Deep Acting</td>
<td>.14</td>
<td>--</td>
<td>-.13</td>
</tr>
<tr>
<td>3. Participant Emotionality</td>
<td>-.10</td>
<td>.03</td>
<td>--</td>
</tr>
<tr>
<td><strong>Phase II</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Surface Acting</td>
<td>--</td>
<td>.43***</td>
<td>-.34*</td>
</tr>
<tr>
<td>2. Deep Acting</td>
<td>.39**</td>
<td>--</td>
<td>-.27†</td>
</tr>
<tr>
<td>3. Participant Emotionality</td>
<td>-.32*</td>
<td>-.30†</td>
<td>--</td>
</tr>
<tr>
<td><strong>Phase III</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Surface Acting</td>
<td>--</td>
<td>.41**</td>
<td>-.28†</td>
</tr>
<tr>
<td>2. Deep Acting</td>
<td>.21</td>
<td>--</td>
<td>-.38*</td>
</tr>
<tr>
<td>3. Participant Emotionality</td>
<td>-.34*</td>
<td>-.14</td>
<td>--</td>
</tr>
</tbody>
</table>

*Note. All cross-correlations were tested with df = 34. Cross-correlations for service
failure are below the diagonal, and cross-correlations for service recovery are above the
diagonal. † p < .10; * p < .05; ** p < .01; *** p < .001.
As shown in Table 12, the relationship between surface acting and deep acting tends to vary as a function of the failure-recovery manipulation and the phase of the call. For the service failure condition, I found no relationship between surface and deep acting during Phase I ($r_{(34)} = .14, n.s.$), a positive relationship during Phase II ($r_{(34)} = .39, p < .01$), and no relationship during Phase III ($r_{(34)} = .21, n.s.$). Though there appeared to be variation, the cross-correlations for these phases were not significantly different from each other ($F_{(2,36)} = 1.41, n.s.$). For the service recovery condition, positive relationships emerged between surface and deep acting occurring during Phase I ($r_{(34)} = .36, p < .05$), Phase II ($r_{(34)} = .43, p < .001$), and Phase III ($r_{(34)} = .41, p < .01$) of the call; the phases also were not significantly different from each other ($F_{(2,48)} = .28, n.s.$). Moreover, utilizing Preacher’s (2002) test of independent correlations, the relationship between surface and deep acting did not differ between conditions during Phase I ($z = -.99, n.s.$), Phase II ($z = -.20, n.s.$), or Phase III ($z = -.93, n.s.$). However, the lack of difference between phases could be attributable to low levels of power in the analyses. With this in mind, it is interesting to note that the relationship between surface acting and deep acting during Phase III of the call (the between-subject manipulation) is significant for those in the service failure condition and non-significant for those in the service recovery condition. Thus, the positive relationship between surface and deep acting could be contingent upon the social information being provided by the confederate callers to the participants. This idea is considered in more depth during tests for Hypotheses 3 and 4.

Finally, I explored the relationship between participant emotionality and both emotion regulation strategies within each of the three study phases and across the failure-recovery conditions (see Table 12). For the service failure condition, participant
emotionality was not significantly related to surface ($r_{(34)} = -.10, n.s.$) or deep acting ($r_{(34)} = -.03, n.s.$) during Phase I. However, during Phase II, participant emotionality was negatively related to surface acting ($r_{(34)} = -.32, p < .05$) and deep acting ($r_{(34)} = -.30, p < .10$), though the cross-correlation with deep acting was only marginally significant. This result suggests that as the performance episode experience became negative, negative emotions increased in participants, enhancing the need for both forms of emotion regulation. Finally, for Phase III, while surface acting was negatively related to participant emotionality ($r_{(34)} = -.34, p < .05$), deep acting was unrelated ($r_{(34)} = -.14, n.s.$). The relationship between participant emotionality and surface acting during the service failure condition did not differ by phase ($F_{(2,36)} = 1.60, n.s.$), though the relationship between participant emotionality and deep acting did differ across phases ($F_{(2,34)} = 3.95, p < .05, \eta_p^2 = .19$). Pairwise comparisons revealed that the cross-correlation for participant emotionality and deep acting during the service failure condition at Phase I was significantly smaller than the cross-correlation found during Phase II ($p < .05$), but not Phase III (see Table 12). Phases II and III also were not significantly different from each other. These findings suggest that for the service failure condition, surface acting consistently correlated with felt emotions, with surface acting increasing as participants felt more negatively. However, deep acting only appears to be related to felt emotions when the tone of the confederate caller initially changed from being neutral to negative.

For the service recovery condition (see Table 12), similar to the service failure condition, there was no relationship of surface acting ($r_{(34)} = -.19, n.s.$) or deep acting ($r_{(34)} = -.13, n.s.$) with participant emotionality for Phase I of the call, which fits with the
neutral social context experienced during Phase I. For Phase II, a negative cross-correlation emerged for surface acting and participant emotionality ($r_{(34)} = -.34, p < .05$); the relationship between deep acting and participant emotionality was also negative, but marginally significant ($r_{(34)} = -.27, p < .10$). Finally, for Phase III, a marginal negative relationship was found between surface acting and participant emotionality ($r_{(34)} = -.28, p < .10$), along with a negative relationship between deep acting and participant emotionality ($r_{(34)} = -.38, p < .05$). This counters the non-significant effect found between deep acting and participant emotionality in Phase III for the service failure condition, suggesting that as the confederate caller became more pleasant, the level of regulation participants needed to put forth declined as their felt emotionality became more positive.

There were no differences between phases for the relationship between participant emotionality and surface acting ($F_{(2,36)} = 1.16, n.s.$), but similar to the service failure condition, there was a marginally significant difference between phases for the relationship between participant emotionality and deep acting ($F_{(2,54)} = 3.15, p = .051, \eta^2_p = .10$). Pairwise comparisons revealed that the cross-correlation between participant emotionality and deep acting at Phase I was marginally different from the cross-correlation for Phase III ($p = .079$), with the relationship being stronger during Phase III (see Table 12). There was no difference between Phases I and II, or Phases II and III.

As a final comparison, I tested whether the cross-correlation of surface acting and deep acting with participant emotionality for each phase differed across the service failure and service recovery conditions. For Phase I, the cross-correlation between surface acting and participant emotionality was not significantly different between conditions ($z = -.39, n.s.$), with a similar non-significant effect being found for the cross-
correlation between deep acting and participant emotionality \( (z = .42, n.s.) \). This pattern of non-significant effects across the recovery-failure conditions also was observed for Phase II (surface acting/participant emotionality, \( z = .09, n.s. \); deep acting/participant emotionality, \( z = -.14, n.s. \)) and Phase III (surface acting/participant emotionality, \( z = -.28, n.s. \); deep acting/participant emotionality: \( z = 1.08, n.s. \)). However, as noted previously, the lack of differences could be due to low power because of the small sample size at this level of analysis. A better understanding of what could be causing these differences in the magnitude of relationships is explored during tests related to Hypotheses 3 and 4.

Overall, across the entire performance episode and within the phases of the episode, those in the service failure and recovery conditions had similar experiences in that they utilized both surface acting and deep acting when interacting with the confederate caller (i.e., a consistent positive cross-correlation between emotion regulation strategies). However, these cross-correlations were not nearly as high as the ones found across all participants ignoring phase of call, suggesting that the social information participants were receiving had an effect on the magnitude of the relationship between surface and deep acting strategies. Thus, the positive relationship between regulation strategies may be weakened to the extent that customers become increasingly negative. Although a positive relationship between the two strategies was found during Phase II for both conditions, the positive relationship remained only for participants in the service recovery condition, meaning that participants in the service failure condition may have begun deploying different levels of the two regulation strategies. A more nuanced understanding of these results will be gained during the test of Hypothesis 5 (i.e.,
switching between surface/deep acting) when I examine whether there is greater variation in which strategy is used most often across the entire call, the different phases of the call, and the between-subject conditions.

Further, although no formal hypotheses were made, across the entire call, a consistent negative relationship was found between participant emotionality and both regulation strategies, suggesting that as felt emotions became increasingly negative, levels of both strategies increased to accommodate the emotional state the participant was experiencing. However, it is important to note that this strong, negative relationship was reduced when looking at participants split between study conditions and phase of performance episode, suggesting that the social context can impact the strength of these relationships. Specifically, both experimental conditions produced negative relationships between participant emotionality and surface acting for Phases II and III of the call (see Table 12). However, for deep acting, a significant negative relationship was found only for the service failure condition during Phase II; there was a significant negative relationship for Phases II and III within the service recovery condition (see Table 12). This pattern of results suggests that participant emotionality and level of effort aimed at actually feeling positively changed as the emotional tone of the caller shifted (i.e., from neutral → negative in Phase II for both conditions; from negative → positive in Phase III for the service recovery condition). The non-significant relationship between deep acting and participant emotionality during Phase III for the service failure condition could be due to the fact that the confederate emotionality did not change as much as occurred in the previous phases of the call (i.e., the caller was still negative in phase III).
Hypothesis 2

Hypothesis 2 proposed that at the episode-level (i.e., continuous ratings aggregated across the entire call), the relationship between surface acting and deep acting would be positive. Correlations for the entire call across both conditions are presented previously in Table 8. A strong, positive correlation was found between average levels of surface acting and deep acting across the entire performance episode ($r = .76$, $p < .01$), supporting Hypothesis 2.

I also tested correlations among average surface acting and deep acting for the entire performance episode split by condition (Table 13), split by phases of the call across both conditions (Table 14), and split by phases of the call and condition (Table 15). For Tables 13 and 15 where analyses were split by condition, service failure correlations are below the diagonal and service recovery correlations are above the diagonal. As a note, these tables only include the averaged values of the continuous rating scales (i.e., participant emotionality, surface acting, deep acting, third-party affect).

Table 13. Means, standard deviations, and correlations for main focal variables split by condition.

<table>
<thead>
<tr>
<th></th>
<th>Mean Failure</th>
<th>SD Failure</th>
<th>Mean Recovery</th>
<th>SD Recovery</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Duration (Seconds)</td>
<td>255.38</td>
<td>51.28</td>
<td>274.13</td>
<td>43.39</td>
<td>--</td>
<td>-.46*</td>
<td>-.08</td>
<td>.15</td>
<td>-.33*</td>
</tr>
<tr>
<td>2. Participant Emo.</td>
<td>10.32</td>
<td>3.21</td>
<td>11.42</td>
<td>2.71</td>
<td>-.04</td>
<td>--</td>
<td>-.29</td>
<td>-.36*</td>
<td>.41**</td>
</tr>
<tr>
<td>3. Surface Acting</td>
<td>8.63</td>
<td>3.83</td>
<td>7.89</td>
<td>3.63</td>
<td>-.01</td>
<td>-.28</td>
<td>--</td>
<td>.75**</td>
<td>.19</td>
</tr>
<tr>
<td>4. Deep Acting</td>
<td>8.03</td>
<td>4.36</td>
<td>7.89</td>
<td>3.46</td>
<td>-.09</td>
<td>-.15</td>
<td>.78*</td>
<td>--</td>
<td>-.09</td>
</tr>
<tr>
<td>5. Third-Party Affect</td>
<td>10.63</td>
<td>0.60</td>
<td>10.67</td>
<td>0.70</td>
<td>-.46**</td>
<td>-03</td>
<td>.11</td>
<td>.20</td>
<td>--</td>
</tr>
</tbody>
</table>

Note. Service failure correlations are above the diagonal, and service recovery correlations are below. Emo. = emotionality. Duration refers to the length of the performance episode (i.e., confederate call) in seconds. * $p < .05$; ** $p < .01$. 

98
Table 14. Means, standard deviations, and correlations for each phase of the performance episode.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. PI: Duration (Seconds)</td>
<td>98.62</td>
<td>24.89</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. PI: Participant Emo.</td>
<td>12.66</td>
<td>2.97</td>
<td>-.23*</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. PI: Surface Acting</td>
<td>4.87</td>
<td>4.09</td>
<td>-.05</td>
<td>-.26*</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. PI: Deep Acting</td>
<td>5.21</td>
<td>4.13</td>
<td>.05</td>
<td>-.10</td>
<td>.76**</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. PI: Third-Party Affect</td>
<td>11.26</td>
<td>0.57</td>
<td>-.29*</td>
<td>.22</td>
<td>.02</td>
<td>-.03</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. PII: Duration (Seconds)</td>
<td>113.91</td>
<td>26.94</td>
<td>.28*</td>
<td>-.12</td>
<td>-.07</td>
<td>.02</td>
<td>-.26*</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>7. PII: Participant Emo.</td>
<td>9.30</td>
<td>3.65</td>
<td>-.06</td>
<td>.55**</td>
<td>-.03</td>
<td>.03</td>
<td>-.02</td>
<td>-.02</td>
<td>--</td>
</tr>
<tr>
<td>8. PII: Surface Acting</td>
<td>10.48</td>
<td>4.74</td>
<td>-.12</td>
<td>-.01</td>
<td>.43**</td>
<td>.40**</td>
<td>.11</td>
<td>-.18</td>
<td>-.41**</td>
</tr>
<tr>
<td>9. PII: Deep Acting</td>
<td>9.94</td>
<td>4.84</td>
<td>-.04</td>
<td>-.03</td>
<td>.38**</td>
<td>.49**</td>
<td>.16</td>
<td>-.08</td>
<td>-.41**</td>
</tr>
<tr>
<td>10. PII: Third-Party Affect</td>
<td>10.31</td>
<td>0.86</td>
<td>-.21</td>
<td>.23*</td>
<td>.06</td>
<td>.01</td>
<td>.71**</td>
<td>-.20</td>
<td>.10</td>
</tr>
<tr>
<td>11. PIII: Duration (Seconds)</td>
<td>52.25</td>
<td>20.71</td>
<td>-.16</td>
<td>.09</td>
<td>.09</td>
<td>-.12</td>
<td>.21</td>
<td>-.27*</td>
<td></td>
</tr>
<tr>
<td>12. PIII: Participant Emo.</td>
<td>10.69</td>
<td>4.97</td>
<td>.01</td>
<td>.37**</td>
<td>-.02</td>
<td>-.04</td>
<td>.03</td>
<td>.02</td>
<td>.63**</td>
</tr>
<tr>
<td>13. PIII: Surface Acting</td>
<td>10.09</td>
<td>6.04</td>
<td>-.07</td>
<td>-.02</td>
<td>.23*</td>
<td>.30**</td>
<td>.13</td>
<td>-.07</td>
<td>-.37**</td>
</tr>
<tr>
<td>14. PIII: Deep Acting</td>
<td>9.28</td>
<td>5.85</td>
<td>-.10</td>
<td>.01</td>
<td>.23*</td>
<td>.44**</td>
<td>.14</td>
<td>-.02</td>
<td>-.31**</td>
</tr>
<tr>
<td>15. PIII: Third-Party Affect</td>
<td>10.24</td>
<td>0.86</td>
<td>-.22</td>
<td>.18</td>
<td>-.04</td>
<td>-.15</td>
<td>.41**</td>
<td>-.09</td>
<td>.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. PII: Surface Acting</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. PII: Deep Acting</td>
<td>.76**</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. PII: Third-Party Affect</td>
<td>.06</td>
<td>.12</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. PIII: Duration (Seconds)</td>
<td>.14</td>
<td>.12</td>
<td>-.29*</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. PIII: Participant Emo.</td>
<td>-.22</td>
<td>-.17</td>
<td>.05</td>
<td>.07</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. PIII: Surface Acting</td>
<td>.62**</td>
<td>.45**</td>
<td>.09</td>
<td>.03</td>
<td>-.62**</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. PIII: Deep Acting</td>
<td>.50**</td>
<td>.58**</td>
<td>.15</td>
<td>-.15</td>
<td>-.62**</td>
<td>.77**</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>15. PIII: Third-Party Affect</td>
<td>-.04</td>
<td>.04</td>
<td>.67**</td>
<td>-.02</td>
<td>.38**</td>
<td>-.32**</td>
<td>-.28*</td>
<td>--</td>
</tr>
</tbody>
</table>

*Note.* Emo. = emotionality. Duration refers to the length of the performance episode (i.e., confederate call) in seconds for each of the three phases. Correlations in bold and italics reference surface acting across the three phases and deep acting across the three phases. *p < .05; **p < .01.
Table 15. Means, standard deviations, and correlations for each phase of the performance episode split by condition.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. PI: Duration</td>
<td>99.08</td>
<td>27.79</td>
<td>98.16</td>
<td>21.98</td>
<td>--</td>
<td>-37**</td>
<td>-14</td>
<td>.02</td>
<td>-30</td>
<td>.30</td>
</tr>
<tr>
<td>2. PI: Part. Emo.</td>
<td>12.66</td>
<td>3.36</td>
<td>12.66</td>
<td>2.57</td>
<td>-14</td>
<td>--</td>
<td>.18</td>
<td>.20</td>
<td>.54**</td>
<td>-12</td>
</tr>
<tr>
<td>3. PI: SA</td>
<td>4.66</td>
<td>4.25</td>
<td>5.07</td>
<td>3.97</td>
<td>.02</td>
<td>.33*</td>
<td>--</td>
<td>.75**</td>
<td>-.07</td>
<td>-.23</td>
</tr>
<tr>
<td>4. PI: DA</td>
<td>5.24</td>
<td>4.83</td>
<td>5.18</td>
<td>3.36</td>
<td>.07</td>
<td>.05</td>
<td>.78**</td>
<td>--</td>
<td>-.16</td>
<td>-.04</td>
</tr>
<tr>
<td>5. PI: T-P Affect</td>
<td>11.33</td>
<td>0.54</td>
<td>11.19</td>
<td>0.60</td>
<td>-.30</td>
<td>-.05</td>
<td>.13</td>
<td>.07</td>
<td>--</td>
<td>-.22</td>
</tr>
<tr>
<td>6. PI: Duration</td>
<td>111.50</td>
<td>28.40</td>
<td>116.32</td>
<td>25.55</td>
<td>.27</td>
<td>.27</td>
<td>.05</td>
<td>.06</td>
<td>-.29</td>
<td>--</td>
</tr>
<tr>
<td>7. PI: Part. Emo.</td>
<td>9.21</td>
<td>3.68</td>
<td>9.40</td>
<td>3.67</td>
<td>.12</td>
<td>.61**</td>
<td>-.18</td>
<td>.01</td>
<td>-.14</td>
<td>.08</td>
</tr>
<tr>
<td>8. PI: SA</td>
<td>10.42</td>
<td>4.77</td>
<td>10.54</td>
<td>4.77</td>
<td>-.09</td>
<td>.02</td>
<td>.49**</td>
<td>.43**</td>
<td>.16</td>
<td>-.24</td>
</tr>
<tr>
<td>9. PI: DA</td>
<td>9.13</td>
<td>4.68</td>
<td>10.75</td>
<td>4.91</td>
<td>-.15</td>
<td>.04</td>
<td>.58**</td>
<td>.64**</td>
<td>-.28</td>
<td>-.18</td>
</tr>
<tr>
<td>10. PI: T-P Affect</td>
<td>10.35</td>
<td>0.82</td>
<td>10.27</td>
<td>0.90</td>
<td>-.30</td>
<td>.01</td>
<td>.02</td>
<td>.02</td>
<td>.62**</td>
<td>-.29</td>
</tr>
<tr>
<td>11. PII: Duration</td>
<td>44.82</td>
<td>19.01</td>
<td>59.68</td>
<td>19.87</td>
<td>-.02</td>
<td>-.19</td>
<td>.22</td>
<td>.16</td>
<td>-.01</td>
<td>.27</td>
</tr>
<tr>
<td>12. PII: Part. Emo.</td>
<td>8.02</td>
<td>4.56</td>
<td>13.35</td>
<td>3.84</td>
<td>.38</td>
<td>.28</td>
<td>-.09</td>
<td>-.01</td>
<td>-.16</td>
<td>.15</td>
</tr>
<tr>
<td>13. PII: SA</td>
<td>13.27</td>
<td>5.43</td>
<td>6.92</td>
<td>4.87</td>
<td>-.12</td>
<td>.17</td>
<td>.18</td>
<td>.28</td>
<td>.16</td>
<td>-.09</td>
</tr>
<tr>
<td>14. PII: DA</td>
<td>11.92</td>
<td>5.94</td>
<td>6.65</td>
<td>4.46</td>
<td>-.22</td>
<td>.13</td>
<td>.24</td>
<td>.40*</td>
<td>.23</td>
<td>-.09</td>
</tr>
<tr>
<td>15. PII: T-P Affect</td>
<td>9.78</td>
<td>0.70</td>
<td>10.70</td>
<td>0.76</td>
<td>-.25</td>
<td>.02</td>
<td>-.17</td>
<td>-.17</td>
<td>.35</td>
<td>-.24</td>
</tr>
</tbody>
</table>

Note. Correlations for the service failure condition are below the diagonal; correlations for the service recovery condition are above. Duration refers to the length of the performance episode (i.e., confederate call) in seconds for each of the three phases. Correlations in bold and italics reference surface acting across the three phases and deep acting across the three phases. Part. Emo. = participant emotionality. SA = surface acting. DA = deep acting. T-P Affect = third-party affect ratings. * p < .05; ** p < .01.

When looking at the entire performance episode separated by study conditions (Table 13), a significant, positive correlation between surface and deep acting was again found for both the service failure ($r = .73$, $p < .01$) and service recovery ($r = .75$, $p < .01$) conditions. This relationship was replicated when looking at the three phases of the call.
across all participants (Phase I: $r = .76, p < .01$; Phase II: $r = .76, p < .01$; Phase III: $r = .77, p < .01$; see Table 14) as well as when looking at the three phases of the call split across service failure (Phase I: $r = .78, p < .01$; Phase II: $r = .77, p < .01$; Phase II: $r = .72, p < .01$; see Table 15) and service recovery (Phase I: $r = .75, p < .01$; Phase II: $r = .76, p < .01$; Phase III: $r = .70, p < .01$; see Table 15) conditions. Together, these results continue to support Hypothesis 2, fitting with previous literature demonstrating a positive relationship between surface and deep acting (Hülsheger & Schewe, 2011). However, it is interesting that these correlations among average levels of surface acting and deep acting are much larger than the momentary cross-correlations between these variables, suggesting that within a given moment, the relationship is much weaker than at a broader, more global level of assessment.

For more insight, additional analyses were conducted to determine whether the aggregate correlations were significantly different from the cross-correlations calculated in testing Hypothesis 1. Given that the dependent nature of the correlations did not match the formula needed for a dependent samples correlation (i.e., correlations of $XY$, $XZ$, and $YZ$ are needed, Bobko, 2001; the current study only had $XY$ and $XY$), I utilized the t-test formula for testing a correlation against a pre-specified correlation value. For instance, when testing the correlation between aggregated surface acting and deep acting ($r = .76$) against the cross-correlation obtained across the entire call for surface acting and deep acting ($r = .64$), the following formula was utilized:

$$
\frac{r_{correlation} - r_{comparison}}{\sqrt{1 - r^{2}_{correlation}}} \rightarrow \frac{.76 - .64}{\sqrt{1 - .76^{2}}} \Rightarrow \sqrt{\frac{.76 - .64}{1 - .76^{2}}} \Rightarrow \sqrt{\frac{.76 - .64}{76 - 2}}
$$
This resulted in a $t$-statistic of 1.59; with degrees of freedom of 74, this value was not significant, suggesting that the overall cross-correlation between surface and deep acting is comparable to the aggregated correlation between the constructs.

However, the correlations and cross-correlations start to differ once the performance episode is separated into the three phases of the call. The correlation for average levels of surface and deep acting across all participants (see Table 14) was significantly larger than the cross-correlation found between surface and deep acting (see Table 11) during Phase I ($t_{(74)} = 6.62, p < .001$), Phase II ($t_{(74)} = 4.63, p < .001$) and Phase III ($t_{(74)} = 5.80, p < .001$). These differences continued to occur when splitting the three phases of the call by condition. For the service failure condition, the correlation for average levels of surface and deep acting (see Table 15) was significantly larger than the cross-correlation of surface and deep acting (see Table 12) during Phase I ($t_{(36)} = 6.13, p < .001$), Phase II ($t_{(36)} = 3.57, p < .001$), and Phase III ($t_{(36)} = 4.41, p < .001$). Similarly, for the service recovery condition, the correlations of aggregate surface and deep acting (see Table 15) were significantly larger than the cross-correlations for surface and deep acting (see Table 12) during Phase I ($t_{(36)} = 3.54, p < .001$), Phase II ($t_{(36)} = 3.05, p < .001$), and Phase III ($t_{(36)} = 2.44, p < .01$). Taken together, these results present an interesting theoretical contribution in that the overall correlations of average levels of surface and deep acting continuous ratings were much larger than the cross-correlations which captured the extent to which surface and deep acting covaried with each other in the moment.

As a final test to understand the nature of the average continuous ratings, I correlated the average continuous ratings of surface acting and deep acting with post-
simulation “in general” surface acting and deep acting collected using standard measures in the literature (see Appendix J for measure information). These measures were taken after the simulation, after all of the continuous ratings had been made. For brevity, Table 16 summarizes the results by study condition and phase of performance episode.

Table 16. Correlations between average continuous surface and deep acting and post-simulation surface and deep acting.

<table>
<thead>
<tr>
<th></th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SR</td>
<td>SF</td>
<td>SR</td>
</tr>
<tr>
<td>Avg. SA / P-S SA</td>
<td>.27</td>
<td>.11</td>
<td>.52**</td>
</tr>
<tr>
<td>Avg. SA / P-S DA</td>
<td>-.32*</td>
<td>.31</td>
<td>-.15</td>
</tr>
<tr>
<td>Avg. DA / P-S SA</td>
<td>.09</td>
<td>.26</td>
<td>.25</td>
</tr>
<tr>
<td>Avg. DA / P-S DA</td>
<td>-.17</td>
<td>-.03</td>
<td>-.05</td>
</tr>
</tbody>
</table>

Note. All values are correlation coefficients (r). SA = surface acting. DA = deep acting. SR = service recovery. SF = service failure. Correlation coefficients for the service failure and service recovery conditions were tested with df = 36. * p < .05; ** p < .01.

Largely, average ratings of continuous surface and deep acting were unrelated to post-simulation deep acting, except for negative relationships (as one would expect) between average levels of surface acting and post-simulation deep acting for the service recovery condition across Phases I (r = -.32, p < .05) and III (r = -.34, p < .05). Also as expected, positive relationships emerged between average continuous surface acting and post-simulation surface acting for the service failure and recovery conditions during Phases II (r = .45, p < .01 and r = .52, p < .01, respectively) and III (r = .45, p < .01 and r = .39, p < .05, respectively). Surprisingly, for the service failure condition, average levels of deep acting were positively correlated with post-simulation surface acting during Phases II (r = .42, p < .01) and III (r = .39, p < .05).
Combined, these results demonstrate the potential limitations of utilizing single-time assessments of surface and deep acting collected at the end of a performance episode. That is, when recalling how one regulated across an entire call, people may rely more on the most recent feelings and experiences than what was experienced throughout the entire call. The amount of strain felt at the end of a performance episode may have a large effect on the ratings of emotion regulation for the entire episode. Thus, the relatively low levels of overlap between the continuous ratings and the end-of-call ratings suggest that each provides unique information about the emotional labor experience.

Hypotheses 3 and 4

Both Hypotheses 3 and 4 dealt with within-subject comparisons of surface acting and deep acting across the three separate phases of the call. Hypothesis 3 stated that (a) surface acting and (b) deep acting would increase from Phase I (i.e., caller being neutral) to Phase II (i.e., caller becoming negative) of the call. Hypothesis 4 further predicted that the levels of (a) surface acting and (b) deep acting would be higher during Phase III of the service encounter for those in the service failure condition (i.e., those exposed to increased negativity from the caller) compared to those in the service recovery condition (i.e., those exposed to increased positivity from the caller). Though no formal hypotheses were made, it was proposed that the levels of surface acting and deep acting would remain the same from Phase II to Phase III for those in the service recovery condition based upon the control theory idea that the discrepancy would not increase and therefore no additional resources would be allocated to managing the situation (Carver & Scheier, 1998; Diefendorff & Gosserand, 2003). A repeated measures ANOVA with a between-subjects factor (3 x 2 design) was used to test Hypotheses 3 and 4, with separate models
run for surface and deep acting. Means for surface and deep acting, as well as participant emotionality, across each phase split by condition are presented in Table 17.

Table 17. Means and standard deviations for participant emotionality, surface acting, and deep acting across performance episode by study condition.

<table>
<thead>
<tr>
<th></th>
<th>(A) Phase I</th>
<th></th>
<th>(B) Phase II</th>
<th></th>
<th>(C) Phase III</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Participant Emotionality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service Failure</td>
<td>12.66 $^B$</td>
<td>3.36</td>
<td>9.21 $^A_C$</td>
<td>3.68</td>
<td>8.02 $^A_B$</td>
<td>4.56</td>
</tr>
<tr>
<td>Surface Acting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service Failure</td>
<td>4.66 $^B_C$</td>
<td>4.25</td>
<td>10.42 $^A_C$</td>
<td>4.77</td>
<td>13.27 $^A_B$</td>
<td>5.43</td>
</tr>
<tr>
<td>Service Recovery</td>
<td>5.07 $^B$</td>
<td>3.97</td>
<td>10.54 $^A_C$</td>
<td>4.77</td>
<td>6.92 $^B$</td>
<td>4.87</td>
</tr>
<tr>
<td>Deep Acting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service Failure</td>
<td>5.24 $^B_C$</td>
<td>4.83</td>
<td>9.13 $^A_C$</td>
<td>4.68</td>
<td>11.92 $^A_B$</td>
<td>5.94</td>
</tr>
<tr>
<td>Service Recovery</td>
<td>5.18 $^B_C$</td>
<td>3.36</td>
<td>10.75 $^A_C$</td>
<td>4.91</td>
<td>6.65 $^A_B$</td>
<td>4.46</td>
</tr>
</tbody>
</table>

Note. $M$ = mean. $SD$ = standard deviation. Means presented in this table are the raw means for each phase within each condition. Values with superscripts indicate that the mean in a particular phase for surface acting/deep acting within the condition was significantly different from another phase.

Prior to interpreting results for Hypotheses 3 and 4 for surface and deep acting, I conducted a repeated measures ANOVA to determine how participant emotionality varied as a function of condition and phase of call. In running these analyses, I noted that the data violated the sphericity assumption for repeated measures analysis of variance (Mauchly’s $W = .85, \chi^2_{(2)} = 11.66, p < .01$). Sphericity occurs when the variances of the differences between repeated measures of a variable are equal (i.e., across phases of the call), which is assumed to be true when performing a repeated measures ANOVA.
Because this assumption was violated, one must make a choice about which correction is more appropriate for the data. According to recommendations from Girden (1992), if the Epsilon statistics are greater than .75, the Huynh-Feldt correction should be used; if they are less than .75, the Greenhouse-Geisser statistic should be used. In the current analyses, the Epsilons were greater than .75 (Greenhouse-Geisser: .87; Huynh-Feldt: .90). As such, the Huynh-Feldt correction was used to account for the unequal variances.

For participant emotionality, results indicated a significant interaction between phase of call and the between-subject factor, or study condition \((F_{(1.805, 133.592)} = 192.70, p < .001, \eta_p^2 = .29)\). To interpret the nature of this interaction, I plotted the means by phase of call and condition in Figure 16 (means and standard deviations are in Table 17).

Figure 16. Interaction between phase of confederate call and study condition predicting changes in participant emotionality.
To further interpret the effect, I ran separate repeated measures ANOVAs for each condition. For the service recovery condition, I found a significant effect for phase of call on participant emotionality ($F(2, 74) = 34.60, p < .001; \eta^2_p = .48$), such that the means for participant emotionality during Phases I ($M = 12.66, SD = 2.57$) and III ($M = 13.35, SD = 3.84$) were significantly higher (i.e., more positive) than the mean for Phase II ($M = 9.40, SD = 3.67$). Phases I and III were not significantly different from each other. Thus, as the confederate caller became angry (Phase II), emotionality declined; as the confederate became more positive (Phase III), participant emotionality improved.

For the service failure condition, there was also a significant effect for phase of call on participant emotionality ($F(1.222, 45.228) = 33.50, p < .001, \eta^2_p = .48$; corrected with the Greenhouse-Geisser statistic). As shown in Figure 16 (means in Table 17), participant emotionality during Phase I ($M = 12.66, SD = 3.36$) was higher (i.e., more positive) than Phases II ($M = 9.21, SD = 3.68$) and III ($M = 8.02, SD = 4.56$). Phases II and III were also significantly different from each other. Thus, as confederate callers became increasingly angry in Phases II and III, participant emotionality declined across the three phases.

Finally, I conducted a series of independent samples t-tests to determine if mean levels of participant emotionality at Phases I, II, and III differed by condition (see Table 17 for means). Results indicated that there was no significant difference between conditions for Phases I ($t(74) = -.01, n.s.$) and II ($t(74) = .22, n.s.$). However, there was a significant difference for Phase III ($t(74) = 5.51, p < .001$), with mean levels of participant emotionality being lower in the service failure condition than in the service recovery condition. In sum, these initial analyses demonstrate that participants felt more negatively when confederate callers were being hostile towards them, and that participants in the
service failure condition were significantly more negative during Phase III of the call than participants who were in the service recovery condition.

In testing Hypotheses 3 and 4 for surface acting, I again utilized the Huynh-Feldt correction due to a sphericity violation (Mauchly’s $W = .76$, $\chi^2_{(2)} = 19.79, p < .001$; Greenhouse-Geisser: .81; Huynh-Feldt: .84). Results indicated a significant interaction between phase of call and the between-subject factor on the level of surface acting $(F_{(1.669, 123.527)} = 25.40, p < .001, \eta^2_p = .26)$. To interpret this effect, I plotted the means for the interaction (see Figure 17; means and standard deviations are presented previously in Table 17) and conducted repeated measures ANOVAs within each condition to determine if the phases were significantly different from each other.

![Figure 17. Interaction between phase of confederate call and study condition predicting changes in surface acting (SA).](#)
For the service failure condition (corrected with the Greenhouse-Geisser statistic) as predicted in Hypothesis 3, there was a significant effect of phase of performance episode on surface acting \( (F_{1.33, 49.06} = 62.30, p < .001; \eta^2_p = .63) \), with pairwise comparisons (see Table 17 for means) revealing that mean levels of surface acting at Phase II \( (M = 10.42, SD = 4.77) \) were significantly higher than mean levels during Phase I \( (M = 4.66, SD = 4.25) \). Further, mean levels of surface acting continued to increase as the call became more hostile, with mean levels of surface acting being significantly higher in Phase III \( (M = 13.27, SD = 5.43) \) than during Phases I and II.

For the service recovery condition, I also found a significant effect for phase of call on surface acting \( (F_{2, 74} = 29.01, p < .001; \eta^2_p = .44) \). However, pairwise comparisons revealed a slightly different story than the one found for the service failure condition. Specifically, Phase II \( (M = 10.54, SD = 4.77) \) was significantly higher than Phases I \( (M = 5.07, SD = 3.97) \) and III \( (M = 6.92, SD = 4.87) \). However, Phases I and III were not significantly different from each other, suggesting that as the social context becomes more pleasant, less emotion regulation may be needed.

Though not hypothesized, to further interpret the results, I conducted a series of independent samples t-tests to determine if mean levels of surface acting at Phases I, II, and III differed by condition (see Table 17 for means and standard deviations). Results indicated that there was no significant difference between conditions for Phases I \( (t_{74} = .44, n.s.) \) and II \( (t_{74} = .11, n.s.) \). However, there was a significant difference for Phase III \( (t_{74} = -5.37, p < .001) \), with mean levels of surface acting being higher in the service failure condition than in the service recovery condition, further supporting Hypothesis 4.
To further explore the effects of condition (failure vs. recovery) and phase on surface acting, I examined the proportion of time participants spent engaging in different levels of the strategy. I separately examined the extent to which participants engaged in (a) no surface acting (i.e., ratings of 1), (b) low levels of surface acting (i.e., scores greater than 1 and up to 4 [very small extent]), mid-levels of surface acting (i.e., scores between 4.20 and 16.80) and (c) high levels of surface acting (i.e., scores of 17 [large extent] and above). These values were chosen based on the anchors used for the continuous ratings scales. The proportions of time for surface acting separated by phase of performance episode and study condition are presented in Table 18.

Table 18. Proportions of time spent utilizing surface acting within performance episode by study condition.

<table>
<thead>
<tr>
<th></th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Service Failure</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Surface Acting</td>
<td>0.39</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td>Low Surface Acting</td>
<td>0.29</td>
<td>0.14</td>
<td>0.08</td>
</tr>
<tr>
<td>Mid-Level Surface Acting</td>
<td>0.29</td>
<td>0.61</td>
<td>0.52</td>
</tr>
<tr>
<td>High Surface Acting</td>
<td>0.03</td>
<td>0.19</td>
<td>0.37</td>
</tr>
<tr>
<td><strong>Service Recovery</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Surface Acting</td>
<td>0.34</td>
<td>0.10</td>
<td>0.28</td>
</tr>
<tr>
<td>Low Surface Acting</td>
<td>0.24</td>
<td>0.12</td>
<td>0.19</td>
</tr>
<tr>
<td>Mid-Level Surface Acting</td>
<td>0.40</td>
<td>0.59</td>
<td>0.50</td>
</tr>
<tr>
<td>High Surface Acting</td>
<td>0.01</td>
<td>0.19</td>
<td>0.03</td>
</tr>
</tbody>
</table>

*Note.* Proportions are the average across participants within a condition (service failure or service recovery).

Fitting with the repeated measures analysis of variance, both conditions reported using relatively little to no levels of surface acting during Phase I of the call. Specifically, for the service failure condition, the proportion of time spent engaging in no surface
acting was .39, which is similar to the proportion of time individuals reported no surface acting during the service recovery condition (.34; \( t(74) = -.48, n.s. \)). Both conditions also reported utilizing low levels of surface acting an equivalent amount (service failure: .29; service recovery: .24; \( t(74) = -.77, n.s. \)). For Phase II, where the call became more difficult for both conditions, participants in both conditions reported using mid-levels of surface acting the most (service failure: .61; service recovery: .58; \( t(74) = -.38, n.s. \)), followed by engaging in high levels of surface acting (service failure: .19; service recovery: .19; \( t(74) = .11, n.s. \)). These results suggest that at the onset of negative emotionality from the customers, participants did not automatically use the maximum amount of surface acting; rather, participants more gradually increased their use of surface acting.

Finally, for Phase III, the greatest divergence in utilizing surface acting is found. In the service failure condition, participants spent a greater proportion of time utilizing high levels of surface acting (.37) in comparison to participants in the service recovery condition (.03; \( t(74) = -4.49, p < .001 \)). Conversely, in the service recovery condition, participants spent a greater amount of time utilizing no surface acting (.27) than their counterparts in the service failure condition (.03; \( t(74) = 3.78, p < .001 \)). These results continue to support Hypotheses 3 and 4 for surface acting: across both conditions, higher levels of regulation were needed in Phase II compared to Phase I (supporting Hypothesis 3). However, this increase only continued into Phase III for the service failure condition (supporting Hypothesis 4); those in the service recovery condition actually had a high proportion of participants not surface act at all.

In testing Hypotheses 3 and 4 for deep acting, the same Huynh-Fundt correction was used given that sphericity was again violated (Mauchly’s \( W = .89, \chi^2(2) = 8.945, p < \))
and the corresponding epsilon values were greater than .75 (Greenhouse-Geisser = .90; Huynh-Feldt: .93). Means and standard deviations were presented previously in Table 17. A significant interaction occurred between phase of call and study condition in predicting deep acting \( (F_{(1.860, 137.618)} = 25.92, p < .001, \eta_p^2 = .26) \). To interpret the nature of the interaction, I again ran two repeated measures ANOVAs for each study condition. The interaction is presented in Figure 18.

![Interaction between phase of confederate call and study condition predicting changes in deep acting (DA).](image)

For the service failure condition, a significant effect of phase of call was found \( (F_{(2,74)} = 41.79, p < .001, \eta_p^2 = .53) \), with pairwise comparisons revealing that mean levels of deep acting across all three phase of the call were significantly different from each
other (see Table 17). In line with Hypothesis 3, mean levels of deep acting during Phase II ($M = 9.13, SD = 4.68$) of the call were significantly higher than mean levels at Phase I ($M = 5.24, SD = 4.83$). Phase III of the call ($M = 11.92, SD = 5.94$), however, revealed the highest mean levels of deep acting, with Phase III being significantly higher than Phases I and II. Thus, as the confederate caller became increasingly hostile, participants felt as though they had to put forth more effort to try and feel the emotions they were required to express.

For the service recovery condition (which was corrected with the Huynh-Feldt correction), I also found a significant effect for phase ($F(1.758, 9.830) = 36.37, p < .001, \eta^2_p = .50$), with all three phases being significantly different from each other (see Table 17). Supporting Hypothesis 3, significantly higher mean levels of deep acting were used during Phase II ($M = 10.75, SD = 4.91$) than Phase I ($M = 5.18, SD = 3.36$). Phase II was also higher than the mean levels of deep acting during Phase III ($M = 6.65, SD = 4.46$), suggesting that as callers became less hostile, lower levels of deep acting were needed. Finally, although Phase III levels of deep acting declined close to where they started in Phase I, Phase III was significantly higher. Combined with the results from the service failure condition, I supported Hypothesis 4, given that mean levels of deep acting increased in the service failure condition during Phase III only.

Further, I conducted a series of independent samples t-tests to determine if mean levels of deep acting during each Phase varied by condition. Results mirrored those found for surface acting, with no differences in mean levels of deep acting being found for Phase I ($t(74) = -0.06, n.s.$) or Phase II ($t(74) = 1.47, n.s.$). A significant difference was found for Phase III ($t(74) = -4.38, p < .001$), with mean levels of deep acting being higher for
those in the service failure condition than those in the service recovery condition. This result further supports Hypothesis 4.

Similar to the supplemental analyses conducted for surface acting, I also conducted analyses to determine the amount of time participants engaged in (a) no deep acting (i.e., ratings of 1), (b) low levels of deep acting (i.e., scores from 1.20 [next value found above 1 indicating some level of regulation] – 4 [very small extent] and below), (c) mid-levels of deep acting (i.e., scores between 4.20-16.80), and (d) high levels of deep acting (i.e., scores of 17 [large extent] and above). These were split by conduction and phase of the performance episode. Proportions, and the corresponding amount of time in seconds for each phase, are presented in Table 19.

Table 19. Proportions of time spent utilizing deep acting within performance episode by study condition.

<table>
<thead>
<tr>
<th></th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Service Failure</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Deep Acting</td>
<td>.39</td>
<td>.13</td>
<td>.10</td>
</tr>
<tr>
<td>Low Deep Acting</td>
<td>.22</td>
<td>.11</td>
<td>.06</td>
</tr>
<tr>
<td>Mid-Level Deep Acting</td>
<td>.34</td>
<td>.64</td>
<td>.57</td>
</tr>
<tr>
<td>High Deep Acting</td>
<td>.05</td>
<td>.12</td>
<td>.28</td>
</tr>
<tr>
<td><strong>Service Recovery</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Deep Acting</td>
<td>.29</td>
<td>.07</td>
<td>.26</td>
</tr>
<tr>
<td>Low Deep Acting</td>
<td>.27</td>
<td>.13</td>
<td>.18</td>
</tr>
<tr>
<td>Mid-Level Deep Acting</td>
<td>.43</td>
<td>.58</td>
<td>.51</td>
</tr>
<tr>
<td>High Deep Acting</td>
<td>.01</td>
<td>.21</td>
<td>.05</td>
</tr>
</tbody>
</table>

*Note.* Proportions are the average across participants within a condition (service failure or service recovery).

Similar to the results for surface acting, for deep acting, both the service failure and service recovery conditions spent the time utilizing no deep acting during Phase I of
the call (service failure: .39; service recovery: .29; t(74) = -1.16, n.s.) in addition to very
low levels of surface acting (service failure: .22; service recovery: .28; t(74) = .69, n.s.).
Thus, when customers were not being difficult, little to no deep acting was needed. For
Phase II of the call, however, both conditions reported utilizing high levels of deep acting
(service failure: .12; service recovery: .21; t(74) = 1.47, n.s.) in addition to low levels of
deep acting (service failure: .11; service recovery: .13; t(74) = .46, n.s.) suggesting again
that at the onset of customer incivility, the highest amounts of emotional labor may not be
required. This is further supplemented by the proportion of time spent utilizing mid-
levels of deep acting during Phase II for both conditions (service failure: .64; service
recovery: .58; t(74) = .75, n.s.). Finally, for Phase III of the call, those in the service
failure condition spent significantly more time utilizing high levels of deep acting (.28)
than those in the service recovery condition (.05; t(74) = -3.14, p < .01), whereas those in
the service recovery condition spent significantly more time utilizing no deep acting (.26;
service failure: .10; t(74) = 2.21, p < .05) or low levels of deep acting (.18; service failure:
.06; t(74) = 2.24, p < .05). Thus, as customers became more pleasant, the amount of time
spent engaging in deep acting reduced. For those who experienced increased customer
hostility (i.e., negative social feedback), however, utilizing moderate to high levels of
deep acting was needed.

Overall, results were fully supportive of Hypotheses 3 and 4 and are consistent
with ideas presented by Côté (2005) regarding the role of social information in emotion
regulation. In line with Hypothesis 3, as customers (i.e., confederate callers) became
increasingly hostile (i.e., providing negative social information from Phase I  Phase II
across both conditions and Phase II  Phase III in the service failure condition), both
types of regulation strategies were used in increasing amounts. Thus, participants were responsive to the negative social feedback and attempted to ‘resolve’ the situation by increasing their use of surface and deep acting. Moreover, in support of Hypothesis 4, the level of emotion regulation only increased for those who experienced higher levels of hostility (i.e., service failure); for participants in the service recovery condition who experienced positive social information, levels of emotion regulation actually decreased. This counters the idea that positive feedback would cause individuals to maintain their regulation at the same level (because participant negative affect would still be present and require regulation). Indeed, it seems that the positive displays of customers resulted in a decrease in regulation to levels lower than the previous state, and closer to the level occurring at the outset of the interaction. This corresponded to a fast decline in negative emotions in response to the civil treatment by the confederate caller.

These results shed light for the first time on how emotion regulation strategies vary across a performance episode depending upon the level of social information being provided by the customer, suggesting that emotion regulation may not always be needed. That is, at the beginning of an interaction, although emotional demands may be present (i.e., employees know that ‘service with a smile’ is required), the situation may not be ‘demanding’ since the interaction has just begun. Thus, even though it is beneficial and important for employees to be friendly with customers at the beginning of an interaction, it may not be challenging to do so. Moreover, for employees who have negative interactions that end on a positive note (i.e., the customer becomes more pleasant; service recovery condition), rather than maintaining high levels of regulation with a customer, regulation may decrease since the customer seems satisfied with the encounter. This
could be due to employees ‘relaxing’ their levels of regulation, or because employees feel as though they can express naturally felt emotions.

To further explore this idea, I considered what occurred in participants when they reported not utilizing surface acting or deep acting. That is, what is the emotionality of participants when they do not feel the need to regulate their emotions using either strategy? Results are presented in Table 20.

Table 20. Participant emotionality ratings when neither surface nor deep acting are utilized.

<table>
<thead>
<tr>
<th>Service Failure</th>
<th>Phase I Emotionality</th>
<th>Phase II Emotionality</th>
<th>Phase III Emotionality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>No SA &amp; DA</td>
<td>12.68</td>
<td>3.18</td>
<td>10.00</td>
</tr>
<tr>
<td>SA &amp; DA Reported</td>
<td>12.48</td>
<td>4.06</td>
<td>9.25</td>
</tr>
<tr>
<td></td>
<td>SA &amp; DA Reported</td>
<td>12.15</td>
<td>3.33</td>
</tr>
</tbody>
</table>

Note. SA = surface acting. DA = deep acting. No SA & DA was indicated by ratings of ‘1’ on the 1-20 continuous rating scale. SA &DA reported indicates that participants reported using any level of surface acting or deep acting (i.e., ratings greater than 1 the 1-20 continuous scale). Analyses were conducted using the 1-second data.

Fitting with calls from Grandey et al. (2013), the results from Table 20 show for the first time what is occurring when employees are not utilizing any form of emotion regulation. For the service failure condition, participants rated their felt emotions as more positive when reporting no emotion regulation in comparison to utilizing some form of emotion regulation during Phases II (no regulation: $M = 10.00, SD = 2.08$; regulation: $M = 9.25, SD = 4.17$; $t_{(4234)} = -2.46, p < .05$) and III (no regulation: $M = 10.71, SD = 0.58$; regulation: $M = 7.85, SD = 4.70$; $t_{(1701)} = -2.28, p < .05$), but not during Phase I (no
regulation: $M = 12.68, SD = 3.18$; regulation: $M = 12.48, SD = 4.06$; $t_{(3763)} = -1.35, n.s.$).

For the service recovery condition, participants also rated their felt emotions as more positive when reporting no emotion regulation in comparison to utilizing some level of emotion regulation during Phases I (no regulation: $M = 13.39, SD = 2.99$; regulation: $M = 12.15, SD = 3.33$; $t_{(3728)} = -10.01, p < .001$), II (no regulation: $M = 14.31, SD = 3.20$; regulation: $M = 9.01, SD = 4.40$; $t_{(4417)} = -17.81, p < .001$) and III (no regulation: $M = 16.28, SD = 3.04$; regulation: $M = 12.40, SD = 4.49$; $t_{(2266)} = -15.98, p < .001$). These results demonstrate that when participants reported not regulating their emotions, it was not because they were feeling negative emotions and were violating the display rules; rather, it was because they felt positively and could, perhaps, express their naturally felt emotions.

**Hypothesis 5**

Hypothesis 5 predicted that participants in the increased service failure condition would be more likely to switch between using surface acting and deep acting strategies in the third phase of the call compared to participants in the service recovery condition. This prediction was made based on the fact that the confederate caller in the service failure condition became increasingly hostile in the third phase and the expectation that participants may try out different strategies (regulation strategies and problem solving strategies) as a result (Côté, 2005); however, the confederate caller in the service recovery condition became more pleasant and positive, likely eliminating the need to continue to experiment with different regulation strategies.

At any given point, participants typically used one emotion regulation strategy at a higher level than the other. Switching was conceptualized as the act of moving from
using one strategy as the dominant approach to using the other strategy as the dominant approach. Switching was operationalized three ways: switching between regulation strategies in either direction, switching from surface acting to deep acting (i.e., the score on deep acting became higher than the score on surface acting), and switching from deep acting to surface acting (i.e., the score on surface acting became higher than the score on deep acting). A frequency count was used to indicate the number of times a participant switched between acting strategies (i.e., a score of ‘7’ indicates that a participant switched their dominant acting strategy seven times). To fully explore the amount of switching that occurred, Table 21 displays the results for switching for the entire call, Phase I, Phase II, and Phase III (the actual test specified by Hypothesis 5). Means for each type of switching in each of the four time frames are presented.
Table 21. Switching between surface acting and deep acting during the performance episode.

<table>
<thead>
<tr>
<th></th>
<th>Switching: Yes or No</th>
<th>Switching: DA → SA</th>
<th>Switching: SA → DA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td><strong>Entire Episode</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service Failure</td>
<td>2.39</td>
<td>(2.39)</td>
<td>1.16</td>
</tr>
<tr>
<td>Service Recovery</td>
<td>3.66</td>
<td>(2.75)</td>
<td>1.76</td>
</tr>
<tr>
<td><strong>Phase I</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service Failure</td>
<td>0.87</td>
<td>(1.18)</td>
<td>0.39</td>
</tr>
<tr>
<td>Service Recovery</td>
<td>1.50</td>
<td>(1.91)</td>
<td>0.71</td>
</tr>
<tr>
<td><strong>Phase II</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service Failure</td>
<td>1.24</td>
<td>(1.60)</td>
<td>0.61</td>
</tr>
<tr>
<td>Service Recovery</td>
<td>1.52</td>
<td>(1.46)</td>
<td>0.74</td>
</tr>
<tr>
<td><strong>Phase III</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service Failure</td>
<td>0.29</td>
<td>(0.61)</td>
<td>0.16</td>
</tr>
<tr>
<td>Service Recovery</td>
<td>0.63</td>
<td>(0.97)</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Note. *M* = mean. *SD* = standard deviation.

Given that Hypothesis 5 only focused on the final phase of the call between conditions, a series of independent samples t-tests were conducted for each of the three conceptualizations of switching. Results indicated that, contrary to Hypothesis 5, there were no significant differences between conditions in the three forms of switching. Specifically, a marginal difference was found for switching ‘in general’ (*t*(74) = 1.84, *p* = .07), with those in the service recovery condition (*M* = 0.63, *SD* = 0.29) engaging in slightly higher levels of switching than those in the service failure condition (*M* = 0.29, *SD* = 0.61). There was no significant difference for switching from deep acting to surface acting (*t*(74) = 1.52, *n.s.*), and another marginal effect was found for switching from surface acting to deep acting (*t*(74) = 1.70, *p* = .09), with the service recovery condition (*M* = 120
= 0.32, \(SD = 0.57\) having slightly higher levels of switching than those in the service failure condition (\(M = 0.13, SD = 0.34\)). In sum, initial analyses testing Hypothesis 5 were not supportive, with marginally significant effects being found in the opposite direction than anticipated.

To further analyze whether differences in switching emerged between conditions and across the three phases of the call, a 3 (phase of call) x 2 (condition) repeated measures ANOVA was conducted for each of the three operationalizations of switching. For switching ‘in general,’ the Huynh-Feldt correction was used. Results indicated that there was no interaction between phase of call and condition (\(F_{(1.876, 138.801)} = .386, n.s.\)). Rather, a main effect for phase of call emerged (\(F_{(1.876, 138.801)} = 10.71, p < .001, \eta^2_p = .13\)), showing that across the three phases for both conditions combined, the lowest amount of switching actually occurred in Phase III of the call (\(M = 0.46, SD = 0.82\)). Phase I (\(M = 1.18, SD = 1.61\)) was significantly higher than Phase III, as was Phase II (\(M = 1.38, SD = 1.53\)); Phases I and II were not significantly different from each other (\(t_{(75)} = -.82, n.s.\)).

There also was a marginally significant effect for condition (\(F_{(1, 74)} = 4.56, p = .058, \eta^2_p = .06\)), with pairwise comparisons showing that over the entire performance episode (i.e., across all three phases), participants in the service recovery condition engaged in higher levels of switching ‘in general’ (\(M = 1.22, S.E. = .14\)) than those in the service failure condition (\(M = 0.80, S.E. = .14, p < .05\)).

For switching from deep acting to surface acting, the Huynh-Feldt correction was again used. Similar to the findings for switching ‘in general,’ there was no significant interaction between phase of call and condition (\(F_{(1.752, 129.670)} = .38, n.s.\)); there was a significant main effect for phase of call (\(F_{(1.752, 129.670)} = 7.80, p < .01, \eta^2_p = .10\)) and the
effect for condition was marginally significant \( (F_{1, 74} = 3.44, p = .068, \eta^2_p = .04) \). For phase of call, post hoc analyses indicated that, across both conditions, the lowest amounts of switching from deep acting to surface acting occurred during Phase III \( (M = 0.24, SD = 0.46) \); this was significantly lower than switching from deep acting to surface acting during Phase I \( (M = 0.55, SD = 0.89) \) and Phase II \( (M = 0.67, SD = 0.82) \). Phases I and II were not significantly different from each other. For condition, post hoc analyses indicated that, across all phases, participants in the service recovery condition \( (M = 1.22, S.E. = 0.14) \) engaged in higher levels of switching from deep to surface acting than those in the service failure condition \( (M = 0.80, S.E. = .14, p < .05) \).

Finally, for switching from surface acting to deep acting (no correction was needed), results again indicated a non-significant interaction for phase of call and condition \( (F_{2, 148} = .23, n.s.) \). There was a significant main effect for phase of call \( (F_{2, 148} = 8.73, p < .001, \eta^2_p = .11) \) and condition \( (F_{1, 74} = 4.67, p < .05, \eta^2_p = .06) \). Post hoc analyses for phase of call demonstrated that the lowest amount of switching from surface acting to deep acting occurred for Phase III \( (M = 0.22, SD = 0.48) \); Phase III was significantly lower than Phase I \( (M = 0.63, SD = 0.89) \) and Phase II \( (M = 0.71, SD = 0.88) \). Phases I and II were not significantly different. For condition, post hoc analyses yielded that, across all phases, those in the service recovery condition \( (M = 0.59, S.E. = 0.08) \) engaged in slightly higher levels of switching from surface acting to deep acting than those in the service failure condition \( (M = 0.39, S.E. = 0.08, p = .068) \).

Overall, Hypothesis 5 was largely unsupported. To illustrate the effects, the means are plotted for switching ‘in general’ in Figure 19; the other two (deep acting \( \rightarrow \) surface acting, surface acting \( \rightarrow \) deep acting) followed the same pattern of results.
Figure 19 helps illustrate that, regardless of condition, the lowest levels of switching occurred during the final portion of the performance episode (Phase III), with Phases I and II not differing. This counters the idea that switching should increase as a function of negative social information from customers (e.g., Côté, 2005; Diefendorff & Gosserand, 2003; Rafaeli & Sutton, 1989). This would suggest that, for those in the service failure condition, switching would be highest during Phase III. Rather, my results suggest that, regardless of the social context, during a service interaction, individuals may ‘settle’ into choosing one type of emotion regulation strategy over the other. Further, given that the mean levels of switching across the entire call were higher for those in the
service recovery condition than those in the service failure condition, this suggests that individuals may utilize varying types of strategies during positive contexts, given that things are going well and they might perceive the ability to ‘experiment’ with their regulation techniques.

Given that switching was unrelated, I considered whether or not one strategy was used more often than the other. To do so, I calculated the proportion of time spent utilizing a greater amount of one strategy versus the other. The proportion of time individuals spent engaging in higher levels of surface acting and deep acting was calculating by identifying the points in time in which surface (deep) acting was higher than deep (surface) acting (i.e., coding this occurrence as a ‘1’ into two separate variables). Thus, values closer to ‘1’ indicate that the individual was primarily using one regulation technique over the other for that portion of the performance episode (i.e., overall, Phase I, Phase II, Phase III). Table 22 presents these proportions split across phases by condition.

Table 22. Proportion of time spent utilizing surface or deep acting a greater amount

<table>
<thead>
<tr>
<th></th>
<th>Entire Call</th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SF</td>
<td>SR</td>
<td>SF</td>
<td>SR</td>
</tr>
<tr>
<td>SA Greater</td>
<td>.41</td>
<td>.37</td>
<td>.28</td>
<td>.31</td>
</tr>
<tr>
<td>DA Greater</td>
<td>.32</td>
<td>.38</td>
<td>.38</td>
<td>.37</td>
</tr>
<tr>
<td>SA/DA Same</td>
<td>.27</td>
<td>.25</td>
<td>.34</td>
<td>.68</td>
</tr>
</tbody>
</table>

*Note.* SA = surface acting. DA = deep acting. SF = service failure condition. SR = service recovery condition.

For Phase I, there were no differences in the proportions of time spent utilizing deep or surface acting more within or between conditions. However, for Phase II, in the
service failure condition, participants utilized surface acting a greater amount of time than deep acting more often (.52) than they used deep acting over surface acting (.29; \( t_{(37)} = 2.33, p < .05 \)). Conversely, in the service recovery condition, the proportion of time spent utilizing higher levels of surface acting (.42) was equal to the amount of time spent utilizing higher levels of deep acting (.42; \( t_{(37)} = .02, n.s. \)). There were no differences between groups in using surface acting more than deep acting (\( t_{(74)} = -1.26, n.s. \)); the service recovery condition, however, used slightly higher levels of deep acting over surface acting than the service failure condition (\( t_{(74)} = 1.75, p = .085 \)). Finally, for Phase III, for the service failure condition (where the social context became increasingly negative), individuals reported spending slightly more time utilizing surface acting over deep acting (.44) than deep acting over surface acting (.25; \( t_{(37)} = 1.76, p = .086 \)). However, in the service recovery condition (where the social context became more positive), the proportions of time were comparable (surface acting higher: .38; deep acting higher: .32; \( t_{(37)} = .60, n.s. \)). There were no differences between conditions in the amount of time spent utilizing greater levels of surface acting (\( t_{(74)} = -.61, n.s. \)) or greater levels of deep acting (\( t_{(74)} = .90, n.s. \)).

Taken together, these results suggest that individuals can utilize equal amounts of time engaging in one strategy over the other, reflecting ideas presented by Beal and Trougakos (2013) and others (e.g., Diefendorff & Gosserand, 2003; Rafaeli & Sutton, 1989). The results also suggest that, in the context of customer incivility (i.e., Phases II and III of the current study) individuals may resort to utilizing surface acting a greater amount of time. This fits with Grandey et al. (2004) who found that individuals exposed to customer aggression were more likely to utilize surface acting and venting in order to
cover the negative emotions they were experiencing. Thus, the social context does have the ability to affect the type and amount of regulation utilized by employees, but not necessarily the extent to which employees switch between strategies.

Hypothesis 6

Hypothesis 6 proposed that the levels of surface acting averaged across the entire performance episode would be positively related to emotional exhaustion. Although the primary test of this hypothesis was for surface acting across the entire performance episode, I also tested the relationships of surface acting and deep acting with emotional exhaustion for each call phase and study condition. These results are presented in Table 23; as a note, mean emotional exhaustion was not significantly different across study conditions (service failure: \( M = 2.63, SD = 1.05 \); service recovery: \( M = 2.37, SD = 1.15 \); \( t(74) = -1.03, n.s. \)).

Table 23. Between-person correlations of focal variables with emotional exhaustion.

<table>
<thead>
<tr>
<th></th>
<th>Entire Episode</th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All SR SF</td>
<td>All SR SF</td>
<td>All SR SF</td>
<td>All SR SF</td>
</tr>
<tr>
<td>EE/Emo.</td>
<td>-.36** -.12</td>
<td>-.56** -.21</td>
<td>-.06 -.34*</td>
<td>-.35** -.20 -.52** -.30** .01 -.57**</td>
</tr>
<tr>
<td>EE/SA</td>
<td>.27* .16 .37*</td>
<td>.13 .11 .16 .23* .10 .38* .30** .16 .40*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE/DA</td>
<td>.25* .17 .32*</td>
<td>.12 .08 .15 .23* .17 .34* .28* .10 .39*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE/T-P Affect</td>
<td>.07 .02 .14</td>
<td>.11 .06 .14 .09 -.00 .18 -.02 -.04 .17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* All values are correlation coefficients (\( r \)). EE = emotional exhaustion. Emo. = participant emotionality. SA = surface acting. DA = deep acting. T-P = third-party. All = all participants (not split by condition). SR = service recovery. SF = service failure. Correlation coefficients for the service failure and service recovery conditions were tested with \( df = 36 \); for all participants, \( df = 74 \). * \( p < .05 \); ** \( p < .01 \).

Supporting Hypothesis 6, across the entire performance episode and all participants, a positive relationship emerged between surface acting and emotional
exhaustion ($r = .27, p < .05$). Though not hypothesized, a positive relationship also emerged between deep acting and emotional exhaustion ($r = .25, p < .05$), suggesting that even though deep acting may be ‘acting in good faith’ (Ashforth & Humphrey, 1993) it still constitutes an effortful process that can be viewed as depleting (Hochschild, 1983).

I then tested whether or not the relationship between emotional exhaustion and both regulation strategies across the entire performance episode differed based upon study condition. Results showed a positive relationship between emotional exhaustion and surface acting for the service failure condition ($r = .37, p < .05$) but not the service recovery condition ($r = .16, n.s.$); however these correlations were not significantly different from each other ($z = -.95, n.s.$). For deep acting, a positive relationship also was observed for the service failure condition ($r = .32, p < .05$) but not the service recovery condition ($r = .17, n.s.$); however, these correlations also were not significantly different from each other ($z = -.67, n.s.$).

Moving into the different phases of the call, for Phase I where customers were neutral with participants, no significant relationships emerged between surface acting or deep acting with emotional exhaustion across all participants ($r = .13, n.s.,$ and $r = .12, n.s.,$ respectively), for participants in the service failure condition ($r = .16, n.s.,$ and $r = .15, n.s.,$ respectively), or for participants in the service recovery condition ($r = .11, n.s.,$ and $r = .08, n.s.,$ respectively). Utilizing Preacher’s (2002) calculator for testing independent correlations, the relationship between surface acting and emotional exhaustion was not different between conditions ($z = .21, n.s.$), nor was there a difference between deep acting and emotional exhaustion between conditions ($z = .30, n.s.$). This
counters the idea that low levels of emotional exhaustion should be tied to low levels of both regulation techniques.

For Phase II of the call, however, relationships began to change. Across all participants, surface acting was positively related to emotional exhaustion \((r = .23, p < .05)\) and so was deep acting \((r = .23, p < .05)\). When splitting participants into conditions, significant relationships only emerged with emotional exhaustion for the service failure condition (surface acting: \(r = .38, p < .01\); deep acting: \(r = .34, p < .01\)) and not for the service recovery condition (surface acting: \(r = .10, n.s.;\) deep acting: \(r = .17, n.s.\)), though the relationships were not different from each other when utilizing Preacher’s (2002) test for independent correlations (surface acting: \(z = -1.25, n.s.;\) deep acting: \(z = -1.76, n.s.\)).

For Phase III, results mirrored those found during Phase II. Across all participants, emotional exhaustion was significantly related to surface acting \((r = .30, p < .01)\) and deep acting \((r = .28, p < .05)\). For the service failure condition, significant relationships with emotional exhaustion were found for surface \((r = .40, p < .05)\) and deep acting \((r = .39, p < .01)\), but no such relationships were found for the service recovery condition (surface acting: \(r = .16, n.s.;\) deep acting: \(r = .10, n.s.\)). Despite the apparent differences in correlations between the conditions, I again found that the relationships for emotional exhaustion with surface and deep acting were not significantly different across the service failure and recovery conditions using Preacher’s (2002) test for independent correlations \((z = -1.10, n.s.\) and \(z = -1.30, n.s.\)). However, it is possible that a lack of power contributed to the difficulty in detecting differences between the conditions.
In sum, I found support for Hypothesis 6, given the positive relationships that emerged between surface acting and emotional exhaustion. However, contrary to previous null findings (e.g., Bono & Vey, 2005), I found a positive relationship between deep acting and emotional exhaustion, suggesting that this effortful process can also place strain on employees (e.g., Hochschild, 1983). Moreover, these effects varied depending upon the phase of the performance episode and the study condition (even though the tests of these differences were non-significant). That is, for the service recovery condition, I found consistent, non-significant relationships between surface acting and deep acting across the three phases of the performance episode. Given that emotional exhaustion was captured post-performance episode, this finding suggests that when individuals have the ability to ‘recover’ from a challenging social interaction (i.e., Phase II of the call), emotional exhaustion is reduced. In contrast, for those in the service failure condition, the incessant negativity from the confederate callers was positively correlated with post-performance episode emotional exhaustion during Phases II and III. These findings suggest that increased negative social feedback from the customer did impact participants’ strain (i.e., emotional exhaustion), consistent with ideas proposed by Côté (2005). This also fits with theory from the emotions as social information model (Van Kleef, 2009), where emotions can communicate information to others that influences subsequent reactions and behaviors (i.e., customers’ emotions communicate information to employees).

Beyond correlation effects, I tested whether surface acting and deep acting, averaged across the entire performance for participants split by study condition, predicted emotional exhaustion above and beyond each other with and without self-rated
participant emotionality. I focused only on participants across the different conditions given that the emotional experiences for the service failure condition were different than the service recovery condition. Results are presented in Table 24.

Table 24. Regression models for emotional exhaustion across the entire episode split by study condition.

<table>
<thead>
<tr>
<th></th>
<th>DV: Emotional Exhaustion</th>
<th>Add Participant Emotionality</th>
<th>ΔR²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Service Failure</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.74 (.40)</td>
<td>3.75 (.67)</td>
<td></td>
</tr>
<tr>
<td>Surface Acting</td>
<td>.09 (.07)</td>
<td>.03 (.06)</td>
<td>.14†</td>
</tr>
<tr>
<td>Deep Acting</td>
<td>.02 (.06)</td>
<td>.04 (.05)</td>
<td></td>
</tr>
<tr>
<td>Participant Emotionality</td>
<td>-.16** (.05)</td>
<td>.23***</td>
<td></td>
</tr>
<tr>
<td><strong>Service Recovery</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.89 (.49)</td>
<td>2.29 (1.16)</td>
<td></td>
</tr>
<tr>
<td>Surface Acting</td>
<td>.03 (.08)</td>
<td>.02 (.08)</td>
<td>.03</td>
</tr>
<tr>
<td>Deep Acting</td>
<td>.03 (.09)</td>
<td>.02 (.09)</td>
<td></td>
</tr>
<tr>
<td>Participant Emotionality</td>
<td>-.03 (.08)</td>
<td>.00</td>
<td></td>
</tr>
</tbody>
</table>

*Note. Values are regression coefficients. Values in parentheses are standard errors.† p < .10; * p < .05; ** p < .01; *** p < .001.

Across the entire call for both conditions, there were no significant relationships for surface or deep acting predicting emotional exhaustion; the only relationship found was a negative relationships between participant emotionality and emotional exhaustion for the service failure condition (b = -.16, p < .01), suggesting that as participant emotionality became more positive, emotional exhaustion decreased. These results fit with the non-significant correlations for the service recovery condition. Further, the non-significant relationships between surface acting and deep acting are due to both variables being positively correlated with each other, ultimately canceling each other out.
To further explore the data, I also tested these relationships across the three phases of the performance episode (i.e., Phase I, Phase II, Phase III) split by study condition in an effort to determine whether the relationships changed depending upon the phase of the phone call. Results are presented in Table 25.
Table 25. Regression models for emotional exhaustion by condition and phase of episode.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Service Failure</th>
<th>Service Recovery</th>
<th>Phase II</th>
<th>Service Failure</th>
<th>Service Recovery</th>
<th>Phase III</th>
<th>Service Failure</th>
<th>Service Recovery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DV: Emotional Exhaustion</td>
<td>Add Participant Emotionality</td>
<td>$\Delta R^2$</td>
<td>Constant</td>
<td>Surface Acting</td>
<td>Deep Acting</td>
<td>Participant Emotionality</td>
<td>Constant</td>
</tr>
<tr>
<td></td>
<td>2.43*** (.27)</td>
<td>4.08*** (.79)</td>
<td></td>
<td>4.08*** (.79)</td>
<td></td>
<td></td>
<td></td>
<td>2.43*** (.27)</td>
</tr>
<tr>
<td></td>
<td>.02 (.07)</td>
<td>-.05 (.07)</td>
<td>.03</td>
<td>.02 (.06)</td>
<td>.06 (.06)</td>
<td>.12* (.06)</td>
<td></td>
<td>.03 (.05)</td>
</tr>
<tr>
<td></td>
<td>.03 (.07)</td>
<td>.04 (.08)</td>
<td>.01</td>
<td>-.00 (.09)</td>
<td>-.00 (.09)</td>
<td>-.02 (.08)</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.72*** (.40)</td>
<td>3.19*** (.60)</td>
<td></td>
<td>2.70* (1.09)</td>
<td></td>
<td></td>
<td></td>
<td>1.98*** (.49)</td>
</tr>
<tr>
<td></td>
<td>.06 (.05)</td>
<td>.04 (.05)</td>
<td>.15</td>
<td>.03 (.06)</td>
<td>.03 (.05)</td>
<td>-.13** (.04)</td>
<td>.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.02 (.06)</td>
<td>-.02 (.06)</td>
<td>.03</td>
<td>.05 (.06)</td>
<td>.03 (.06)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.05 (.07)</td>
<td>.02</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.47** (.47)</td>
<td>3.03*** (.69)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.91*** (.39)</td>
</tr>
<tr>
<td></td>
<td>.04 (.04)</td>
<td>.03 (.03)</td>
<td>.18</td>
<td>.05 (.03)</td>
<td>.01 (.03)</td>
<td>-.11** (.04)</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Values are regression coefficients. Values in parentheses are standard errors. 
† $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$. 

132
As can be seen, there were no significant relationships between surface acting and emotional exhaustion, as well as deep acting and emotional exhaustion, across the three phases of the call within each of the conditions. This null effect was found both with and without controlling for participant emotionality, and again fits with the non-significant correlations for the service recovery condition, and with the non-significant correlations for Phase I of the call for the service failure condition. However, for the service failure condition, the non-significant effects in Phase II and Phase III are due to surface acting and deep acting being positively correlated and canceling each other out in analyses.

Given the unique nature of the continuous data collection I utilized, another possibility in determining what promotes emotional exhaustion was looking at the proportion of time individuals spent engaging in no regulation, low levels of regulation, and high levels of surface and deep acting separately, and simultaneously. All analyses were conducted within the phases of the call with participants split by study condition to enhance the interpretation of effects. As a note, it is important to interpret results below with caution: although the proportions of times spent surface acting and deep acting at varying levels were averaged across the entire call and within each of the three phases, emotional exhaustion was measured only once at the end of the performance episode. Thus, levels of emotion regulation during Phase III of the call were the most proximal predictors of the emotional exhaustion ratings. Results are in Table 26.
Table 26. Correlations between emotional exhaustion and proportion of time spent surface/deep acting at different levels.

<table>
<thead>
<tr>
<th>Emotional Exhaustion with:</th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SR</td>
<td>SF</td>
<td>SR</td>
</tr>
<tr>
<td>No SA</td>
<td>.00</td>
<td>.21</td>
<td>-.03</td>
</tr>
<tr>
<td>Low SA</td>
<td>-.17</td>
<td>-.31</td>
<td>-.14</td>
</tr>
<tr>
<td>Mid-Level SA</td>
<td>.08</td>
<td>-.11</td>
<td>.01</td>
</tr>
<tr>
<td>High SA</td>
<td>.32</td>
<td>.44**</td>
<td>.11</td>
</tr>
<tr>
<td>No DA</td>
<td>-.08</td>
<td>.08</td>
<td>.01</td>
</tr>
<tr>
<td>Low DA</td>
<td>-.03</td>
<td>-.12</td>
<td>-.07</td>
</tr>
<tr>
<td>Mid-Level DA</td>
<td>.08</td>
<td>-.17</td>
<td>-.23</td>
</tr>
<tr>
<td>High DA</td>
<td>.23</td>
<td>.35*</td>
<td>.28</td>
</tr>
<tr>
<td>No SA &amp; DA</td>
<td>-.03</td>
<td>.17</td>
<td>-.03</td>
</tr>
<tr>
<td>Low SA &amp; DA</td>
<td>-.16</td>
<td>-.24</td>
<td>-.28</td>
</tr>
<tr>
<td>Mid-Level SA &amp; DA</td>
<td>.09</td>
<td>-.06</td>
<td>-.20</td>
</tr>
<tr>
<td>High SA &amp; DA</td>
<td>.27</td>
<td>.43**</td>
<td>.07</td>
</tr>
</tbody>
</table>

Note. All values are correlations with emotional exhaustion and the proportion of time spent utilizing a specific regulation strategy. SA = surface acting. DA = deep acting. SR = service recovery condition. SF = service failure condition. * p < .05; ** p < .01; *** p < .001.

During Phase I, the only significant correlations occurred between high levels of surface acting \((r = .44, p < .01)\), deep acting \((r = .35, p < .05)\) and joint levels of high surface and deep acting \((r = .43, p < .01)\) for the service failure condition only; there were no such relationships for the service recovery condition. Similarly, for Phase II, although there was no relationship between high levels of surface acting and deep acting for the service failure condition, there were positive correlations between high levels of deep acting and emotional exhaustion \((r = .43, p < .01)\) and joint levels of high surface and deep acting \((r = .44, p < .01)\). This suggests that, perhaps at the onset of customer incivility, putting forth effort to actually be positive with the customer is more problematic than just hiding how one is truly feeling. There were no such effects found for the service recovery condition during Phase II.
Finally, for Phase III, for the service failure condition significant positive relationships again emerged between emotional exhaustion and high levels of surface acting ($r = .35, p < .05$), high levels of deep acting ($r = .40, p < .05$), and joint levels of high surface and deep acting ($r = .39, p < .05$). Interestingly, for the first time in the service recovery condition, a significant negative relationship occurred between joint levels of low surface and deep acting at Phase III with emotional exhaustion ($r = -.35, p < .05$). This finding suggests that, under circumstances where customers are being pleasant, low levels of both forms of regulation may benefit individual well-being. Taken together, the results across the three phases suggest that when customers are difficult, high levels of either form of regulation are particularly problematic for employee well-being. However, during positive social contexts, such findings do not emerge (i.e., no form or level of emotion regulation was problematic for emotional exhaustion after the call); rather, results suggest that utilizing low levels of regulation was linked to lower emotional exhaustion when customers were pleasant.

Besides the proportion of time spent engaging in high/low levels of surface and deep acting, another possibility was that the amount of switching between strategies would impact emotional exhaustion. To test this idea, I ran correlational analyses between emotional exhaustion and the three conceptualizations of switching (in general, from surface acting to deep acting, from deep acting to surface acting) across the entire call and within each phase. Participants were split by condition in all analyses. Results are presented in Table 27.
Table 27. Correlations between emotional exhaustion and switching between regulation strategies.

<table>
<thead>
<tr>
<th>Emotional Exhaustion with:</th>
<th>Entire Episode</th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SR</td>
<td>SF</td>
<td>SR</td>
<td>SF</td>
</tr>
<tr>
<td>Switching In General</td>
<td>-.15</td>
<td>.08</td>
<td>.02</td>
<td>-.07</td>
</tr>
<tr>
<td>Switching DA → SA</td>
<td>-.18</td>
<td>.10</td>
<td>-.04</td>
<td>-.07</td>
</tr>
<tr>
<td>Switching SA → DA</td>
<td>-.11</td>
<td>.04</td>
<td>.08</td>
<td>-.06</td>
</tr>
</tbody>
</table>

Note. All values are correlations with emotional exhaustion and switching between regulation strategies. SA = surface acting. DA = deep acting. SR = service recovery condition. SF = service failure condition. * \( p < .05 \); ** \( p < .01 \); *** \( p < .001 \).

There were no significant correlations between emotional exhaustion and any form of switching across the entire performance episode or within the phases of the call for either condition. Thus, switching between regulation strategies was not related to the experience of emotional exhaustion.

Hypothesis 7

Hypothesis 7 proposed that at the within-episode level of analysis, (a) deep acting would be positively related to the hedonic tone of employee emotional expressions, and (b) surface acting would be negatively related to the hedonic tone of employee emotional expressions, with a third-party evaluating the emotional expressions of participants (i.e., evaluating vocal tone continuously throughout the performance episode). Analyses were across the entire call, across the entire call split by condition, across the three different phases, and across the three different phases with participants split by condition. Results are presented in Table 28.
Table 28. Cross-correlations for third-party affect ratings.

<table>
<thead>
<tr>
<th></th>
<th>Entire Episode</th>
<th></th>
<th></th>
<th>Phase I</th>
<th></th>
<th></th>
<th>Phase II</th>
<th></th>
<th></th>
<th>Phase III</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>SR</td>
<td>SF</td>
<td>All</td>
<td>SR</td>
<td>SF</td>
<td>All</td>
<td>SR</td>
<td>SF</td>
<td>All</td>
<td>SR</td>
</tr>
<tr>
<td>Emo/TPA</td>
<td>.44***</td>
<td>.39**</td>
<td>.49***</td>
<td>.10</td>
<td>.14</td>
<td>.06</td>
<td>.10</td>
<td>.09</td>
<td>.12</td>
<td>.40***</td>
<td>.46***</td>
</tr>
<tr>
<td>SA/TPA</td>
<td>-.49***</td>
<td>-.44**</td>
<td>-.53***</td>
<td>-.06</td>
<td>-.25</td>
<td>-.11</td>
<td>-.17</td>
<td>-.11</td>
<td>-.24</td>
<td>-.28**</td>
<td>-.29†</td>
</tr>
<tr>
<td>DA/TPA</td>
<td>-.40***</td>
<td>-.41**</td>
<td>-.40***</td>
<td>-.19†</td>
<td>-.12</td>
<td>.03</td>
<td>-.16</td>
<td>-.18</td>
<td>-.13</td>
<td>-.26*</td>
<td>-.34*</td>
</tr>
</tbody>
</table>

Note. All values are cross-correlations, with \( df = 74 \) across all participants and \( df = 36 \) within phases. Emo = participant emotionality. SA = surface acting. DA = deep acting. TPA = third-party affect ratings. SR = service recovery condition. SF = service failure condition. \( * \ p < .10; \ * * \ p < .05; \ * * * \ p < .01; \ * * * * \ p < .001 \)

Across the entire performance episode, the cross-correlation between third-party rated emotionality (i.e., hedonic tone of employee emotional expressions) with deep acting was \( r_{74} = -.40 \ (p < .001) \) and with surface acting was \( r_{74} = -.49 \ (p < .001) \). These results support Hypothesis 7b only, suggesting that increases in surface acting (i.e., hiding felt emotions) led to lower ratings of hedonic tone. Counter to my hypothesis, for deep acting, the more effort individuals put into feeling the emotions they were supposed to express with the caller, the less positive their hedonic tone sounded. Though not hypothesized, the cross-correlation between participant emotionality and third-party affect across the entire call was actually positive (\( r_{74} = .44, p < .001 \)), suggesting that the more positive participants felt, the more third-party raters were able to detect this.

For the entire call split by conditions, results mirrored those found in testing Hypothesis 7: negative relationships for surface acting across both conditions (service failure: \( r_{34} = -.53, p < .001 \); service recovery: \( r_{34} = -.44, p < .01 \); not significantly different from each other, \( z = -.49, n.s. \)), as well as negative relationships for deep acting across both conditions (service failure: \( r_{34} = -.40, p < .01 \); service recovery: \( r_{34} = -.48, p < .001 \); not significantly different from each other, \( z = .42, n.s. \)). Further, for participant
emotionality, there were positive relationships with third-party affect ratings for the service failure ($r_{(34)} = .49, p < .001$) and service recovery conditions ($r_{(34)} = .39, p < .001$), and these relationships were not significantly different from each other ($z = -.52, n.s.$).

When looking at all participants split across the three phases, the only instance that third-party rated emotionality related to surface or deep acting was in the final phase of the call (Phase III), with surface acting ($r_{(74)} = -.28, p < .01$) and deep acting ($r_{(74)} = -.26, p < .05$) both exhibiting negative effects. There was a marginally significant effect for surface acting in Phase I ($r_{(74)} = -.19, p < .10$) but no effect for deep acting ($r_{(74)} = -.06, n.s.$); no effects were found in Phase II (surface acting: $r_{(74)} = -.17, n.s.$; deep acting: $r_{(74)} = -.16, n.s.$). By utilizing the same follow-up analyses from Hypothesis 1 (i.e., running repeated measures ANOVAs on the Fisher’s Z transformed cross-correlations), I found that the cross-correlation for surface acting and third-party affect ratings was not significantly different across phases of the call ($F_{(2,84)} = 0.65, n.s.$), nor was the cross-correlation between deep acting and third-party affect ($F_{(2,84)} = 2.46, n.s.$).

Further, for participant emotionality and third-party ratings of affect, third-party affect ratings were positively correlated with participant emotionality only during Phase III of the call ($r_{(74)} = .40, p < .001$); there were no relationships for Phase I ($r_{(74)} = .10, n.s.$), or Phase II ($r = .10, n.s.$). A repeated measures ANOVA indicated a significant effect of phase ($F_{(2,94)} = 10.07, p < .001; \eta^2_p = .18$), with the cross-correlation in Phase III being significantly different from the cross-correlations in Phases I and II. Phases I and II were not significantly different from each other.

Given the very different caller experiences of participants across the three phases, results were further interpreted across the three phases split by condition (Table 27). For
the service failure condition, during Phase I there was no relationship between surface acting and third-party rated emotionality ($r_{(34)} = -.11, n.s.$); there also was a non-significant relationship for deep acting and third-party rated emotionality ($r_{(34)} = .03, n.s.$). Similar findings emerged for Phase II, with non-significant cross-correlations between third-party rated emotionality and surface acting ($r_{(34)} = -.24, n.s.$) and deep acting ($r_{(34)} = -.13, n.s.$). Finally, for Phase III, a significant cross-correlation was found between surface acting and third-party rated emotionality ($r_{(34)} = -.27, p < .10$), though this effect was marginal; there was no relationship for deep acting and third-party rated emotionality ($r_{(34)} = -.15, n.s.$). The relationship between surface acting and third-party rated emotionality did not differ across the three phases for those in the service failure condition ($F_{(2,34)} = 1.47, n.s.$). The relationship between deep acting and third-party rated emotionality also did not differ across the three phases of the call during the service failure condition ($F_{(2,34)} = 1.66, n.s.$). Finally, for third-party affect and participant emotionality, a marginally significant relationship was found during Phase III only ($r_{(34)} = .29, p < .10$), with no effects present for Phase I ($r_{(34)} = .06, n.s.$) or Phase II ($r_{(34)} = .12, n.s.$). The cross-correlations were not significant different from each other ($F_{(2,34)} = 2.72, n.s.$).

For participants in the service recovery condition, non-significant relationships were found during Phase I between third-party rated emotionality and both surface ($r_{(34)} = -.25, n.s.$) and deep acting ($r_{(34)} = -.12, n.s.$). For Phase II, non-significant relationships emerged between surface acting and third-party rated emotionality ($r_{(34)} = -.18, n.s.$) and between deep acting and third-party rated emotionality ($r_{(34)} = -.11, n.s.$). Finally, for Phase III, a marginally significant relationship was found between third-party rated
emotionality and surface acting \( r_{(34)} = -.29, p < .10 \), and a significant relationship was found with deep acting \( r_{(34)} = -.34, p < .05 \). Though there was variation in the significance of relationships, for third-party rated emotionality and surface acting, the relationship did not differ between phases of the call \( F_{(2,34)} = 2.12, n.s. \). Similarly, there were no differences across phases for the relationship between deep acting and third-party rated emotionality \( F_{(2,34)} = 1.20, n.s. \). For third-party affect ratings and participant emotionality, a significant, positive cross-correlation was found during Phase III \( r_{(34)} = .46, p < .001 \) only (Phase I: \( r_{(34)} = .14, n.s. \); Phase II: \( r_{(34)} = .09, n.s. \). The cross-correlation during Phase III was significantly different from Phases I and II \( F_{(2,58)} = 7.47, p < .01, \eta^2_p = .21 \); Phases I and II were not significantly different from each other.

As a final analysis, I compared whether the correlation between third-party rated emotionality and surface (deep) acting was significantly different within a specific phase between conditions. For third-party rated emotionality and surface acting, there was no difference between conditions during Phase I \( z = -.11, n.s. \), Phase II \( z = .46, n.s. \), and Phase III \( z = .05, n.s. \). In regards to third-party rated emotionality and deep acting, there also were no differences across the three phases (Phase I: \( z = .67, n.s. \); Phase II: \( z = .61, n.s. \); Phase III: \( z = 1.04, n.s. \).

In sum, results were consistently supportive of Hypothesis 7b, but not for 7a, given that surface acting did, as hypothesized, negatively relate to third-party rated emotionality. Deep acting, however, was proposed to have a positive relationship with third-party rated emotionality, the opposite of which was found. These relationships were consistent across the entire call and Phase III of the call, both across conditions and between conditions. Thus, even in voice-to-voice interactions with individuals who were
not witness to the original interaction and were unable to hear the customer-side of the
dialogue, both surface and deep acting negatively covaried with hedonic tone.

As a supplemental analysis, I tested whether switching between surface and deep
acting strategies exhibited relationships with average third-party affect ratings across the
three phases of the call. This tested the idea that high rates of switching between
strategies may cause truly felt emotions (i.e., negative emotions) to ‘leak’ out and
negatively affect third-party affect ratings. Results are depicted in Table 29.

<table>
<thead>
<tr>
<th>Third-Party Affect with:</th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switching In General</td>
<td>-.13</td>
<td>.01</td>
<td>-.30</td>
</tr>
<tr>
<td>Switching DA → SA</td>
<td>-.30</td>
<td>.14</td>
<td>.27</td>
</tr>
<tr>
<td>Switching SA → DA</td>
<td>.05</td>
<td>-.12</td>
<td>.20</td>
</tr>
</tbody>
</table>

Note. All values are correlations; third-party affect ratings were averaged within each of
the three phases. SA = surface acting. DA = deep acting. TPA = third-party affect ratings.
SR = service recovery condition. SF = service failure condition. † p < .10; * p < .05;
** p < .01; *** p < .001

As shown in Table 29, there were no relationships between average third-party
affect ratings during Phase I and Phase II. During Phase III, however, there was a positive
relationship between switching ‘in general’ and third-party affect ratings for the service
failure condition (r = .33, p < .05) as well as between switching from surface acting to
deep acting and third-party affect ratings (r = .33, p < .05). Given the context of Phase III
for the service failure condition (i.e., confederate callers were becoming angry), these
results suggest that when individuals were actively switching their regulation techniques
to try and manage a) their felt emotions and b) the social situation, third-party affect
ratings improved. More importantly, ratings improved when individuals in the service failure condition switched from surface acting to deep acting, suggesting that this switch may have enhanced the positive emotionality expressed by participants.

Conversely, for the service recovery condition, a negative relationship was found only between switching from surface acting to deep acting and third-party affect ratings \((r = -.37, p < .05)\). These results suggest that as participants decreased their switching from surface acting to deep acting, third-party affect ratings improved. Part of this relationship may largely be explained by the fact that participants during Phase III of the call for the service recovery condition were experiencing high levels of positive emotions (see Table 17). Regardless, the results suggest that when employees under negative social contexts in particular try and a) switching between strategies and b) switch from surface acting to more authentic deep acting, third-party affect ratings may actually improve.

**Surface Acting and Deep Acting Trajectories**

Although Hypotheses 3 and 4 addressed the extent to which individuals varied in their use of surface and deep acting across phases, these results were based solely on the mean, or average, level of surface acting and deep acting between phases of the call. As such, these analyses do not fully capture within-episode dynamics, such as the trajectory of surface acting and deep acting, or whether the surface acting or deep acting gradually or rapidly increased or decreased at different points in the call center simulation.

To answer this question, I used growth curve modeling to examine the nature of the within-person trends in the data. Prior to performing these analyses, I rescaled the time variable for all participants so that they would be on approximately the same time scale, ranging from 0 to 100. In doing this, I also attempted to ensure that the proportion
of time spent in each of the three phases was fairly constant across participants, with
about 36% of the call spent in phase I, 44% of the call spent in phase II, and 20% of the
call spent in phase III. This standardization procedure was used in an attempt to ensure
that each person’s time variable was on the same scale and that proportion of time spent
at each phase was approximately equal so that movement between phases would occur at
roughly the same point for all participants.

As a preliminary step, I first determined the amount of within-person variability
for participant emotionality, surface acting, and deep acting by running unconditional
means models (Kreft, DeLeeuw, & Aiken, 2006) in Hierachical Linear Modeling (HLM)
6.0 (Raudenbush & Byrk, 2002). Results are presented in Table 30.

Table 30. Unconditional means models for focal variables.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Within-Person Variance ($\sigma^2$)</th>
<th>Between-Person Variance ($\tau_{00}$)</th>
<th>% of total variance that is within persons (ICC(1))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant Emotionality</td>
<td>11.61</td>
<td>9.00</td>
<td>56.33%</td>
</tr>
<tr>
<td>Surface Acting</td>
<td>22.74</td>
<td>13.79</td>
<td>62.25%</td>
</tr>
<tr>
<td>Deep Acting</td>
<td>18.97</td>
<td>15.25</td>
<td>55.44%</td>
</tr>
</tbody>
</table>

Note. % of total variance that is within-person was computed using the formula
$\sigma^2 / (\sigma^2 + \tau_{00})$.

The results in Table 30 demonstrate that there was a substantial amount of within-
person variability for the three main continuous rating variables, ranging from 55.44% –
62.25%. To visually capture this within-person variability, I plotted the average
trajectories of participant emotionality, surface acting, and deep acting for participants in
the recovery and failure conditions (see Figures 20 and 21).
Figure 20. Within-person experience of emotionality, surface acting, and deep acting for the service failure condition.
Across both conditions, Phases I and II of the call were very similar. During Phase I, participants experienced fairly consistent neutral levels of affect, paired with low levels of emotion regulation. At the onset of Phase II (i.e., the beginning of customer incivility), however, participant emotionality began to decline (i.e., become more negative) and the amount of emotion regulation increased. For the first two phases, surface and deep acting were used roughly equivalent amounts. However, during Phase III, there were two key differences between conditions. First, in the service recovery
condition, participant emotionality increased (i.e., became more positive) as the customer became more pleasant, with levels of emotion regulation declining; for the service failure condition, participant emotionality became more negative, and emotion regulation continued to increase. Second, for the service failure condition, the levels of surface acting and deep acting appear to diverge in phase III, such that levels of deep acting plateaued, whereas surface acting continued to rise. This counters Phase III for the service recovery condition, where both surface and deep acting declined at similar levels.

For illustrative purposes, it is important to note that these average depictions of the service failure and service recovery condition do not fit each participant’s experience. Rather, some participants experienced a great deal of variability in continuous ratings, whereas other participants were more stable. To illustrate this point, I chose two participants from the service failure condition who exhibited low or high variability on all three of the constructs. I then mapped their ratings onto the average ratings to show the divergences that occurred in some cases across participants. For ease of interpretation, a separate plot was created for participant emotionality (Figure 22), surface acting (Figure 23) and deep acting (Figure 24).
Figure 2. Examples of variability in participant emotionality ratings.
Figure 23. Examples of variability in participant surface acting ratings.
As can be seen, whereas one person engaged in highly variable surface acting and deep acting throughout the entire performance episode and had highly variable emotional experiences, another participant was much more stable with their emotionality and use of regulation. These differences speak to the benefit of the continuous rating approach for understanding emotional labor processes: although the results thus far have depicted emotional labor as being variable at the momentary-level of analysis, this does not necessarily apply to all emotional labor actors.

Figure 24. Examples of variability in participant deep acting ratings.
In an attempt to model the surface acting, deep acting, and participant emotionality trajectories, I utilized Growth Curve Modeling (GCM) in HLM 6.0 (Raudenbush & Bryk, 2002) by adding linear and nonlinear terms to the unconditional models previously specified. According to Ployhart and Vandenberg (2010), GCM allows researchers to understand how constructs change over time and model non-linear effects.

At Level-1, I modeled the linear effect of time as well as various nonlinear terms (e.g., squared, cubed) in predicting participant emotionality, surface acting, and deep acting. I also modeled study condition (i.e., service failure, service recovery) as a Level-2 predictor of the intercepts and slopes. Level-1 predictors were uncentered and fixed; Level-2 predictors were grand-mean centered (Enders & Tofghi, 2007). Initial analyses suggested that a model with a cubed time predictor would not run, so all results below include only the linear time and squared time Level-1 predictors. Results are in Table 31.

Table 31. HLM estimates for growth curve models.

<table>
<thead>
<tr>
<th></th>
<th>Participant Emotionality</th>
<th>Surface Acting</th>
<th>Deep Acting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>γ</td>
<td>SE</td>
<td>γ</td>
</tr>
<tr>
<td>Level-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>14.24***</td>
<td>.34</td>
<td>2.00***</td>
</tr>
<tr>
<td>Time</td>
<td>-.14***</td>
<td>.00</td>
<td>.22***</td>
</tr>
<tr>
<td>Time²</td>
<td>.001***</td>
<td>.00</td>
<td>-.001***</td>
</tr>
<tr>
<td>Level-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>1.81*</td>
<td>.69</td>
<td>-1.56†</td>
</tr>
<tr>
<td>Condition x Time</td>
<td>-.14***</td>
<td>.01</td>
<td>.18***</td>
</tr>
<tr>
<td>Condition x Time²</td>
<td>.002***</td>
<td>.00</td>
<td>-.002***</td>
</tr>
</tbody>
</table>

Note. SE = standard error; robust standard errors were not used because effects were entered as fixed instead of random (Raudenbush & Bryk, 2002). Condition was coded as ‘0’ for service failure, and ‘1’ for service recovery. Level 1 n = 101; Level-2 n = 76. † p < .10; * p < .05; ** p < .01; *** p < .001.
For participant emotionality (see Table 31), the linear term for time was significant ($\gamma = -0.14, p < .001$) and accounted for approximately 4.9% of the within-person variance. Addition of the squared term for time resulted in a significant coefficient ($\gamma = .001, p < .001$) accounting for an additional 8.10% of the within-person variance in participant emotionality. Further, condition was a significant predictor of the intercept and both time predictor slopes, suggesting cross-level moderation. Figure 25 presents the latent curves for the two experimental conditions.

Figure 25. Growth curve model for participant emotionality.
In an attempt to better understand the nature of these effects, separate models were tested for each experimental condition with time and time-squared entered as Level-1 predictors. For the service recovery condition, time was a significant predictor ($\gamma = -.22$, $p < .001$; pseudo $r^2 = .00$) as was time-squared ($\gamma = .002$, $p < .001$; pseudo $r^2 = .214$). As shown in Figure 25, for those in the service recovery condition, although there was a decline in felt affect as the call became difficult, participants felt increasingly positive once the call began to ‘recover.’ For the service failure condition, time was a significant predictor ($\gamma = -.007$, $p < .001$; pseudo $r^2 = .233$) as was time-squared ($\gamma = .002$, $p < .001$; pseudo $r^2 = .002$). Combined, these results suggest that the change in participant emotionality was primarily linear and negative for the service failure condition, but was curvilinear for the service recovery condition, declining in the early part of the call and then becoming positive in the later part of the call.

For surface acting (see Table 31), the linear term of time was significant ($\gamma = .22$, $p < .001$) accounting for 20.8% of the within-person variance. Additionally, the squared term was significant ($\gamma = -.001$, $p < .001$), accounting for an additional 6.09% of the overall within-person variance. Further, condition was a significant predictor of the intercept and both time predictor slopes, again suggesting cross-level moderation. Figure 26 presents the latent curves.
Growth curve model for surface acting.

To understand the effects within each condition, I again ran separate models just for the service recovery and service failure conditions. For the service recovery condition, there was a significant effect for the linear term of time ($\gamma = .31, p < .001$, pseudo-$r^2 = .713$) and the squared term of time ($\gamma = -.002, p < .001$, pseudo-$r^2 = .187$). As shown in Figure 26, similar to participant emotionality, the relationship between surface acting and time was non-linear for those in the service recovery condition, peaking mid-way through the performance episode, and declining as the performance episode became more pleasant. For the service failure condition, time ($\gamma = .14, p < .001$, pseudo $r^2 = .427$) and
the squared term of time ($\gamma = -.0002, p < .001$, pseudo $r^2 = .0009$) were also positive predictors of surface acting, though the relationship between time and surface acting was positive and linear (i.e., surface acting increased across the entire performance episode).

Finally, for deep acting (see Table 31), I also found a positive effect for the linear time term ($\gamma = .20, p < .001$) that accounted for 15.6% of the within-person variance, and the squared time term ($\gamma = -.001, p < .001$) that accounted for 6.83% of the within-person variance across all participants. Condition was again a significant predictor of the intercept and both time predictor slopes, suggesting cross-level moderation. This relationship is depicted in Figure 27.

![Growth Curve Model for Deep Acting](image)

Figure 27. Growth curve model for deep acting.
Within the service recovery condition, both the linear time ($\gamma = .31$, $p < .001$, pseudo $r^2 = .071$) and squared time terms ($\gamma = -.002$, $p < .001$, pseudo $r^2 = .206$) were significant, with the same type of non-linear relationship emerging that was found for surface acting. For the service failure condition, the linear time ($\gamma = .09$, $p < .001$, pseudo $r^2 = .233$) and squared time terms ($\gamma = -.0001$, $p < .05$, pseudo $r^2 = .002$) were also significant, and also followed a similar pattern as was found in surface acting. Thus, for the service recovery condition, deep acting increased with customer incivility and decreased when the customer became pleasant; for the service failure condition, deep acting continuously increased as the call became more hostile.

Inspection of these curves and the plots presented earlier suggests that the rate of change in the continuous ratings was not constant across different experimental conditions. As such, another way in which the various experimental conditions might differ is in the rate of change in the continuous ratings. A gradual change suggests that participants may be experiencing gradual changes in their circumstances or that they are reacting slowly to such changes. In contrast, a rapid change suggests that participants are experiencing their environment as changing quickly and, as a result, their affective experience or way of regulating their affect is also changing rapidly.

As shown previously in Figure 20, for those in the service failure condition, the average trajectory of participant emotionality, surface acting, and deep acting over time, shows that deep acting plateaus during Phase III in comparison to surface acting, which appears to continually increase. Given the large amount of within-person variability uncovered, I calculated the amount of change per second that occurred for participants within each condition and within each of the three phases of the performance episode.

155
Positive values indicate an increase in a variable, whereas negative values indicate a decrease. I then calculated the average change per second for each of the phases and split by study condition. Results are in Table 32.

Table 32. Average amount of change per second in participant emotionality, surface acting, and deep acting ratings.

<table>
<thead>
<tr>
<th></th>
<th>Phase I</th>
<th></th>
<th>Phase II</th>
<th></th>
<th>Phase III</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SF</td>
<td>SR</td>
<td>SF</td>
<td>SR</td>
<td>SF</td>
<td>SR</td>
</tr>
<tr>
<td>Participant Emotionality</td>
<td>-.00</td>
<td>.01</td>
<td>-.02</td>
<td>.00</td>
<td>-.04</td>
<td>.08</td>
</tr>
<tr>
<td>Surface Acting</td>
<td>.05</td>
<td>.07</td>
<td>.05</td>
<td>-.00</td>
<td>.06</td>
<td>-.08</td>
</tr>
<tr>
<td>Deep Acting</td>
<td>.06</td>
<td>.06</td>
<td>.04</td>
<td>.00</td>
<td>.04</td>
<td>-.10</td>
</tr>
</tbody>
</table>

Note. SF = service failure condition. SR = service recovery condition. Negative values = increase; positive values = decrease.

The change scores presented in Table 32 largely reflect what was uncovered in the GCM analyses in terms of participant emotionality, surface acting, and deep acting increasing or decreasing across the phases of the call. What is new is that these analyses show the rate of change per second as increasing or decreasing across the three phases. To explore this, I conducted a series of repeated measures ANOVAs for participant emotionality, surface acting, and deep acting.

For participant emotionality, there was a significant interaction between condition and phases of call ($F_{(1.489, 110.189)} = 11.75, p < .001, \eta^2_p = .14$), with post-hoc analyses revealing a significant effect for phase of call for the service recovery condition ($F_{(1.339, 49.559)} = 9.67, p < .001; \eta^2_p = .21$) and a marginally significant effect for service failure condition ($F_{(1.669, 61.766)} = 3.31, p = .052; \eta^2_p = .08$). For the service recovery condition, pairwise comparisons revealed that the rate of change during Phase III (.08) was
significantly different from Phases I (.01) and II (.00), such that the rate of change in participant emotionality was greater during Phase III in comparison to Phases I and II (Phases I and II were not significantly different from each other). In other words, participant emotionality ratings would be expected to improve by about 1 point on the scale approximately every 12 seconds during Phase III compared to about every 100 seconds in Phase I (and even longer in Phase II).

For service failure, pairwise comparisons revealed that the rate of change in Phase III (-.04) was significantly different from Phase I (.00) but not Phase II (-.02); Phases I and II were not significantly different from each other. Specifically, affect would be expected to worsen by 1 point on the scale every 25 seconds in Phase III compared to every 50 seconds in Phase II and 100 seconds in Phase I. Thus, for the service failure condition, results indicated that emotionality was changing to become more negative, as opposed to becoming more positive in the service recovery condition.

For surface acting, there was a significant interaction between phase of call and study condition ($F(1.879, 139.023) = 11.84, p < .001, \eta^2_p = .14$), with post-hoc analyses demonstrating a significant effect for phase of call for the service recovery condition only ($F(2,74) = 21.67, p < .001; \eta^2_p = .37$); there was no effect for the service failure condition ($F(2,74) = .03, n.s.$). For the service recovery condition, pairwise comparisons revealed that the rate of change in Phase III (-.08) was significantly different from the rate of change during Phases I (.07) and II (-.00), and that Phases II and I were also significantly different from each other. These results indicate that surface acting was increasing during Phase I (by 1 point on the scale approximately every 14 seconds), remained steady during Phase II, and decreasing during Phase III (by 1 point approximately every 12 seconds).
Finally, for deep acting, I also found a significant interaction between phase of call and study condition \( F(2,148) = 6.91, p < .01, \eta^2 = .09 \), with post-hoc analyses demonstrating a significant effect for phase of call for the service recovery condition \( F(1.255, 46.420) = 10.61, p < .01, \eta^2_p = .22 \) and not the service failure condition \( F(2,74) = .79, n.s. \). Pairwise comparisons revealed that the rate of change during Phase I (.06) was significantly different from Phases II (.00) and III (-.10), and that Phases II and III were not significantly different. These results indicate that deep acting was increasing by one point approximately every 16 seconds in Phase I, remaining steady during Phase II, and decreasing one point every 10 seconds in Phase III. Thus, the rate of change for deep acting was significantly increasing during Phase I, but then remained fairly steady and declined in use during Phases II and III.

In sum, the supplemental analyses helped demonstrate that a) there is substantial within-person variability in emotionality, surface acting, and deep acting that can be non-linear under certain social contexts (i.e., negative → positive, as opposed to negative → increasingly negative), and b) the rate of change can vary depending upon the social context. I include these findings in my overall discussion in the following chapter. A summary of all of the results discussed in Chapter 4 can be found in Table 33.
Table 33. Summary of results.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hypothesis 1</strong>&lt;br&gt;At the momentary level of analysis, there is a negative correlation between surface acting and deep acting.</td>
<td>• Minimal Support&lt;br&gt;• The cross-correlation between surface/deep acting was positive across all participants for entire performance episode, for the entire performance episode with individuals split by condition, and across the time segments for all participants/participants split by condition&lt;br&gt;• The magnitude of this relationship did vary as a function of the condition and the time segment (range: .14 - .43), suggesting certain aspects of the social context (e.g., low levels of social information, high levels of customer incivility) may cause individuals to utilize one strategy more so than the other&lt;br&gt;• Suggests that surface/deep acting may be more complimentary of each other than opposing constructs</td>
</tr>
<tr>
<td><strong>Hypothesis 2</strong>&lt;br&gt;Episode-level mean surface acting and mean deep acting are positively correlated across subjects.</td>
<td>• Supported&lt;br&gt;• Across all ways of analyzing the data (all participants, participants split by condition, across the entire performance episode, within each of the time segments), surface/deep acting were positively correlated&lt;br&gt;• The episode-level correlation was stronger in magnitude than the momentary-level correlation, suggesting that ‘on average’ assessments may not be as reflective of what is happening on a momentary basis</td>
</tr>
<tr>
<td><strong>Hypothesis 3</strong>&lt;br&gt;In response to a service failure (i.e., heightened negative affective displays from the customer), mid-interaction levels of reported (a) surface acting and (b) deep acting are higher than those reported at the beginning of the interaction.</td>
<td>• Supported&lt;br&gt;• For participants in the service failure condition, levels of surface/deep acting increased from Phase I to Phase II&lt;br&gt;• For participants in the service recovery condition, levels of surface/deep acting increased from Phase I to Phase II&lt;br&gt;• There were no differences in the amount of surface/deep acting during Phase I and Phase II between conditions&lt;br&gt;• Though not hypothesized, participant emotionality decreased in Phase II (i.e., became more negative) from Phase I; this effect occurred within both of the conditions</td>
</tr>
<tr>
<td><strong>Hypothesis 4</strong>&lt;br&gt;The levels of (a) surface acting and (b) deep acting are higher during the third part of the service encounter for those participants in the increased service failure condition compared to participants in the service recovery condition.</td>
<td>• Supported&lt;br&gt;• For the service failure condition, surface/deep acting significantly increased from Phase II to Phase III; levels of participant emotionality continued to decline (i.e., became more negative)&lt;br&gt;• For the service recovery condition, mean levels of surface/deep acting declined close to Phase I levels; participant emotionality also improved (i.e., became more positive)&lt;br&gt;• Participants spent more time using maximal levels of surface/deep acting in the service failure condition; in the service recovery condition, more time was spent using little to no surface/deep acting&lt;br&gt;• Additional supplemental analyses indicated that when participants reported not using surface/deep acting, they were feeling more positively than when they were using surface/deep acting&lt;br&gt;• Supplemental growth curve analyses indicated a linear relationship for surface acting, deep acting, and participant emotionality was more appropriate for the service failure condition (i.e., acting more, feeling more negative); a curvilinear relationship was more appropriate for the service recovery condition (i.e., increasing acting and then decreasing again; feeling more negative and then feeling more positive again)</td>
</tr>
</tbody>
</table>
Table 33. Summary of results (cont.)

<table>
<thead>
<tr>
<th>Hypothesis 5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants in the increased service failure condition are more likely to switch between surface acting and deep acting strategies in the third phase compared to participants in the service recovery condition.</td>
<td></td>
</tr>
<tr>
<td>• Not Supported</td>
<td></td>
</tr>
<tr>
<td>• Switching did not increase as the call became more negative for those in the service failure condition (i.e., Phase III)</td>
<td></td>
</tr>
<tr>
<td>• Participants switched more during Phase I/II than during Phase III</td>
<td></td>
</tr>
<tr>
<td>• Participants switched more in the service recovery condition than in the service failure condition</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothesis 6</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Levels of surface acting averaged across a performance episode are positively related to emotional exhaustion.</td>
<td></td>
</tr>
<tr>
<td>• Supported</td>
<td></td>
</tr>
<tr>
<td>• Average levels of surface acting positively correlated with emotional exhaustion across participants; contrary to past findings, average levels of deep acting also positively correlated with emotional exhaustion across participants</td>
<td></td>
</tr>
<tr>
<td>• Split by condition, average levels of surface acting and deep acting overall and across the three phases positively correlated with emotional exhaustion for the service failure condition only; no such relationships emerged for the service recovery condition, suggesting recovery experiences can buffer detriments to one’s level of well-being</td>
<td></td>
</tr>
<tr>
<td>• Supplementary analyses demonstrated that, for the service failure condition, spending a large proportion of time utilizing maximal levels of surface/deep acting (as opposed to low or moderate levels, or no acting at all) positively correlated with emotional exhaustion only, indicating that high levels of surface/deep acting are what is problematic for employee well-being</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothesis 7</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>At the within-episode level of analysis, (a) deep acting is positively related to the hedonic tone of employee emotional expressions and (b) surface acting is negatively related to the hedonic tone of employee emotional expressions.</td>
<td></td>
</tr>
<tr>
<td>• Partially Supported</td>
<td></td>
</tr>
<tr>
<td>• Across the entire performance episode, there was a negative relationship between surface acting and third-party affect ratings; counter to expectations, across the entire performance episode, there was also a negative relationship between deep acting and third-party affect ratings</td>
<td></td>
</tr>
<tr>
<td>• Across the entire performance episode, participant emotionality was positively correlated with third-party affect ratings</td>
<td></td>
</tr>
<tr>
<td>• When the call was split by condition, for participants in the service failure condition, there was only a marginally significant negative relationship between surface acting and third-party affect ratings, in addition to a marginally significant positive relationship between participant emotionality and third-party affect ratings in Phase III only</td>
<td></td>
</tr>
<tr>
<td>• When the call was split by condition, for participants in the service recovery condition, there was a significant negative relationship between deep acting and third-party affect ratings and a marginally significant relationship between surface acting and third-party affect ratings; these suggest that third-party affect ratings became more positive as surface/deep acting decreased; the service recovery condition also exhibited a positive relationship between participant emotionality and third-party affect ratings, with all of these effects only being exhibited during Phase III</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER V
SUMMARY

Research on emotional labor has been pervasive in the organizational literature, resulting in many major overview articles, chapters, and meta-analyses (e.g., Bono & Vey, 2005; Grandey, 2000; Hülsheger & Schewe, 2011; Mesmer-Magnus et al., 2012) and even an edited book dedicated to the topic (Grandey et al., 2013). One of the primary emotional labor ‘topics’ has been the study of surface acting (i.e., faking how one is feeling with a customer) and deep acting (i.e., trying to actually feel positively in order to be pleasant with a customer) (Diefendorff et al., 2005), with an emphasis on determining if one regulation strategy is better than the other (e.g., Grandey, 2003). This accruement of this research, however, has come with methodological shortcomings and conflicting theory and results.

The current study is the first in the emotional labor literature to integrate continuous rating assessments (Ruef & Levenson, 2007; Mauss et al., 2005, 2011) in order to understand the moment-to-moment processes that occur in service interactions. In doing so, the within-person dynamics of emotional labor (i.e., utilizing surface acting, deep acting, or a combination of the two) was finally explored after being theoretically suggested in the literature (Beal & Trougakos, 2013). These within-person dynamics were studied within a carefully designed call center simulation, which manipulated the momentary social context that participants experienced, furthering our understanding.
of whether emotional labor depends on moment-to-moment variations in the social
c context (e.g., Côté, 2005; Van Kleef, 2009; Van Kleef et al., 2010). To date, no research
has been able to explore whether the social context matters at the within-episode level of
analysis, with most research focusing instead on the overall effect of a particular
customer or the typical effects of customers, in general (e.g., Grandey et al., 2004; Rupp
& Spencer, 2006; Sliter et al., 2010). Finally, the current study also utilized third-party
affect ratings (i.e., emotional performance) and participant-rated emotional exhaustion to
determine how momentary variations in surface acting and deep acting influence both
well-being and performance. Though previous research has utilized some form of third-
party emotional performance rating (e.g., Barger & Grandey, 2006), no work has
examined whether third-party raters are able to detect momentary fluctuations in
employee (i.e., participant) hedonic tone, which fails to address whether fluctuations in
hedonic ratings correspond to actual shifts or changes in emotion regulation.

The results of the current study theoretically advance the emotional labor
literature in several ways. First, I found that individuals do, in fact, utilize surface acting
and deep acting simultaneously within the same service interaction, an idea that has been
suggested (Beal & Trougakos, 2013) but dismissed as scholars continue to suggest
surface and deep acting as antipodes of each other (e.g., Mesmer-Magnus et al., 2012).
Further, I found that this relationship is stronger in magnitude when aggregating across
the performance episode (and ignoring within-episode variability) and making between-
persons comparisons as opposed to analyzing the relationships at the momentary level of
analysis. Second, the results showed that manipulated aspects of the social context shaped
participants’ momentary felt emotions and emotion regulation strategies in predictable
ways that conform to dynamic models of social interaction (Côté, 2005; Van Kleef, 2009). These results suggest that employees attend to and are affected by the emotional states of customers, and that they attempt to alter their emotion regulation strategies in response to how the customer is treating them (e.g., negative contexts increased emotion regulation and positive contexts decreased emotion regulation).

Third, I found that only under certain social contexts is emotion regulation problematic for end-of-the-call well-being. That is, when employees utilize high levels of emotion regulation and do not have the opportunity to recover within an interaction and have a positive social experience before the call is complete, both surface and deep acting positively relate to emotional exhaustion. Finally, cross-correlational results showed that third-party raters are able to detect variations in employee emotion regulation even during voice-to-voice interactions, such that participant variation in surface and deep acting negatively corresponded to third-party affect ratings. I explore each of these main ideas with respect to my primary research questions in the subsequent sections.

Can Customers Influence Employee Emotion Regulation?

Although emotional labor theory often posits the importance of employees shaping customers’ emotional states (Grandey et al., 2013), less work has paid attention to the role customers have in shaping employee emotion regulation. Building from theory presented by Rafaeli and Sutton (1989) and Côté (2005), I explored whether social information from customers caused participants to alter their emotion regulation strategy. I theorized that if the social feedback (i.e., customer emotional expression) is positive, no changes are needed; if the social feedback is negative, changes in emotion regulation are used to try to ‘fix’ the social interaction.
In the current study, the social information from the confederate callers significantly affected the use of surface and deep acting within a single performance episode. At the start of the call where minimal social information was available due to confederate callers being neutral (i.e., Phase I), participants felt fairly positive, exhibiting low levels of surface acting and deep acting. For Phase II, when the confederate callers acted irritated and provided negative social feedback (e.g., “I’m on a time crunch here with this stupid experiment and now you’re telling me there are extra steps? Why didn’t you tell me that before?!??!”), participants reported feeling significantly more negative affect and levels of both surface acting and deep acting increased. This trend of increasing levels of negative affect, surface acting, and deep acting continued during Phase III for those in the service failure condition, given that the confederate callers provided even more negative social information (i.e., became increasingly hostile, irate). Thus, as the social information became increasingly negative, individuals felt more negative affect and increased their level of regulation to manage their own feelings and expressions so as to ‘fix’ the interaction and manage the customer’s negative emotionality. Conversely, in the service recovery condition where the social feedback became positive during Phase III, participants decreased both regulation strategies.

Moreover, when considering the proportion of time individuals spent utilizing different strategies, when exposed to high levels of negative customer treatment in the service failure condition (Phase III), participants spent greater amounts of time using high levels of both regulation strategies in comparison to the previous phases of the call. For participants who experienced positive social feedback in the service recovery condition (Phase III), a greater proportion of time was spent utilizing little to no emotion regulation.
in comparison to what had to be used during the second, more challenging phase. This finding also occurred in the context of employees experiencing larger proportions of positive emotions, suggesting that customers, when acting pleasantly to employees, can improve employee emotionality and decrease the need for employees to utilize emotion regulation. Though this may be of little relevance to customers, there are significant implications from the side of management trying to enhance the well-being and performance of their employees (see Practical Implications for further discussion). Theoretically, the results also support many of the tenets in the Emotions as Social Information model (EASI; Van Kleef, 2009), such that customer emotions can communicate valuable information to employees (i.e., satisfaction with service, dissatisfaction with service), ultimately influencing employee emotionality and performance. Van Kleef (2009) articulates that this process largely operates via emotional contagion (i.e., mimicry, mirror-neuron activity, physiological feedback from vocal quality and intonation) or via interpersonal liking. Given that participant emotionality fluctuated in line with the confederate emotional expressions (i.e., positive in Phase I, neutral/negative in Phase II, and negative or positive in Phase III depending upon study condition), fitting with previous work (Van Kleef & Côté, 2007), participants may have ‘caught’ the negative or positive emotionality from callers, influencing their own emotionality and the need for emotion regulation.

Supplemental analyses of the trajectory of surface and deep acting helped shed light on these effects. First, there was a substantial amount of variability at the momentary-level of analysis, though, as demonstrated in Figures 22-24, there was individual variability in the amount of within-episode variability. Second, as shown in the
growth curve model (GCM) analyses, there was a curvilinear relationship in the service recovery condition between surface acting, deep acting, and participant emotionality, such that participants increased (decreased) their use of emotion regulation (emotionality) as the call reached the midway point (i.e., became more negative) and decreased (increased) their use of emotion regulation (emotionality) as the call became more positive. A linear relationship more accurately reflected emotion regulation and emotionality for the service failure condition, with emotion regulation steadily increasing, and emotionality steadily decreasing, as the performance episode continued. Finally, when looking at the rate of change, participants in the service failure changed their ratings at a fairly consistent rate across the call, such that deep acting and surface acting were consistently increasing; in the service recovery condition, participants increased their use of strategy at a more rapid rate during Phase I and decreased their use of either strategy at a more rapid rate during Phase III.

Another hypothesis related to social information was that participants would ‘switch’ between surface acting and deep acting more often when faced with an increasingly difficult customer. However, this finding was not supported, suggesting that participants switched relatively similar amounts across phases. Results did indicate that participants seemed to switch more during the service recovery condition, in addition to generally switching more during Phases I and II of the call (compared to Phase III). Largely, these results suggest that participants may tend to adopt a single strategy of choice, or, in the case of a positive social interaction, may perceive more flexibility in their use of emotion regulation strategies. One potential explanation for the lack of findings for strategy switching is that the difficult situation at hand (i.e., interacting with
a negative customer) may not have been as novel as anticipated. That is, participants may have had experience interacting with a rude customer and may not have needed to try out different strategies when attempting to handle the challenging interaction.

What is the Relationship Between Surface Acting and Deep Acting?

Though theory suggests that surface and deep acting are antipodes (e.g., Austin, et al., 2008; Kruml & Geddes, 2000; Mesmer-Magnus et al., 2012; Zapf, 2002; Zapf et al., 1999), a recent meta-analysis found that the population correlation between surface and deep acting is positive (ρ = .22; Hülsheger & Schewe, 2011). In light of this conflict between theory and findings, I proposed that at the person-level (i.e., across an entire performance episode) the established positive relationship would emerge, but at the momentary-level (i.e., within a performance episode), I argued that the often theorized negative relationship would emerge as individuals may be likely to only pursue one regulation strategy at a time.

Results were partially supportive of the idea that the nature of the surface acting and deep acting correlation would depend on the level of analysis. At the person-level, I found that average surface acting and average deep acting were positively correlated across the various experimental conditions (ranging from r = .70 to .78). At the momentary-level (using cross-correlations), I also found a positive relationship between surface and deep acting, but the strength and significance of this relationship varied as a function of the experimental condition and call phase (ranging from .14 to .43). For instance, in the service recovery condition, a positive relationship between surface acting and deep acting occurred for Phases II and III of the call, suggesting that individuals could and would simultaneously increase or decrease surface acting and deep acting
efforts. Indeed, supplemental analyses demonstrated that people reported simultaneously engaging in both strategies roughly 72.7% of the call, suggesting that surface acting and deep acting are not mutually exclusive regulation strategies.

Though these results are counter to my initial expectations, they were not fully unanticipated given that researchers have found positive relationships (e.g., Beal et al., 2011; Hülsheger & Schewe, 2011). However, they do suggest that theoretical models that present surface acting and deep acting as opposite or alternative strategies should be modified by allowing the strategies to be simultaneously used. There were fairly consistent findings across levels of analysis that surface acting and deep acting are positively related, though the magnitude of this relationship was weaker at the momentary level than at the aggregate level. Further, the relationship even became non-significant under certain conditions (e.g., when individuals were experiencing low levels of social information from confederate callers [Phase I] and when participants were experiencing high levels of negative social information [Phase III]), but in no instance was the relationship negative, as implied by prior theoretical work.

Does Participant Emotionality Covary with Momentary Emotion Regulation?

Although not a main hypothesis in the current study, I had the opportunity to explore how momentary variations in participant emotionality related with momentary use of surface and deep acting. Analyses revealed that both surface and deep acting exhibited negative relationships with participant emotionality ratings under certain circumstances. Cross-correlation analyses demonstrated that surface acting and deep acting largely negatively covaried with participant emotionality. These results suggest that the implementation of surface and deep acting is related to how one is currently
feeling. In the current challenging context, the most interesting exception to this general finding occurred during Phase III, where participant emotionality was unrelated to deep acting for the service failure condition, but negatively related to deep acting in the service recovery condition (the relationship was negative in Phase II for both conditions). This finding suggests that the relationship between participant emotionality and deep acting may be dependent on shifts in customer (i.e., confederate) emotionality. That is, in Phase III for the service failure condition, the non-significant relationship could indicate a lack of variability in deep acting and affect, meaning that the utilization of deep acting did not increase, even if confederate emotionality became worse. This lack of change in deep acting would generate the non-significant cross-correlation. Conversely, in the service recovery condition, when the confederate caller became more positive, the amount of deep acting utilized by the participant decreased (hence the negative relationship). It is also worth noting that Phase III of the service failure condition is the experimental condition in which the pattern of surface acting diverged the most from the pattern of deep acting; in particular, surface acting continued to increase in response to the rude customer but deep acting plateaued, suggesting that participants may have been putting more effort into suppression than into changing their feelings.

Are Surface and Deep Acting Harmful to Employee Well-Being?

A central focus of the emotional labor literature has been on understanding the influence of surface and deep acting on employee well-being (e.g., Bono & Vey, 2005; Grandey, 2000; Grandey et al., 2013; Hülsheger & Schewe, 2011). More specifically, scholars have been concerned with how surface acting and deep acting relate to employee emotional exhaustion (a key dimension of burnout). Previous results have indicated that
surface acting is problematic for emotional exhaustion (Brotheridge & Grandey, 2002; Brotheridge & Lee, 2002; Grandey, 2003; Hülsheger & Schewe, 2011; Judge et al., 2009; Pugliesi, 1999; Totterdell & Holman, 2003), with non-significant effects often emerging for deep acting (Brotherlidge & Lee, 2002; Brotheridge & Grandey, 2002; Grandey, 2003; Hülsheger & Schewe, 2011). Although the results of the current study found positive effects for average surface acting and deep acting throughout the course of the entire performance episode on end-of-the-call emotional exhaustion, the strength of this effect actually depended on the specific study condition and call phase.

Not surprisingly, average surface acting and deep acting from Phase I (in which the customer was neutral and both surface acting and deep acting were low) was unrelated to emotional exhaustion after the call for both experimental conditions. However, for Phases II and III of service failure condition, significant, positive correlations emerged for surface acting and deep acting with emotional exhaustion. In contrast, no significant correlations emerged for the service recovery condition in these phases. Thus, surface and deep acting were associated with emotional exhaustion in two out of the six experimental conditions tested in this study, and these conditions represented a customer progression moving from neutral to negative to even more negative. For these employees, the amount of surface and deep acting was linked to how emotionally taxed they felt after the call was complete. In contrast, when participants went through the same progression of customer treatment, but the customer became pleasant in the latter third of the call, there was no effect of ill effect of surface acting or deep acting on subsequent exhaustion.
Further, when considering the proportion of time individuals spent engaging in no surface and deep acting, as well as low, moderate, and high levels of surface and deep acting, results indicated that it was not just engaging in any level of surface acting or deep acting that was problematic for well-being. Rather, engaging in high levels of either form of regulation (or both simultaneously) was related to increased emotional exhaustion for those in the service failure condition (during all phases of the call). Thus, as long as individuals utilized low levels of emotion regulation, the negative influence on well-being could be minimized.

These results have three important implications. First, when employees are able to recover during a performance episode (i.e., satisfy a customer concern, see a positive change in customer demeanor), the harmful effect of surface and deep acting on emotional exhaustion is removed. Thus, even though individuals in the service recovery condition were exposed to customer incivility during Phase II, the positive experience at the end of the call mitigated the detrimental effect on well-being. Second, the results suggest that although in most situations deep acting was unrelated to end-of-call exhaustion, it can be challenging and exhausting for employees in social contexts in which employees are interacting with negative customers. Although it is possible that participant rated emotionality influences some of these effects, the results are suggestive of surface and deep acting not *always* being problematic for employees. Finally, if employees utilize low levels of (or no) surface and deep acting, the strategies are not problematic for well-being.
Are Surface and Deep Acting Harmful to Employee Emotional Performance?

In regards to performance, the current study is the first to collect continuous assessments of participant’s hedonic tone from third party raters, providing another unique contribution to the emotional labor literature. This allowed me to test for the first time whether third-party raters were able to detect momentary variations in vocal quality that corresponded to participants’ emotion regulation strategies. The emotional labor literature has proposed two ‘competing’ ideas: 1) both surface acting and deep acting would positively be related to emotional performance since both are designed to enhance outward positive displays (Hochschild, 1983), or 2) only deep acting will be linked positively to hedonic tone, given that surface acting may allow actual felt (negative) emotions to ‘leak’ out to customers (e.g., Elfenbein & Ambady, 2004). Fitting with previous meta-analytic results (e.g., Hülsheger & Schewe, 2011), I proposed that surface acting would be negatively related to third-party affect ratings (i.e., felt emotions would ‘leak’ out), whereas deep acting would be positively related to third-party affect ratings since they should be the more genuine regulation approach, with individuals actually feeling the positive emotions they were required to express.

Results showed that surface and deep acting were both negatively related to third-party affective tone ratings within the third phase of the performance episode for the service recovery condition, with only a significant relationship between third-party ratings and surface acting emerging for the service failure condition. However, the regulation strategies were unrelated to third party affect ratings in Phases I and II. For Phase I, the non-significant relationships may be occurring given that there was very little need for participants to regulate their emotions. During Phase II, although the need to
regulate was increased because the customer became more difficult, participants’ outward emotional expressions may not have been dramatically altered; this lack of variability could help explain the null result. Yet, during Phase III, the interaction either became more taxing (service failure) or becoming easier to handle (service recovery).

As such, the negative relationship of third party affect ratings with surface acting only for Phase III in the service failure condition suggests that participants were increasing their variability in the use of surface acting, but not deep acting (this is visually depicted in Figure 20). In contrast, for the service recovery condition, participant affect became more positive and the amount of emotion regulation decreased; thus, the negative relationship with both strategies in this situation suggests that participants were sounding more positive while also regulating to a lesser extent. Thus, results showed that third-party observers were able to detect variations in employee hedonic vocal tone that covaried with emotion regulation strategies. Moreover, results also showed that third-party observers might be sensitive to detecting when individuals are switching between strategies; more specifically, they may be able to detect when individuals switch from surface to deep acting. Also of interest, results demonstrated that third-party affect ratings positively covaried with participant self-reported emotionality, suggesting that customers can not only detect variations in emotion regulation, but can also detect the actual emotions being experienced by employees.

These findings demonstrate that third party observers may be able to detect emotion regulation in voice-to-voice interactions. Further, they suggest that both regulation strategies may be ‘leaky’ and allow customers to detect whether regulation strategy use is increasing (service failure) or decreasing (service recovery). Though these
relationships may be determined in part by self-reported emotions, they begin to garner support for the idea that surface and deep acting may relate to third party evaluations of employee affective quality.

Implications for Emotional Labor Theory

The current findings have significant implications for revamping emotional labor theory. For example, Grandey’s (2000) model of emotional labor is one of the most prominent in the literature, articulating a linear process of external factors (i.e., positive/negative events from the customer, emotional display rules) affecting surface and deep acting, which in turn influence employee well-being and performance. What is lacking from this model, however, is an acknowledgement that the process is dynamic with reciprocal effects operating. For instance, this model ignores the fact that the displays from the employee could serve as social feedback to customers, influencing customer emotional reactions, which creates a new emotional event to which the employee needs to respond. Other theoretical models have begun to describe dynamic and reciprocal processes in emotional labor theory (Beal & Trougakos, 2013; Diefendorff & Gosserand, 2003), but this is the first study to provide empirical evidence for these within episode moment-to-moment changes in emotional labor.

The two models in the emotional labor literature that do reference this dynamic nature are Côté’s (2005) social interaction model of emotion regulation, as well as Diefendorff and Gosserand’s (2003) control theory-based model of emotional labor. Although both theoretical models acknowledge that the emotional labor process is dynamic and can change based upon feedback from the situation or the customer, neither addresses the possibility that surface acting and deep acting can be used simultaneously.
The results of the current study support this idea given that at the momentary-level there was a positive relationship between surface acting and deep acting, though the magnitude varied as a function of the social context. Therefore, a new type of emotional labor model is needed that accommodates the view that these strategies can be used jointly. Further, a model that paints a truly dynamic within-episode view of emotional labor is needed.

There are several main propositions that emerge from the current study that will help inform this ‘new’ model of emotional labor. First, emotional labor varies on a momentary basis within a single performance episode. Fitting with Beal and colleagues (2005), employees encounter a variety of performance episodes within a single working day; Beal and Trougakos (2013) further stated that these performance episodes could represent customer interactions. The current study suggests that within a single performance episode, the use of surface and deep acting will likely occur, sometimes simultaneously, and sometimes separately. This result fits with the second change in emotional labor theory: deep acting and surface acting can be used simultaneously. Especially in negative contexts, employees may feel the need to put forth effort to actually be positive with callers, while simultaneously attempting to hide the frustration they are experiencing. This idea contradicts theory that places surface and deep acting as antipodes (e.g., Mesmer-Magnus et al., 2012), yet fits with the positive relationships emerging between surface and deep acting in the literature (e.g., Hülsheger & Schewe, 2011). From a practical perspective, employees may be able to simultaneously tackle the emotional labor problem through multiple means.

Another way in which emotional labor theory may be modified based on the results of this study is that more attention should be placed on the role of the social
context in dynamically shaping emotional labor. Building from ideas from Rafaeli and Sutton (1989) and Côté (2005), in the beginning of some social contexts, little to no surface or deep acting may be needed, suggesting that enforcing emotional display rules on employees also may not always be needed. Of course, in the current study, the social context was manipulated to be emotionally neutral and less demanding at the beginning of the call, minimizing the need for emotion regulation. Nonetheless, it is possible that if employee-customer exchanges start off on a similar neutral/positive tone, the amount of effort required by the employee to meet the emotional demands will be minimized. However, as interactions with customers continue, the amount of regulation required changes. During scenarios that may be characterized as service recovery experiences, increased regulation will be utilized when customers are difficult, but levels of regulation will decrease as customers become pleasant. Thus, employees may feel less pressure to ‘maintain what works’ in terms of emotion regulation, and instead feed off of the positive emotions from the customer. Conversely, in scenarios where service failure is occurring, emotion regulation will continuously increase, placing a great deal of strain on employees. Thus, the overall need to regulate, and amount of emotion regulation needed, must be articulated as more dependent on the social context.

Finally, the literature has often proclaimed that emotion regulation is problematic for employees; specifically, surface acting negatively impacts well-being, whereas deep acting often has mixed results. The current study demonstrates that a) both strategies can be problematic at certain levels and b) positive social feedback from customers can mitigate the negative effects of surface acting and deep acting on strain. Spending a large proportion of time utilizing high levels of either strategy in the service failure condition
positively related to the experience of emotional exhaustion, but a low to moderate level of these strategies was not problematic. In fact, for the service recovery condition, there was one result in which low levels of surface and deep acting combined was negatively related to emotional exhaustion, suggesting that low levels of regulation could, under specific social contexts, be beneficial for employees. These results suggest that the problematic stigmas surrounding emotional labor may be contingent on the level of surface/deep acting being utilized in addition to the social context surrounding the situation. Moreover, under circumstances where breaks or gaps between calls can encourage recovery, emotion regulation tactics may not cause as many problems for employees. Thus, overall, the social context and variability of surface and deep acting strategies should be better incorporated into future emotional labor theory and research.

Practical Implications

The results of the current study have significant implications for both managers and employees in service oriented jobs. First, the results of the current study demonstrate that employees do simultaneously use both strategies within performance episodes, which can be problematic given that emotional labor is taxing for employees (Hochschild, 1983) and both strategies have been linked to negative employee outcomes (Hülscheger & Schewe, 2011). In the current study, engaging in low to moderate levels of surface or deep acting was not problematic for employee well-being. Rather, it was the proportion of time spent engaging in high to maximal levels of either strategy (or both strategies at the same time) that strongly correlated with higher levels of emotional exhaustion. This effect was found for those in the service failure condition only, suggesting that employee well-being improves if employees can recover from customer hostility within the
performance episode. From the employee’s perspective, this result means that individuals need to be cognizant that they may be regulating too much, and try to maintain low to moderate levels of regulation if possible.

Moreover, even though display rules may be communicated on the job and employees will be encouraged to behave appropriately, employees could try to be mindful of instances where they actually feel positive emotions (i.e., fitting with positive display rules) and just communicate how they are actually feeling with the customer. Supplementary analyses indicated that when employees are not regulating, they are feeling higher levels of positive emotions, which would likely fit the emotional expectations of the job. Further, expressing naturally felt emotions could have a positive effect on employee well-being, as well as a positive influence on performance as referenced by the relationship between regulation and third-party ratings during the third phase for participants in the service recovery condition (i.e., as regulation decreased, positive ratings of employee hedonic tone increased).

From management’s perspective, these results suggest that respite breaks between interactions (e.g., Trougakos, Beal, Green, & Weiss, 2008) may help employee well-being and performance, especially when customers are being hostile. In the current study, individuals in the service recovery condition were buffered from the negative effects of customer incivility on emotional exhaustion. This service recovery experience was short in duration (lasting 59.68 seconds on average) suggesting that brief moments of positivity throughout the workday, either triggered by the customer or put in place by management, could have beneficial effects for employees. For instance, in call center settings, as opposed to taking calls one after the other rapidly, employees could be put on a 60-90
second waiting period between calls to compose themselves and ‘regroup’ if necessary. Similar types of experiences could be implemented in front-line service occupations (e.g., food sales, coffee shop barristas) where employees are encouraged to temporarily step away from the register and allow another employee to handle the next customer order. Moreover, if employees are exposed to high levels of incivility within a given performance episode, policies should be in place to allow them to temporarily leave the phone (or the floor if a front-line occupation) in order to ensure that well-being is not completely harmed.

Further, the results of the current study have unique implications for how to train emotional labor in the workplace. A study by Gabriel, Moran, Diefendorff, Leung, and Benedetti (under review) found that managers do report utilizing training on the job to communicate and enforce emotional labor demands (i.e., express positive emotions, hide negative emotions). However, the emphasis of Gabriel et al.’s study was on emotional labor ‘in general,’ meaning that employees were told to be positive with customers across all situations, ultimately encouraging the use of deep acting, primarily. The results from Hypotheses 3 and 4 suggest that employees, in actuality, may use varying levels of surface and deep acting depending upon the type of interaction they are engaging in: for difficult aspects of an interaction, both strategies may be needed, but at the onset of an interaction, or when customers are being pleasant, minimal regulation may be needed. Thus, training for emotional labor could be more nuanced, in that management could convey to employees the need to utilize regulation strategies only when customers are being challenging.
Moreover, the results of the current study suggest that employees could be trained on how to respond and adapt to customer incivility. Gabriel et al.’s (2013) focus was on training that encouraged certain types of emotional displays; they did not consider training for how employees handle customers who become rude or hostile. Employees could be encouraged to disengage from customer incivility by reappraising the situation differently (e.g., attributing the customer’s hostility as the situation the customer is in, as opposed to being ‘caused’ by the employee) to try and minimize the impact it has on their employee well-being. Additionally, given that service recovery experiences are possible, employees could be encouraged to try their best to ‘stick’ with the phone call (or service interaction), given that customers could become pleasant later in the service interaction, ultimately improving the employee’s emotional experience. These types of implementation tactics could help employees a) become more prepared for possible customer incivility and b) have novel ways of handling such experiences.

Further, managers need to be cognizant that, even in voice-to-voice interactions, customers may be able to detect variations in vocal quality that are attributable to employees’ felt emotions and emotion regulation strategies. Thus, managers must communicate to employees that, even when they are trying to regulate their emotions appropriately, especially during challenging interactions, the level of effort put forth can have a negative impact on evaluations of employees’ emotional performance.

Interestingly, third-party affect ratings were positively related to participant emotionality, suggesting that as participants were feeling more positively, third-party raters were able to detect this in their emotional performance evaluations. Fitting with Diefendorff et al.’s (2005) finding that expressing naturally felt emotions is a distinct regulation strategy,
perhaps managers should consider encouraging employees to display how they truly feel to the customer, as long as it fits with the emotional expectations of the job.

Limitations and Future Directions

As with any study, there are limitations. In regards to sample, the current study relied on undergraduate students, which may not generalize to other populations (e.g., older adults, full-time service employees, call center employees). Future research would be well suited to consider older, full-time employees as their sample to try and replicate the effects found, or determine if a different pattern of results emerge. For instance, research has established that older adults utilize emotion regulation strategies differentially than younger adults (Dahling & Perez, 2010; Sliter, Chen, Withrow, & Sliter, in press); as such, older adults may have different levels of variability in the extent to which they use both strategies within a given performance episode.

Moreover, in the current simulation, participants were only exposed to one ‘real’ call from a confederate caller. In an actual call center, employees would be exposed to a much higher volume of calls, slightly detracting from the participants’ experience of being in a ‘real’ call center. The current study did highlight the basic structure of a difficult call center interaction, and participants felt as though the experience was realistic (participants’ rated the experience as fairly realistic: $M = 3.83$ [5-point scale], $SD = 0.80$; see Appendix J for the measure designed for this study). Nonetheless, an important extension of this work would be for actual call center employees to provide continuous assessments of their call center experience. For instance, in collaboration with a call center, recordings of employee phone calls could be provided and put through the same E-Prime 2.0 computer program used in the current study. A wide range of calls could be
rated, from pleasant to difficult, in order to look at the different types of social feedback being provided. Moreover, by providing continuous ratings of multiple calls that occurred back-to-back, we would be able to gain insight into how employees regulate their emotions over multiple phone calls, and if techniques change the longer the employee is ‘on the floor.’

From an analytic standpoint, I looked at three distinct phases of the call, creating three averages of emotional experience for each participant. Although this choice fit my study design (see previous Figure 2), another choice would be to identify specific ‘events’ within the call. For instance, one could identify portions of time the customer is talking versus when the employee (i.e., participant) is talking to look for changes in participant emotion regulation (e.g., do people regulate more when they are speaking compared to when another person is speaking?). Additionally, one could mark each instance of specific customer incivility, determining if each incivility experience is unique on participant emotion regulation, or if their effects are more multiplicative. These types of ‘data marking’ would be possible within a continuous rating framework.

In terms of research design, in the current study I focused on one aspect of surface acting and one aspect of deep acting for the participants to rate. Although these two aspects (hiding and changing one’s emotions, respectively) are central to the definitions of surface and deep acting (Diefendorff et al., 2005), other choices could have been made. For instance, future research could look at ‘faking’ or ‘masking’ for surface acting, as well as the extent to which participants ‘actually tried to experience the required emotions’ when assessing deep acting (see Appendix J for full measures). Additionally, Diefendorff et al. (2005) included a dimension of naturally felt emotions in their study of
emotional labor, finding that this was distinct from surface and deep acting. Future research may consider adding a continuous rating of naturally felt emotions, to see if it corresponds to when participants report simultaneously low levels of surface and deep acting during an exchange.

Further, although I opted to utilize the ‘traditional’ emotional labor measures of surface and deep acting for the current study, Gross (1998) has highlighted a variety of emotion regulation constructs, some of which have correspondence to surface and deep acting (Grandey, 2000). For instance, Gross (1998) has a process-based emotion regulation model starting with an emotional cue, which triggers antecedent-focused and response-focused emotion regulation strategies. Reappraisal is a main antecedent-focused strategy, defined as an effortful process where individuals’ attempt to reinterpret a potentially emotionally arousing situation in unemotional terms. The idea of reappraisal, though not directly parallel, mirrors ideas found in deep acting, wherein individuals attempt to consciously alter their emotional experience in the hopes of having appropriate emotional reactions. Conversely, suppression occurs when individuals actively try to inhibit an emotional reaction even though they are experiencing it (i.e., inhibiting outward displays of frustration or anger when feeling negatively). Suppression closely imitates surface acting, given that both involve a ‘hiding’ or ‘faking’ component. A unique follow-up study would be to see how these two related yet distinct emotion regulation concepts operate in a call center simulation on a momentary basis.

Additionally, although past precedent (e.g., Goldberg & Grandey, 2007) calls for all confederate callers to be male for consistency, this prohibited testing whether emotional labor effects depend on the gender of the caller. For instance, it may be the
case that participants would interpret anger differentially coming from a male, rather, than a female, confederate caller due to stereotypes about males being more agentic and females being more nurturing and compassionate (Hess, Thibault, Adams, & Kleck, 2010). Another possibility is that participants may be less influenced by hostility if it is coming from someone of the same sex (i.e., female participants may respond more positively to female confederates) given biases such as the ‘similar-to-me’ effect (e.g., Anderson & Shackleton, 2011). These questions remain unanswered, and warrant future explanation.

Finally, in the current study, all continuous ratings were presented on a single, horizontal axis. Although this allowed for continuous ratings of surface acting and deep acting fairly efficiently, it precluded a full assessment of the participant’s affective experience by restricting participants to only rating the valence (i.e., positive, neutral, negative) of their emotionality. That is, work by Russell (1980) identified that felt emotions can be placed onto an affect circumplex, with varying levels of valence (i.e., positive – negative) and activation (i.e., high – low). Thus, even though participants were able to distinguish their positive or negative emotional experiences, their level of overall emotionality activation was not captured with the current computer software. Indeed, this limitation was highlighted by Ruef and Levenson (2007), who stated that all scales used in continuous rating assessments must be placed on a two-pole continuum; this is regardless of whether one uses a rating dial, a slider, or a mouse-movement recording computer program like the current study. As a solution, Lizdek, Sadler, Woody, Ethier, and Malet (in press) proposed that individuals could move the cursor throughout a two dimensional rating space indicating both the valence and activation of their felt affect.
This would allow a full affective circumplex to be rated, as well as an intersection of constructs of interest to emotional labor. For instance, Lizdek and colleagues provide the example of a four-quadrant model, with unfriendly – friendly behavior on the x-axis, and submissive – dominant behavior on the y-axis. In an emotional labor study, participants could provide ratings of customer emotionality (i.e., positive – negative) and their own felt emotionality (i.e., positive – negative) to name two options. In sum, though the current research design had many advantages above and beyond the research paradigms typically used, next steps are present.

Conclusion

Though interest in emotional labor is flourishing, research has been largely conducted using ‘in general’ assessments of emotional labor that fail to a) capture the dynamic nature of emotional labor processes and b) enhance the study of emotional labor both theoretically and methodologically. The current study integrated ideas from the social information model of emotional labor (Côté, 2005) with continuous rating assessments (Ruef & Levenson, 2007; Mauss et al., 2005, 2011) to understand what happens in the moment for service employees as they are exposed to different social contexts (i.e., service failure, service recovery). In sum, I found that individuals do vary in the extent to which they use both surface and deep acting strategies within a single performance episode, with the momentary-level cross-correlation being much weaker than the person-level correlation. This suggests that on a moment-to-moment basis, individuals are varying in their regulation strategy choice. Moreover, I found that it is not just engaging in surface or deep acting that is challenging for employees. Rather, it is engaging in high levels of surface and deep acting that creates problems for employee
well-being only in contexts where customers are increasingly hostile. Thus, in service
contexts where customers are challenging, management must be cognizant of the amount
of time employees have to recovery from their experience.
REFERENCES


APPENDIX A

TRAINING MATERIALS – SURFACE ACTING FIRST

PRIOR TO PARTICIPANT ARRIVAL:

- **In the Confederate Lab:**
  - Load up Skype on the confederate computer; on the confederate computer, have ‘UTSp社会实践’ (password: XYYYYYYYYY) loaded.
  - Let the confederate know which script he will be using.
  - Make sure you have the confederate phone number so you can call him at the end of the training session (i.e., “Hey Pete, we’re live now!”)

- **In the Participant Lab:**
  - Load up Skype on the participant computer (UTSrep14; password: XYYYYYYYYY).
  - Double click the ‘PowerGramo’ icon on the desktop to make sure PowerGramo is recording in the Skype software.
  - TURN DOWN THE VOLUME ON THE PARTICIPANT COMPUTER!!!! It should be as low as possible (around 2-3%).
  - Make sure the following are out for the participant:
    - All physiological equipment, complete with new sensors
    - Headset
    - Caller ID paper slips
    - Pen
    - Participant Instructions
    - PowerPoint booklet
    - Call instruction guide (should be taped to the desk)
  - Get E-Prime ‘started’ on the rating computer.
  - Turn on the BioData system and make sure:
    - The participant’s name is entered into the BioData system
    - The BioData system/Bluetooth system is working!

Once the participant arrives, have the participant sit at the computer where Skype is already loaded on the screen.

“Hi, my name is ________________, and I am the primary researcher for today’s study. Thank you for participating in our research. The goal of this study is to explore customer service in call centers. We are actually partnered with a local call center here in Akron who asked us to assist them with this project. Over the next hour, you will play the role of a service representative at University Tech Support and answer some calls via
Skype, which is an internet-based phone service that is used in the call center we are partnered with.

Today’s study is unique because we are partnering with another study occurring right now at another university in Ohio. Students in that study are creating a PowerPoint presentation and have been provided the phone number of our University Tech Support call center if they have any questions. You will be fielding calls from these students and assisting them through step-by-step instructions in the Microsoft PowerPoint manual we have created. Students enrolled in that study are competing for being entered into a lottery for $100 if they are the first to finish in their session, so it is important that you help them as best as you can. Because they think they are calling a real call center, it is imperative that you are happy, helpful, and respectful. Be sure to be positive and enthusiastic! They think they are calling a real technology help hotline.

In front of you are key materials for the study:
- **Caller ID Slips:** Before you can help the student, you must make sure that you ask for their participant ID number. You cannot help the student if they do not provide you their number. The researchers in the other study will use this ID number to see which of their participants called the hotline. When you get their ID number, just write it down on one of these caller slips and put the date and time.
- **Greeting and Closing Guidelines:** Taped to the desk is a list of instructions reviewing the steps that are to be followed with each call. Please make sure you remember to always say the greeting, follow-up, and closing statements.
- **PowerPoint Manual:** In front of you is the PowerPoint manual that University Tech Support has provided. Inside, you’ll find step-by-step instructions that you can run student callers through. University Tech Support piloted this training manual, and found that callers rarely asked questions that were not contained somewhere in this manual. When a caller asks for something, simply flip through the manual, find the correct heading, and read the instructions step-by-step.

Because you are working in a call center, it is important that you follow University Tech Support guidelines. **We want you to be happy, positive, and respectful for all customers** who are calling into the hotline. To help ensure this, **your calls will be monitored and recorded** via Skype for service quality. This is standard in all call centers. Customers who call into the call center will also be **rating you on the customer service you provide them.** We will be making sure that you **express happy and positive emotions no matter** what happens during the call. In order to show how sincere we are about instilling good customer service, **participants who have great customer service will be entered into a raffle to win one of two $50.00 gift cards to Amazon.com.** Your entry will be contingent on your performance, so **make sure to be as positive and enthusiastic as you can!**

I’m now going to leave you for a few minutes to review the training materials and the PowerPoint manual. I’ll be back in about five minutes.”
Leave the room. Enter after five minutes have passed.

“ I hope you found the materials helpful. Do you have any questions?”

Answer any questions the participants have.

“Okay, to make sure you’re all set, I’m going to go into another room and call you and run through a brief scenario with you. This will let you practice the mandatory greeting and closing lines, and it will make sure you’re being as pleasant as we need you to be. When the phone rings on Skype, please answer it. We’ll keep it short.”

Leave and call participant from confederate lab.

Participant:  Thank you for calling University Tech Support. My name is ______ and I am happy to serve you. Can I have your participant number please?

YOU:  Sure. It’s 8714.

Participant:  Thank you. How can I help you today?

YOU:  Oh, I just need to insert a table.

Participant:  OK – make sure you count now and find out how many columns or rows you need before we create your table.

YOU:  No problem. I just need four columns and four rows.

Participant:  Click “Insert” at the top of the screen. Go down and click on “Table.”

YOU:  Great. What’s next?

Participant:  Use the up and down arrows next to “Columns” to increase or decrease the number of columns. Then use the up and down arrows next to “Rows” to increase or decrease the number of rows. Then, click “OK.” You can enter text into each box in the table as needed.

YOU:  OK – great. It’s in.

Participant:  Is there anything else that I can assist you with today? We are always happy to help you through any problem.

YOU:  Nope!
Participant: Thank you so much for calling University Tech Support! It has been my pleasure to assist you. Have a great day!

After call, return to lab room. USE THIS TIME TO CORRECT THE PARTICIPANT IF THEY DID SOMETHING WRONG (i.e., missed a key “required” line from the manual, didn’t list instructions step by step, etc.).

“I think you can be even happier. Let your smile show through the phone!

Today’s study actually has a second component to it that I would like to walk you through today. After you are done taking calls, you’ll be providing some continuous ratings. To practice, I want to use the call we just made. Please move over to the computer next to you.”

IMPORTANT!! As participant moves, pull the sound file off of the ‘PowerGramo Player’ and export it as a ‘.wav’ file entitled one.wav. This should be exported to the hard drive. Unplug the hard drive, and plug it into the USB extension cord.

On the rating computer, pull up data file and put it into the ‘Calls’ folder on the desktop. You are then ready to run the ePrime program. To start the program, click the purple man on the tool bar.

“For these continuous ratings, you’ll be making three ratings per phone call. The first rating is about how you feel emotionally during the call. You will see a scale on the computer screen ranging from negative emotions on the left to positive emotions on the right, with the numbers 1-20 on the scale. During the call, you will be able to move the mouse as much or as little as you want along the scale depending on how you’re feeling. It is important to make sure you keep the mouse within the boxes at all times. Do you have any questions? If not, you can press the space bar to make your first rating.”

Direct participant toward the instructions on the computer screen, and then instruct them to proceed with providing their continuous ratings of emotions. When they tell you they are done, hit ‘F1’ to go to the blank screen.

“Great! Do you have any questions so far?”

Answer participant questions.

“Since the current study is interested in the emotional experience of call center workers, the next two ratings you are going to provide have to do with how you regulated your emotions during the call. In the first rating, I want you to rate the extent to which you felt varying levels of surface acting during the encounter. Surface acting means that you felt like you had had to hide how you were really feeling from the customer. For example, you might have felt negative emotions, but you hid them and displayed positive emotions to the customer instead. As you listen to the call, you can move the mouse along the scale
depending on how much you felt you were hiding your emotions with the caller at that point in time.

For these ratings, you will also have a continuous scale ranging from 1-20. A rating of “1” reflects not having to surface act at all. For a rating of “1,” you could say, “I am not hiding my emotions at this time with the caller.” For a rating of “10,” you could say, “I am hiding my emotions some at this time with the caller, but I am not completely acting with them.” For a rating of “20,” you could say, “I am hiding what I really feel so the customer thinks I’m really happy when I’m not.”

Again, you can move the mouse as much or as little as you like. This means you can stay in one place on the scale, or you can move it a lot. This depends upon your personal emotional experience! Please stay within the boxes, and make sure you are hovering over one of the numbers at all times.

Do you have any questions about what you’re supposed to be rating?

*IF YES: Clarify. IF NO: Press F1 to bring up the surface acting screen.*
*Let participant practice making the surface acting rating.*

“Now that you have rated the extent to which you used surface acting, I would like to ask you to provide continuous ratings of your experience of deep acting. Deep acting is when you put forth effort to try and really feel positively about a situation, or an interaction with a customer. For example, you might really try to feel positively so you can be pleasant with the customer. As you listen to the call, you can move the mouse along the scale depending on how much you felt you were trying to genuinely feel the emotions needed with the caller at that point in time.

For these ratings, a rating of 1 reflects not deep acting at all. For a rating of “1,” you could say, “I am not experiencing any effort trying to change my emotions.” For a rating of “10,” you could say, “I am putting in a moderate amount of effort to try to feel the emotions I am supposed to express with the caller at this time.” For a rating of “20,” you could say, “I am making a full effort to actually feel the emotions that I need to express with the call at this time.”

As a reminder, you can move the mouse as much or as little as you want depending upon your experience with the caller. Please stay within the boxes, and make sure you are hovering over one of the numbers at all times.

Do you have any questions about what you’re supposed to be rating?

*IF YES: Clarify. IF NO: Press F1 to bring up the deep acting screen.*
*Let participant practice making the surface acting rating.*

“You’re all finished now with the practice ratings. Please go back to the call center computer.
At the request of the call center we are partnered with, we are also going to be recording physiological measures today. This is fairly new for the company, so I appreciate your cooperation. Is it okay for me to proceed? We have tested these measures on ourselves, and we can assure you that they do not create any discomfort.

Before we put on the physiological sensors, you’ll first want to put on the headset.

These three chest sensors are also going to track heart rate and pulse. For the red sensor, please place the adhesive under your rib cage on the left. For the white sensor, please place it on your collar bone on the left above the red sensor. For the black sensor, please place it on your collar bone on the right parallel to the white sensor.

On your left hand, we are going to attach three sensors. The first two (VELCRO) go on your ring and middle finger. These detect skin conductance, which is an indicator of stress. The finger grip (GREY CLAMP) goes on your index finger. This detects heart rate and pulse.

For the physiological measures to work, it’s really important that you leave your hand we are using completely still, and that you don’t move around in your seat a lot. If for any reason you feel uncomfortable with the physiological measurements, please let me know.

Hook participant up to sensors and hand devices.

OK – you’re set! I’ll call the other university to let them know your line is open – their students have been working for about 15 minutes now, so you might get some calls pretty quickly. In order to help monitor the physiological data, I will be staying in the room with you. However, I am unable to help you with any portion of the task.

Remember: we’re monitoring you for your emotional expressions, so be happy and don’t let students in the other study hear any negative emotions! Please keep answering the calls until I return. We’ll do more ratings once you are finished with all the calls.”

Sit at the physiological computer and load the physiological program. Create a new participant session with the participant’s name and ID number.

THEN: CALL THE CONFEDERATE! Say: “Hey, Pete. Yep, we’re live.” Confederates should wait 20 seconds from that point to begin the first call. The physiological data should start recording immediately (Click ‘Record’). When the participant answers the phone, press the power button on the wireless device quickly one to ‘mark’ the physiological data for call start.
APPENDIX B

TRAINING MATERIALS – DEEP ACTING FIRST

PRIOR TO PARTICIPANT ARRIVAL:

- **In the Confederate Lab:**
  - Load up Skype on the confederate computer; on the confederate computer, have ‘UTSpracticecall’ (password: XXXXXXXXXXX) loaded
  - Let the confederate know which script he will be using
  - Make sure you have the confederate phone number so you can call him at the end of the training session (i.e., “Hey Pete, we’re live now!”)

- **In the Participant Lab:**
  - Load up Skype on the participant computer (UTSrep14; password: XXXXXXXXXXX)
  - Double click the ‘PowerGramo’ icon on the desktop to make sure PowerGramo is recording in the Skype software
  - TURN DOWN THE VOLUME ON THE PARTICIPANT COMPUTER!!!! It should be as low as possible (around 2-3%)
  - Make sure the following are out for the participant:
    - All physiological equipment, complete with new sensors
    - Headset
    - Caller ID paper slips
    - Pen
    - Participant Instructions
    - PowerPoint booklet
    - Call instruction guide (should be taped to the desk)
  - Get E-Prime ‘started’ on the rating computer
  - Turn on the BioData system and make sure:
    - The participant’s name is entered into the BioData system
    - The BioData system/Bluetooth system is working!

*Once the participant arrives, have the participant sit at the computer where Skype is already loaded on the screen.*

“Hi, my name is ________________, and I am the primary researcher for today’s study. Thank you for participating in our research. The goal of this study is to explore customer service in call centers. We are actually partnered with a local call center here in Akron who asked us to assist them with this project. Over the next hour, you will play the role of a service representative at University Tech Support and answer some calls via
Skype, which is an internet-based phone service that is used in the call center we are partnered with.

Today’s study is unique because we are partnering with another study occurring right now at another university in Ohio. Students in that study are creating a PowerPoint presentation and have been provided the phone number of our University Tech Support call center if they have any questions. You will be fielding calls from these students and assisting them through step-by-step instructions in the Microsoft PowerPoint manual we have created. Students enrolled in that study are competing for being entered into a lottery for $100 if they are the first to finish in their session, so it is important that you help them as best as you can. Because they think they are calling a real call center, it is imperative that you are happy, helpful, and respectful. Be sure to be positive and enthusiastic! They think they are calling a real technology help hotline.

In front of you are key materials for the study:

- **Caller ID Slips**: Before you can help the student, you must make sure that you ask for their participant ID number. You cannot help the student if they do not provide you their number. The researchers in the other study will use this ID number to see which of their participants called the hotline. When you get their ID number, just write it down on one of these caller slips and put the date and time.

- **Greeting and Closing Guidelines**: Taped to the desk is a list of instructions reviewing the steps that are to be followed with each call. Please make sure you remember to always say the greeting, follow-up, and closing statements.

- **PowerPoint Manual**: In front of you is the PowerPoint manual that University Tech Support has provided. Inside, you’ll find step-by-step instructions that you can run student callers through. University Tech Support piloted this training manual, and found that callers rarely asked questions that were not contained somewhere in this manual. When a caller asks for something, simply flip through the manual, find the correct heading, and read the instructions step-by-step.

Because you are working in a call center, it is important that you follow University Tech Support guidelines. **We want you to be happy, positive, and respectful for all customers** who are calling into the hotline. To help ensure this, **your calls will be monitored and recorded** via Skype for service quality. This is standard in all call centers. Customers who call into the call center will also be **rating you on the customer service you provide them**. We will be making sure that you **express happy and positive emotions no matter** what happens during the call. In order to show how sincere we are about instilling good customer service, **participants who have great customer service will be entered into a raffle to win one of two $50.00 gift cards to Amazon.com.** Your entry will be contingent on your performance, so **make sure to be as positive and enthusiastic as you can!**

I’m now going to leave you for a few minutes to review the training materials and the PowerPoint manual. I’ll be back in about five minutes.”
Leave the room. Enter after five minutes have passed.

“I hope you found the materials helpful. Do you have any questions?”

Answer any questions the participants have.

“Okay, to make sure you’re all set, I’m going to go into another room and call you and run through a brief scenario with you. This will let you practice the mandatory greeting and closing lines, and it will make sure you’re being as pleasant as we need you to be. When the phone rings on Skype, please answer it. We’ll keep it short.”

Leave and call participant.

Participant: Thank you for calling University Tech Support. My name is ______ and I am happy to serve you. Can I have your participant number please?

YOU: Sure. It’s 8714.

Participant: Thank you. How can I help you today?

YOU: Oh, I just need to insert a table.

Participant: OK – make sure you count now and find out how many columns or rows you need before we create your table.

YOU: No problem. I just need four columns and four rows.

Participant: OK – Click “Insert” at the top of the screen. Go down and click on “Table.”

YOU: Great. What’s next?

Participant: Use the up and down arrows next to “Columns” to increase or decrease the number of columns. Then use the up and down arrows next to “Rows” to increase or decrease the number of rows. Then, click “OK.” You can enter text into each box in the table as needed.

YOU: OK – great. It’s in.

Participant: Is there anything else that I can assist you with today? We are always happy to help you through any problem.

YOU: Nope!
Participant: Thank you so much for calling University Tech Support! It has been my pleasure to assist you. Have a great day!

After call, return to lab room. USE THIS TIME TO CORRECT THE PARTICIPANT IF THEY DID SOMETHING WRONG (i.e., missed a key “required” line from the manual, didn’t list instructions step by step, etc.).

“I think you can be even happier. Let your smile show through the phone!

Today’s study actually has a second component to it that I would like to walk you through today. After you are done taking calls, you’ll be providing some continuous ratings. To practice, I want to use the call we just made. Please move over to the computer next to you.”

IMPORTANT!! As participant moves, pull the sound file off of the ‘PowerGramo Player’ and export it as a ‘.wav’ file entitled one.wav. This should be exported to the hard drive. Unplug the hard drive, and plug it into the USB extension cord.

On the rating computer, pull up data file and put it into the ‘Calls’ folder on the desktop. You are then ready to run the ePrime program. To start the program, click the purple man on the tool bar.

“For these continuous ratings, you’ll be making three ratings per phone call. The first rating is about how you feel emotionally during the call. You will see a scale on the computer screen ranging from negative emotions on the left to positive emotions on the right, with the numbers 1-20 on the scale. During the call, you will be able to move the mouse as much or as little as you want along the scale depending on how you’re feeling. It is important to make sure you keep the mouse within the boxes at all times. Do you have any questions? If not, you can press the space bar to make your first rating.”

Direct participant toward the instructions on the computer screen, and then instruct them to proceed with providing their continuous ratings of emotions. When they tell you they are done, hit ‘F1’ to go to the blank screen.

“Great! Do you have any questions so far?”

Answer participant questions.

“Since the current study is interested in the emotional experience of call center workers, the next two ratings you are going to provide have to do with how you regulated your emotions during the call. In the first rating, I would like to ask you to provide continuous ratings of your experience of deep acting. Deep acting is when you put forth effort to try and really feel positively about a situation, or an interaction with a customer. For example, you might really try to feel positively so you can be pleasant with the customer. As you listen to the call, you can move the mouse along the scale depending on how
much you felt you were trying to genuinely feel the emotions needed with the caller at that point in time.

For these ratings, a rating of 1 reflects not deep acting at all. For a rating of “1,” you could say, “I am not experiencing any effort trying to change my emotions.” For a rating of “10,” you could say, “I am putting in a moderate amount of effort to try to feel the emotions I am supposed to express with the caller at this time.” For a rating of “20,” you could say, “I am making a full effort to actually feel the emotions that I need to express with the call at this time.”

As a reminder, you can move the mouse as much or as little as you want depending upon your experience with the caller. Please stay within the boxes, and make sure you are hovering over one of the numbers at all times.

Again, you can move the mouse as much or as little as you like. This means you can stay in one place on the scale, or you can move it a lot. This depends upon your personal emotional experience! Please stay within the boxes, and make sure you are hovering over one of the numbers at all times.

Do you have any questions about what you’re supposed to be rating?

*IF YES: Clarify. IF NO: Press F1 to bring up the surface acting screen.*

Let participant practice making the surface acting rating.

Now that you have rated deep acting, I would like you to provide ratings of your experience of surface acting. *Surface acting means that you felt like you had had to hide how you were really feeling from the customer.* For example, you might have felt negative emotions, but you hid them and displayed positive emotions to the customer instead. As you listen to the call, you can move the mouse along the scale depending on how much you felt you were hiding your emotions with the caller at that point in time.

For these ratings, you will also have a continuous scale ranging from 1-20. A rating of “1” reflects not having to surface act at all. For a rating of “1,” you could say, “I am not hiding my emotions at this time with the caller.” For a rating of “10,” you could say, “I am hiding my emotions some at this time with the caller, but I am not completely acting with them.” For a rating of “20,” you could say, “I am hiding what I really feel so the customer thinks I’m really happy when I’m not.”

Do you have any questions about what you’re supposed to be rating?

*IF YES: Clarify. IF NO: Press F1 to bring up the deep acting screen.*

Let participant practice making the surface acting rating.

“You’re all finished now with the practice ratings. Please go back to the call center computer.
At the request of the call center we are partnered with, we are also going to be recording physiological measures today. This is fairly new for the company, so I appreciate your cooperation. Is it okay for me to proceed? We have tested these measures on ourselves, and we can assure you that they do not create any discomfort.

Before we put on the physiological sensors, you’ll first want to put on the headset.

These three chest sensors are also going to track heart rate and pulse. For the red sensor, please place the adhesive under your rib cage on the left. For the white sensor, please place it on your collar bone on the left above the red sensor. For the black sensor, please place it on your collar bone on the right parallel to the white sensor.

On your left hand, we are going to attach three sensors. The first two (VELCRO) go on your ring and middle finger. These detect skin conductance, which is an indicator of stress. The finger grip (GREY CLAMP) goes on your index finger. This detects heart rate and pulse.

For the physiological measures to work, it’s really important that you leave your hand we are using completely still, and that you don’t move around in your seat a lot. If for any reason you feel uncomfortable with the physiological measurements, please let me know.

*Hook participant up to sensors and hand devices.*

OK – you’re set! I’ll call the other university to let them know your line is open – their students have been working for about 15 minutes now, so you might get some calls pretty quickly. In order to help monitor the physiological data, I will be staying in the room with you. However, I am unable to help you with any portion of the task.

Remember: we’re monitoring you for your emotional expressions, so be happy and don’t let students in the other study hear any negative emotions! Please keep answering the calls until I return. We’ll do more ratings once you are finished with all the calls.”

*Sit at the physiological computer and load the physiological program. Create a new participant session with the participant’s name and ID number.*

THEN: CALL THE CONFEDERATE! Say: “Hey, Pete. Yep, we’re live.” Confederates should wait 20 seconds from that point to begin the first call. The physiological data should start recording immediately (Click ‘Record’). When the participant answers the phone, press the power button on the wireless device quickly one to ‘mark’ the physiological data for call start.
APPENDIX C

PARTICIPANT INFORMATION

Thank you for your participation in this experiment!

The goal of this research is to explore customer service in call centers. This project is in partnership with a local call center in Akron that is interested in improving the customer service experience in their business. Your participation today is very helpful in this endeavor.

For this project, the investigators of this study are collaborating with a team of researchers who are conducting a concurrent study at another university in the Midwest. Your role is to answer calls (via Skype, an internet-based phone service) from other students participating in the other experiment. The callers (i.e., the participants in the other experiment) have been told that the University Tech Support (UTS, i.e., you) will be on hand to assist them with potential problems during the task they need to complete. The first five people to finish the other experiment will be entered in a drawing to win $100 so it is important that you help them as best as you can. You will be provided with thorough instructions for answering any questions they might have and helping them complete their task.

To ensure realism, you are asked to treat the callers as if you were an actual customer service representative working for University Technology Services. Please remember that the “customers” calling you do not know that you are participating in an experiment. They believe that you are an employee of UTS. Therefore, it is imperative that you treat participants as actual customers and not inform them that you are also part of a study.

You will ask each caller at the beginning of the call to give you their ID number. Since this study is linked to another study, the investigators need to be able to link a “UTS Rep” to a “customer.” The other researchers are interested in which of their participants called one of our participants for help. Participants in the other experiment have been told that the UTS Rep will ask them for their ID number.

As part of your role as a customer service representative, please treat these interactions as if they were actual conversations with real customers. We want positive, friendly service all the time and every time! Remember to:
- Maintain a pleasant demeanor at all times.
- Remember, a smile can be transmitted even through one’s tone of voice!
- Do not act irritated or unfriendly in any way to the customers.
- Always remain courteous and pleasant toward the customers.
- Remember, the customer is *always* right!

Your performance in this task will be audio-recorded. As a reminder, the top performers of this experiment will be entered to win one of two $50 gift cards to Amazon.com. Therefore, make sure you are providing service with a smile!

Please remember to continue taking the phone calls until the researcher returns to the room. If you have any questions, please feel free to find the researcher at any time.
University Tech Support (UTS) is dedicated to providing service excellence to all faculty, students, and staff members. Our motto is “Better service, better tech support!”

**Job Description for “Customer Service Representative”**

1. You work in the customer service department in the “Help Desk” division. Your primary job duty is to provide help with any programs our customers are using to ensure that they continue to use UTS sponsored software and help services.

2. You interact with customers via Skype, an internet-based phone service. Because we are committed to providing high quality customer service, all phone calls are audio taped through the Skype system. This allows us to play back the calls and monitor performance at any time.

3. In providing high quality customer service, we ask that you maintain a positive demeanor at all times. We want our customers to know you are smiling at them even though they are a phone line away!
Customer Service Phone Etiquette

1. **Greeting**: always greet the customers when they call you. Our standard greeting for all customers is:

   “Thank you for calling University Tech Support! My name is _____ and I am happy to serve you. May I please have your ID number?”

2. **ID Number**: It is imperative that you write down the ID number of each person who calls. Do not proceed with the phone call until you receive this number.

   “Thank you for your ID number. How may I assist you today?”

3. **Always Follow-Up**: during the phone call, customers may have multiple questions or requests. After each question, we ask our customer service representatives to politely ask the following:

   “Is there anything else that I can assist you with today? We are always happy to help you through any problem.”

4. **Closing**: At the end of each phone call, we want our customers to know that it has been a pleasure helping them. Our standard closing to all customers is:

   “Thank you so much for calling University Tech Support! It has been my pleasure to assist you. Have a great day!”
**PowerPoint Frequently Asked Questions**

The following FAQ is designed to help you handle frequently asked customer questions in regard to the PowerPoint program. We have conducted pilot studies about the questions that emerge from customers, and it is very rare that customers will call with problems other than the ones listed below.

When a customer calls, simply walk them through the step-by-step instructions outlined below to help the customer with his or her problem. You do not need any prior knowledge about PowerPoint to help them!

**How do I insert a text box into my presentation?**
1. Go to “Insert” at the top of the computer screen.
2. Scroll down to “Text Box.”
3. Your mouse on the screen should now change shape.
4. In the PowerPoint slide you want to draw the text box, click the mouse.
5. Holding the mouse down, draw your text box to the desired size.
6. Once at the desired size, release the mouse.
7. Proceed to enter in the desired text.

**How do I change the slide layout?**
1. Click on “Format” at the top of the screen.
2. Go down and click on “Slide Layout.”
3. There are four general categories of layouts to choose from.
   a. Ask customers if they are making a slide with text, pictures, or both.
4. FOR TEXT: Go to the subheading that says “Text Layouts” and choose.
5. FOR PICTURES: Go to the subheading that says “Context Layouts” and choose.
6. FOR BOTH: Go to the subheading that says “Text and Content Layouts” and choose.
7. Exit out of “Slide Layout” on the top right of the screen once done selecting.

**How do I put a border around my text box?**
1. Click onto the text in the text box; the border should become selected.
2. Right click on the border around the text box.
3. Click on “Format Text Box.”
4. A number of tabs will appear on the top that you can select.
5. Click on “Colors and Line.”
7. Click the drop down menu that says “No Line” and select a color option.
   Automatic would be black.
8. Click OK to add the border.
How do I insert a picture from the Internet?
1. PRIOR TO BEGINNING, MAKE SURE CUSTOMERS HAVE SAVED THE DESIRED IMAGE ONTO THE DESKTOP OF THEIR COMPUTERS.
2. Images on the desktop should be saved as a .gif or .jpg file.
3. Click on “Insert” at the top of the screen.
4. Go down and click on “Picture.”
5. Once clicked on “Picture,” click on “From File.”
6. Click on “Desktop” on the right side of the screen.
7. Select your image.
8. Click on “Insert.”
9. TO CHANGE THE SIZE: Click on picture, and go to the corner of the picture. A diagonal arrow should appear. Drag arrow in towards the center of the picture to make the picture smaller. Drag arrow away from the center of the picture to make the picture larger.

How do I insert a Table?
1. PRIOR TO BEGINNING, MAKE SURE CUSTOMERS KNOW HOW MANY COLUMNS OR ROWS THEY WANT TO CREATE IN THEIR TABLE.
2. Click on “Insert” at the top of the screen.
3. Go down and click on “Table.”
4. Use the up and down arrows next to “Columns” to increase or decrease the number of columns.
5. Use the up and down arrows next to “Rows” to increase or decrease the number of rows.
6. Click “OK.”
7. Enter text into each box in the table as needed.

How do I change the width of the border around my text box?
1. Click onto the text in the text box; the border should become selected.
2. Right click on the border around the text box.
3. Click on “Format Text Box.”
4. A number of tabs will appear on the top that you can select.
5. Click on “Colors and Line.”
7. Before changing the width, make sure you have selected a color.
   a. IF THEY HAVE NOT SELECTED A COLOR: Click on the drop down menu that says “No Line” and select a color option.
8. In the “Line” subsection, there will be a box that says “Weight.”
9. Use the up and down arrows next to “Weight” to change the width of the text box.
10. Click OK to change the weight.
**How do I insert additional columns and/or rows into a Table?**

1. Click on “View” at the top of the screen.
2. Go down and click on “Toolbars.”
3. Select “Tables and Borders.”
4. A new toolbar should appear – click on the tab that says “Table” at the bottom of the toolbar.
5. Four options will appear to either insert columns to the left/right of where the mouse is in the table or to insert rows above/below where the mouse is.
   a. IF THE OPTIONS ARE NOT IN BOLD: You must click your mouse in a box inside the table in order to get these options to appear.
   b. IF A COLUMN/ROW IS INSERTED IN THE WRONG PLACE: With your mouse, highlight all of the boxes/columns in the row you wish to delete. Once highlighted, right click the boxes and click on “Delete Columns/Rows”
6. Once inserted, enter in text as needed.

**How do I animate the slides?**

1. Click on “Slide Show” at the top of the screen.
2. Go down and click on “Animation Schemes.”
3. A side bar will appear on the right side of the computer screen.
4. Choose the animation that best suits your needs.
   a. IF ASKED FOR RECOMMENDATION: Subtle animation is usually more appropriate for professional presentations than the animation schemes labeled “Exciting.”
5. Click “Apply to All Slides” to animate all of the PowerPoint Slides the same.
APPENDIX E

SERVICE FAILURE SCRIPT

In this condition, your goal is to make the subject’s task as difficult as possible. However, you need to start neutral and build up to being more negative. As the call goes on, you can sound irritated, rushed, angry, etc. Don’t insult the subjects or get personal, and don’t use profanity or inappropriate name-calling. However, be difficult and hard to deal with the entire call. You should be at your highest level of negativity towards the end. The call should last 3-4 minutes.

Participant: Thank you for calling University Tech Support. My name is _____ and I am happy to serve you. Can I have your participant number please?

YOU: Umm, yeah. It’s 2890.

Participant: Okay, great. How can I help you today?

YOU: Okay, I need to insert some text boxes and I just can’t remember how to do this. Do you have the steps to do that?

Participant: Yes – first…

YOU: (CUT PARTICIPANT OFF) Wait, how many steps does it take? (LET PARTICIPANT FULLY EXPLAIN ALL STEPS TO YOU, ONLY AGREEING AS HE/SHE GOES THROUGH THEM)

Participant: Seven. It shouldn’t be too bad. First, go to “Insert at the top of the computer screen. Then scroll down to “Text Box.” Your mouse on your screen should now change shape. In the PowerPoint slide, move your mouse to where you want the box, click down, and draw the box to the desired size. When you’re at the desired slide, you can release the mouse. You can then enter the desired text.

YOU: Wait a second… why isn’t there a border?

Participant: Well, there are other steps for creating the border.
YOU:  (START GETTING ANGRY) WHAT?! Are you kidding me? I’m on a time crunch here with this stupid experiment and now you’re telling me there are extra steps? Why didn’t you tell me that before?!?!

Participant:  I’m sorry, I didn’t know…

YOU:  Seriously?! Who makes a text box with no border? Now I can’t even see where the box is on the screen! This is ridiculous… what a waste of time. (PAUSE) WELL? Are you going to help me or not?

Participant:  Yes. Click on the text box…

YOU:  Easier said than done since I can’t even see the text box on the stupid screen anymore. Ughhh… now I have to redo the textbox ALL OVER AGAIN. Don’t you know how to do your job at all?? (UNDER YOUR BREATH) This is completely ridiculous…

Participant:  I’m sorry, I am trying to help.

YOU:  Well you’re doing a pretty terrible job at it, don’t you think?!

Participant:  I’m sorry, I didn’t mean to.

YOU:  Whatever. I’m over this! Just tell me how to make the line appear on the text box. And don’t start at the beginning! Just tell me how to change the line color. I found the “Colors and Line” tab already.

Participant:  Click on “Colors and Line” on the tabs. Then go to the subsection entitle “Line.” Click the drop down menu that says “No Line” and select a color option. Click OK.

YOU:  Okay – fine. There’s the border. How do I change the width of this thing?

Participant:  You’re going to right click the border of the text box and click on “Format Text Box…”

YOU:  (GETTING IRATE) ARE YOU KIDDING ME?! I am being TIMED and you’re telling me we could have done this already?!!! Why do you keep doing this?! Why can’t you just tell me everything at once? Haven’t you made a PowerPoint before?

Participant:  I’m sorry. I was just doing what you asked.

YOU:  You know, this doesn’t feel a lot like a help line right now. I feel like you are wasting my time here with all these steps!!
Participant: I’m sorry. I was just trying to help you.

YOU: You know, if YOU were going through this study that I am right now you wouldn’t have time for such terrible service either! I can’t believe you’re not helping me change the width of this stupid text box. Are you going to do ANYTHING?!

Participant: You can change the width of the box by going back to the “Format Text Box” and changing the width…

YOU: You know what? Thanks for nothing. I’ll figure this out on myself. (HANG UP)
APPENDIX F

SERVICE RECOVERY SCRIPT

In this condition, your goal is to make the subject’s task as difficult as possible but then become positive towards the end. You need to start neutral (pleasant, but not overly nicely), become angry, and then become positive. As you get angry, you can sound irritated, rushed, angry, etc. Don’t insult the subjects or get personal, and don’t use profanity or inappropriate name-calling. However, be difficult during the call until you start becoming positive again. As you become positive, you can laugh, sigh in relief, sound excited, etc. The call should last 3-4 minutes.

Participant: Thank you for calling University Tech Support. My name is ______ and I am happy to serve you. Can I have your participant number please?

YOU: Umm, yeah. It’s 2890.

Participant: Okay, great. How can I help you today?

YOU: Okay, I need to insert some text boxes and I just can’t remember how to do this. Do you have the steps to do that?

Participant: Yes – first…

YOU: (CUT PARTICIPANT OFF) Wait, how many steps does it take? (LET PARTICIPANT FULLY EXPLAIN ALL STEPS TO YOU, ONLY AGREEING AS HE/SHE GOES THROUGH THEM)

Participant: Seven. It shouldn’t be too bad. First, go to “Insert at the top of the computer screen. Then scroll down to “Text Box.” Your mouse on your screen should now change shape. In the PowerPoint slide, move your mouse to where you want the box, click down, and draw the box to the desired size. When you’re at the desired slide, you can release the mouse. You can then enter the desired text.

YOU: Wait a second… why isn’t there a border?

Participant: Well, there are other steps for creating the border.
YOU: (START GETTING ANGRY) WHAT?! Are you kidding me? I’m on a time crunch here with this stupid experiment and now you’re telling me there are extra steps? Why didn’t you tell me that before?!?!

Participant: I’m sorry, I didn’t know…

YOU: Seriously?! Who makes a text box with no border? Now I can’t even see where the box is on the screen! This is ridiculous… what a waste of time. (PAUSE) WELL? Are you going to help me or not?

Participant: Yes. Click on the text box…

YOU: Easier said than done since I can’t even see the text box on the stupid screen anymore. Ughhh… now I have to redo the textbox ALL OVER AGAIN. Don’t you know how to do your job at all?? (UNDER YOUR BREATH) This is completely ridiculous…

Participant: I’m sorry, I am trying to help.

YOU: Well you’re doing a pretty terrible job at it, don’t you think?!

Participant: I’m sorry, I didn’t mean to.

YOU: Whatever. I’m over this! Just tell me how to make the line appear on the text box. And don’t start at the beginning! Just tell me how to change the line color. I found the “Colors and Line” tab already.

Participant: Click on “Colors and Line” on the tabs. Then go to the subsection entitled “Line.” Click the drop down menu that says “No Line” and select a color option. Click OK.

YOU: Okay – fine. There’s the border. How do I change the width of this thing?

Participant: You’re going to right click the border of the text box and click on “Format Text Box…”

YOU: Wait… it’s back under that tab again? (Calming down)

Participant: Yes…

YOU: Okay, well, at least I know where to go already, right? (laugh slightly to indicate positive emotional tone)

Participant: Yes. You just go back there and you should see a section where you can click “Weight.”
YOU: Yep! I see it. I appreciate you staying so calm. I’m just tired from this experiment.

Participant: Okay, just click which width you want right there.

YOU: Great! There are a lot of options. When I make my choice, do I just click okay?

Participant: Yep – click okay and you’re done.

YOU: Phew. Okay. That’s done. Wow! I guess that’s it! You were a great help.

Participant: Thank you so much for calling University Tech Support! It has been my pleasure to assist you. Have a great day!
APPENDIX G

POST-SIMULATION SCRIPT

After the call ends, wait 15 seconds and then say:

“I just received a message from the coordinator of the other study saying that was their last participant. So, you’re done taking phone calls for today! Please remove the sensors and the headset and we’ll continue with the continuous ratings.”

Remove sensors.

“We’re now going to finish up your continuous ratings. Do you have any questions about the scales before you use them again? You’ll be prompted with the same instructions on the screens to refresh your memory about the different scales. You can press the space bar to begin.”

Have participants go through the three scales. Stay in the room to make sure they complete everything and there are no questions. Before each rating scale, restate what the scale measured and ask if they would like to hear the instructions again.

“Great! You’re done with those. To finish, we have some final scales for you to complete. After that, you’ll receive your debriefing and you’re free to go.”

DEBRIEFING

'Thank you for participating in today's study. Call center simulations are important for studying and understanding employee well-being at work due to the high levels of emotional demands call center workers experience. Today, you did not actually receive calls from real customers, but rather confederate callers who were trained to act in the manner they did. You are automatically entered into the raffle for one of two $50.00 gift cards to Amazon.com, and your HPR credit will be posted within the next 48 hours.

We greatly appreciate your participation and cooperation in today's study.'

224
APPENDIX H

PRE-SIMULATION MEASURES

Demographics

- Age (in years)
- Sex
- Ethnicity (Caucasian, African American, Asian American, American Indian, Hispanic, Other [please specify])
- Do you work full-time or part-time?
  - IF YES: How many hours per week do you work on average?
  - IF YES: How long have you worked with your current company?
  - IF YES: What is your job title?
  - IF YES: Please list your primary job duties involved in this position.
- Have you ever had any job experience in a service occupation?
  - IF YES: Please list the name of each service job/job title, the primary job duties, and the length of time worked for each job.
  - Repeat for up to five occupations.
- Have you ever worked in a call center?
  - IF YES: Please provide the length of time you worked at the call center, as well as your primary job duties.

Chronic Emotion Regulation

Emotion Regulation Questionnaire (Gross & John, 2003)
Instructions: Please read the following items and rate the extent to which you agree or disagree on a 7-point scale (1 = “strongly disagree;” 7 = “strongly agree”).

Reappraisal Items (Cronbach’s alpha = .77):
1. I control my emotions by changing the way I think about the situation I’m in.
2. When I want to feel less negative emotion, I change the way I’m thinking about the situation.
3. When I want to feel more positive emotion, I change the way I’m thinking about the situation.
4. When I want to feel more positive emotion (such as joy or amusement), I change what I’m thinking about.
5. When I want to feel less negative emotion (such as sadness or anger), I change what I am thinking about.
6. When I’m faced with a stressful situation, I make myself think about it in a way that helps me stay calm.
Suppression Items (Cronbach’s alpha = .80):
7. I control my emotions by not expressing them.
8. When I am feeling negative emotions, I make sure not to express them.
9. I keep my emotions to myself.
10. When I am feeling positive emotions, I am careful not to express them.

Emotional Expressivity
Emotional Expressivity Scale (Gross & John, 1995)
Instructions: Please read the following items and rate the extent to which you agree or disagree with the following items on a 7-point scale (1 = “strongly disagree;” 7 = “strongly agree”).

Positive Emotional Expressivity (Cronbach’s alpha = .66):
1. When I’m happy, my feelings show.
2. Whenever I feel positive emotions, people can easily see exactly what I am feeling.
3. I laugh out loud when someone tells me a joke that I think is funny.
4. I am an emotionally expressive person.
*“I laugh out loud when someone tells me a joke that I think is funny” was problematic for the scale; with its removal, Cronbach’s alpha = .72.

Negative Emotional Expressivity (Cronbach’s alpha = .65):
5. People often do not know what I am feeling. (R)
6. Whenever I feel negative emotions, people can easily see exactly what I am feeling.
7. What I’m feeling is written all over my face.
8. No matter how nervous or upset I am, I tend to keep a calm exterior. (R)
9. It is difficult for me to hide my fear.
10. I’ve learned it is better to suppress my anger than to show it. (R)
*“I’ve learned it is better to suppress my anger than to show it” was problematic for the reliability of the scale; with its removal, Cronbach’s alpha = .73.

Action State Orientation
Action-State Orientation Scale Revised (Diefendorff, Hall, Lord, & Strean, 2000)
Instructions: For each item, please select which statement best represents you.

Preoccupation (Cronbach’s alpha = .61):
1. If I’ve worked for weeks on one project and then everything goes completely wrong with the project:
   a. It takes me a long time to adjust myself to it.
   b. It bothers me for a while, but then I don’t think about it anymore.
2. If I had just bought a new piece of equipment (for example, a MP3 player) and it accidentally fell on the floor and damaged beyond repair:
   a. I would manage to get over it quickly.
   b. It would take me a long time to get over it.
3. If I have to talk to someone about something important and, repeatedly, can’t find him or her at home:
   a. I can’t stop thinking about it, even when I’m doing something else.
   b. I easily forget about it until I see the person.
4. When I am told that my work has been completely unsatisfactory:
   a. I don’t let it bother me for too long.
   b. I feel paralyzed.
5. If I’m stuck in traffic and miss an important appointment:
   a. At first, it’s difficult for me to start to do anything else at all.
   b. I quickly forget about it and do something else.
6. When something really gets me down:
   a. I have trouble doing anything at all.
   b. I find it easy to distract myself by doing other things.
7. When several things go wrong on the same day:
   a. I usually don’t know how to deal with it.
   b. I just keep on going as though nothing had happened.
8. When I have put all my effort into doing a really good job on something and the whole thing doesn’t work out:
   a. I don’t have too much difficult starting something else.
   b. I have trouble doing anything else at all.

Hesitation (Cronbach’s alpha = .67):
1. When I know I must finish something soon:
   a. I have to push myself to get started.
   b. I find it easy to get it done and over with.
2. When I don’t have anything in particular to do and I am getting bored:
   a. I have trouble getting up enough energy to do anything at all.
   b. I quickly find something to do.
3. When I am getting ready to tackle a difficult problem:
   a. It feels like I am facing a big mountain that I don’t think I can climb.
   b. I look for a way that the problem can be approached in a suitable manner.
4. When I have to solve a difficult problem:
   a. I usually don’t have a problem getting started on it.
   b. I have trouble sorting things out in my head so that I can get down to working on the problem.
5. When I have a lot of important things to do and they must all be done soon:
   a. I often don’t know where to begin.
   b. I find it easy to make a plan and stick with it.
6. When I have to take care of something important which is also unpleasant:
   a. I do it and get it over with.
   b. It can take a while before I can bring myself to it.
7. When I am facing a big project that has to be done:
   a. I often spend too long thinking about where I should begin.
   b. I don’t have any problems getting started.
8. When I have an obligation to do something that is boring and uninteresting:
   a. I do it and get it over with.
   b. It can take a while before I can bring myself to do it.

**Volatility** (Cronbach’s alpha = .43):
1. When I have learned a new and interesting game:
   a. I quickly get tired of it and do something else.
   b. I can really get into it for a long time.
2. When I read an article in the newspaper that interests me:
   a. I usually remain so interested in the article that I read the entire article.
   b. I still often skip to another article before I’ve finished the first one.
3. When one of my co-workers brings up an interesting topic for discussion:
   a. It can easily develop into a long conversation.
   b. I soon lose interest and want to go do something else.
4. When I am busy working on an interesting project:
   a. I need to take frequent breaks and work on other projects.
   b. I can keep working on the same project for a long time.
5. When I read something I find interesting:
   a. I sometimes still want to put the article down and do something else.
   b. I will sit and read the article for a long time.
6. When I am trying to learn something new that I want to learn:
   a. I’ll keep at it for a long time.
   b. I often feel like I need to take a break and go do something else for a while.

**Dispositional Positive and Negative Affect**
The Positive and Negative Affectivity Schedule (Watson, Clark, & Tellegen, 1988)
Instructions: Please rate the extent to which you feel the following emotions *in general* on a 5-point scale (1 = “very slightly to not at all;” 5 = “extremely”).

**Positive Affect** (Cronbach’s alpha = .85):
Interested
Alert
Excited
Inspired
Strong
Determined
Attentive
Enthusiastic
Active
Proud

**Negative Affect** (Cronbach’s alpha = .88):
Irritable
Distressed
Ashamed
Upset
Nervous
Guilty
Scared
Hostile
Jittery
Afraid

Personality – Neuroticism and Conscientiousness
Big-10 Personality “Mini-Markers” from Saucier (1994)
Instructions: Rate the extent to which you agree or disagree that each of the following items describes who you are as a person (1 = “strongly disagree;” 5 = “strongly agree”).

Conscientiousness (Cronbach’s alpha = .89):
1. Organized
2. Efficient
3. Systematic
4. Practical
5. Disorganized (R)
6. Sloppy (R)
7. Inefficient (R)
8. Careless (R)

Neuroticism (Cronbach’s alpha = .78):
1. Unenvious (R)
2. Relaxed (R)
3. Moody
4. Jealous
5. Temperamental
6. Envious
7. Touchy
8. Fretful

Self-Monitoring
Revised Self-Monitoring Scale (Lennox & Wolfe, 1984)
Instructions: For each statement indicate your general perception of your abilities on the following 7-point scale (1 = “very strongly disagree;” 7 = “very strongly agree”).

Ability to Modify Self-Presentation (Cronbach’s alpha = .75):
1. Once I know what the situation calls for, it’s easy for me to regulate my actions accordingly.
2. I have found that I can adjust my behavior to meet the requirements of any situation I find myself in.
3. I have trouble changing my behavior to suit different people and different situations. (R)
4. In social situations, I have the ability to alter my behavior if I feel that something else is called for.
5. I have the ability to control the way I come across to people, depending on the 
impression I wish to give them.
6. When I feel that the image I am portraying isn’t working, I can readily change to 
something that does.
7. Even when it might be to my advantage, I have difficulty putting up a good front. (R) 
**Sensitivity to Expressive Behaviors of Others** (Cronbach’s alpha = .82)
1. I can usually tell when I’ve said something inappropriate by reading it in the listener’s 
eyes.
2. My powers of intuition are quite good when it comes to understanding others’ 
emotions and motives.
3. If someone is lying to me, I usually know it at once from that person’s manner of 
expression.
4. In conversations, I am sensitive to even the slightest change in facial expression of the 
person I’m conversing with.
5. I am often able to read people’s true emotions correctly through their eyes.
6. I can usually tell when others consider a joke to be in bad taste, even though they may 
laugh accordingly.
7. I tend to be attentive to the reactions of others to my behavior.

**Customer Orientation**
Customer Orientation Scale – Intrinsic Motivation items only (Allen, Pugh, Grandey, & 
Groth, 2010); adapted from Donavan, Brown, & Mowen (2004) (Cronbach’s alpha = .95)
Instructions: For each statement, please indicate the extent to which you agree or disagree 
(1 = “strongly disagree;” 5 = “strongly agree”).
1. I enjoy nurturing my service customers.
2. I take pleasure in making every customer feel like he/she is the only customer.
3. Every customer’s problem is important to me.
4. I thrive on giving individual attention to each customer.
5. I enjoy delivering the intended services on time.
6. I find a great deal of satisfaction in completing tasks precisely for customers.
7. I enjoy having the confidence to provide good service.
8. I enjoy getting to know my customers personally.
APPENDIX I

PHYSIOLOGICAL MEASURES

Physiological measures were recorded using BioTrace+ Software for NeXus-10 Physiological Monitoring and Feedback System. This data system had three sets of sensors that were used to record physiological data continuously from participants during the confederate phone call.

Three sensors with electrodes were attached to the participant’s chest for electrocardiogram monitoring (ECG/EKG); sensors were adhered below the left collar bone, right collar bone, and on the lower left hand side of the participant’s chest below the rib cage. Data were collected at 256 sps (samples per second).

A finger clamp was attached to the participant’s index finger on his/her non-dominant hand. This clamp assessed blood volume pulse (BVP) at 128 sps. Blood volume pulse assesses heart rate and measures relative blood flow. This sensor is often used out of ease in comparison to ECG/EKG sensors; however, it is more susceptible to measurement artifact given that any movement will put ‘noise’ in the data.

Finally, two velcro sensors were attached to the third and fourth fingers on the participant’s non-dominant hand (i.e., three sensors total were on the participant’s non-dominant hand). These sensors monitored galvanic skin response (GSR) at 32 sps. GSR measures the sweat response in the skin, which is a measure of sympathetic nervous system arousal.
APPENDIX J

POST-SIMULATION MEASURES

**Emotional Exhaustion**
Adapted Job-Related Emotional Exhaustion Scale Items (Erickson & Ritter, 2001)
Instructions: When thinking about the call center experiences you just had, please rate the extent to which you agree or disagree with the following statements (1 = “strongly disagree;” 7 = “strongly agree”).
1. I feel emotionally drained from this call center task.
2. I feel used up right now.
3. This call center task really put a lot of strain on me.
4. I feel burned out from this call center task.
5. This call center task put too much stress on me.
6. I feel I worked too hard on this call center task.

**In General Emotional Labor**
Emotional Labor Strategy Scale by Diefendorff, Croyle, and Gosserand (2005); certain items were from Kruml and Geddes (2000) and Grandey (2003)
Instructions: For the following items, rate the extent to which you agree or disagree with each statement in general in regards to your experience in today’s call center. (1 = “strongly disagree;” 5 = “strongly agree”)

**Surface Acting** (Cronbach’s alpha = .92)
1. I put on an act in order to deal with customers in an appropriate way.
2. I fake a good mood when interacting with customers.
3. I put on a “show” or “performance” when interacting with customers.
4. I just pretend to have the emotions I need to display for my job.
5. I put on a “mask” in order to display the emotions I need for the job.
6. I express feelings to customers that are different from what I feel inside.
7. I fake the emotions I show when dealing with customers.

**Deep Acting** (Cronbach’s alpha = .91)
8. I try to actually experience the emotions that I must show to customers.
9. I make an effort to actually feel the emotions that I need to display toward others.
10. I work hard to feel the emotions that I need to show to customers.
11. I work at developing the feelings inside of me that I need to show to customers.
Realism of Call Center Experience
Designed for this study (Cronbach’s alpha = .78).
Instructions: As you reflect back on your call center experience today, please state how much you agree with the following questions on the below 5-point scale (1 = “strongly disagree;” 5 = “strongly agree”).
1. This call center experience was realistic.
2. I felt like a real call center worker.
3. My interaction with the student caller was believable.

Post-Simulation Affect
The Positive and Negative Affectivity Schedule (Watson, Clark, & Tellegen, 1988)
Instructions: Please rate the extent to which you currently feel the following emotions on a 5-point scale (1 = “very slightly to not at all;” 5 = “extremely”).
Positive Affect (Cronbach’s alpha = .88):
Interested
Alert
Excited
Inspired
Strong
Determined
Attentive
Enthusiastic
Active
Proud

Negative Affect (Cronbach’s alpha = .88):
Irritable
Distressed
Ashamed
Upset
Nervous
Guilty
Scared
Hostile
NOTICE OF APPROVAL

September 21, 2012

Alison S. Gabriel
1212 Monument Road NW
Contact, Ohio 44103

From: Sharon McWhorter, IRB Administrator

Re: IRB Number 201200977 "Continuous Ratings of Emotion Regulation"

Thank you for submitting an IRB Application for Review of Research Involving Human Subjects for the referenced project. Your protocol represents minimal risk to subjects and has been approved under Expedited Categories A/F.

<table>
<thead>
<tr>
<th>Approval Date:</th>
<th>September 20, 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuation Application Due:</td>
<td>September 20, 2013</td>
</tr>
</tbody>
</table>

In addition, the following is/are approved:

☐ Waiver of documentation of consent
☐ Waiver or alteration of consent
☐ Research involving children
☐ Research involving prisoners

Please adhere to the following IRB policies:

- IRB approval is given for no more than 12 months. If your project will be active for longer than one year, it is your responsibility to submit a continuation application prior to the expiration date. We request submission two weeks prior to expiration to ensure sufficient time for review.
- A copy of the approved consent form must be submitted with any continuation application.
- If you plan to make any changes to the approved protocol, you must submit a continuation application for change and it must be approved by the IRB before being implemented.
- Any adverse reactions/incidents must be reported immediately to the IRB.
- If this research is being conducted for a master's thesis or doctoral dissertation, you must file a copy of this letter with the thesis or dissertation.
- When your project terminates, you must submit a Final Report Form in order to close your IRB file.

Additional information and all IRB forms can be accessed on the IRB website at: http://www.usnews.edu/research/areas/compliance/IRBForms.php

Cc: James Difandorff - Advisor
Cc: Valerie Cardenas - IRB Chair

[Approved consent form enclosed]

The University of Akron is an Equal Education and Employment Institution.